Prospective Reserves of Life Insurance Policies under the Heath-Jarrow-Morton Framework

Åsmund Hausken Sande

Thesis submitted for the degree of
Master in Stochastic Modelling, Statistics and Risk Analysis
(Finance, insurance and risk)
30 credits

Department of Mathematics
Faculty of mathematics and natural sciences

UNIVERSITY OF OSLO

Spring 2021

Prospective Reserves of Life Insurance Policies under the Heath-Jarrow-Morton Framework

Åsmund Hausken Sande

© 2021 Åsmund Hausken Sande

Prospective Reserves of Life Insurance Policies under the Heath-Jarrow-Morton Framework

http://www.duo.uio.no/

Printed: Reprosentralen, University of Oslo

Abstract

In this thesis we consider a general stochastic interest rate under the HJM (Heath-Jarrow-Morton) framework. We further present a general model for the pricing of life insurance policies allowing for a wider range of stochastic policy functions than previously done within the HJM framework. This is carried out by modelling them under a general financial market model with Gaussian noise.

Furthermore, we develop standard pricing formulas based on financial arbitrage methods for both current time and future time-points. It is worth noting that these equations are contingent on formulas pricing the instantaneous values of the policy functions as financial claims.

Lastly, we give an example where the theory is applied to exactly evaluate the price of reserves within a new theoretical pension scheme with stochastic policy functions tied to the interest rate. As a part of this example we develop small generalizations of some financial pricing formulas for call and digital options on zero-coupon bonds.

In order to rigorously justify these results the thesis covers a large amount of background material. This includes measure and probability theory, as well as using these to introduce important concepts in interest rate, finance and classical insurance theory.

Acknowledgements

I would like to thank my friends and family for their support, in particular my father for sharing his academic experience and my sister for offering her insight into the world of economics. Lastly, I want to extend my gratitude to David R. Baños, my supervisor, for his support and guidance.

Contents

Chapter 1

Introduction

1.1 General Background

The discipline of actuarial science, on which this thesis will primarily focus, applies mathematical and statistical methods to fields such as insurance and finance. Mainly, we will be considering rigorous mathematical modelling of uncertainty in the field of life insurance.

The job of an actuary is to apply such actuarial methods in order to compute and analyse risk. In life insurance, for example, actuaries are often tasked with the study of mortality, which includes computing survival probabilities for different groups of people. Another important task in life insurance, which will be a major part of this thesis, is the pricing of insurance contracts. The payouts of these agreements, referred to as reserves, are determined by policy functions which state how much the insured is owed under different circumstances.

The methods used for such analysis has changed significantly over time. While dating back at as far as antiquity the field of actuarial science had its scientific beginnings in the 17th century. Here personal risk, like mortality, was studied alongside annuities, compound interest and eventually probability theory, all of which are essential for the field's foundation.

Early actuarial models were generally deterministic, but with the formalization of stochastic processes in the 20th century much more general methods have been introduced. The advent of modern computers further revolutionized the field and allowed more complicated models to be practically utilized. This has led to the possibility of stochastic models for both interest rates and policy functions. Along with these advancements there have been many concepts borrowed from mathematical finance, primarily in regards to bond and option pricing.

While such methods are available, many insurance models still utilize deterministic, or even constant, interest rates in their analysis. Even so, much work has been done with the application of stochastic short-rate models in the realm of life insurance, see Persson [10] or Zaglauer [12] for examples. However, a common problem with short-rate models, mentioned in Filipović [3], is the difficulty in fitting the model to the forward curve.

There are many ways of resolving this problem, for example one could use the BDT (Black-Derman-Toy) short-rate model. As mentioned in Gaillardetz

[5] "the BDT model is used by practitioners because it matches the current term structure of interest rates and the volatilities", where it is referred to [8] and others for a detailed justification.

Solutions like the BDT model, however, assumes a positive short-rate, and as recent times have shown, this is not always the case. We will in this thesis present a framework that allows for both easy integration of the entire forward curve in model calibration as well as allowing for negative interest rates. The HJM (Heath-Jarrow-Morton) forward rate framework allows for both of these as mentioned in Filipović [3], which is there justified by the initial forward curve explicitly appearing in the definition.

Unfortunately, as far as we are aware, not much work has been done with the pricing of reserves under direct modelling of forward rates within the general HJM framework. In Baccinello [1] this class of interest rate models is used to compute the current price of reserves for a limited class of stochastic policy functions, utilizing policy functions based on stock price models with constant drift and volatility. We will aim to generalize this setting by allowing for a much broader class of policy functions.

1.2 General Outline

In this thesis we will similarly consider a general stochastic interest rate under the HJM framework. We will create a general model allowing for a wider range of stochastic policy functions by modelling them under a general financial market model with Gaussian noise.

Furthermore, we will develop pricing formulas based on financial arbitrage methods for both current time and future time-points. It is worth noting that these equations will be contingent on formulas pricing the instantaneous values of the policy functions as financial claims. Lastly, we give an example where the theory is applied to exactly evaluate the price of reserves within a new theoretical pension scheme. In this example we will consider stochastic policy functions based on the price of zero-coupon bonds, thereby tying them directly to the forward rate.

In order to rigorously justify these results, the thesis will cover a large amount of background material. This will include measure and probability theory, as well as using these to introduce important concepts in interest rate, finance and classical insurance theory.

1.3 Structure of the Thesis

- We start in Chapter 2 by covering the necessary theoretical background material. Firstly, basic concepts in measure and probability theory are introduced, along with the main results needed to work with them. These concepts are then used to formulate the standard Lebesgue integral as well as giving our own thorough construction of the Lebesgue-Stieltjes integral based on a finite variation approach. Several important results regarding integration theory is also given, including the monotone and dominated convergence theorems. The final part of the chapter covers stochastic processes, conditional expectations and the Itô integral. This is then used

to define stochastic differential equations and several important classes of stochastic processes. Some relevant examples and applications are given along with our own proofs of several known results.

- Chapter 3 introduces many important concepts from mathematical finance as well as defining our market model. Arbitrage theory is then introduced leading into the two first fundamental theorems of asset pricing. These are applied to define arbitrage-free prices for attainable claims which will be essential for our later evaluation of reserves.

- We move on to Chapter 4 where we specify the HJM framework for forward rates together with some important results. Using this framework we introduce the concept of forward measures and apply these to formulate pricing formulas for European and digital options which generalizes earlier known results.

- In Chapter 5 we cover classical life insurance based on deterministic interest rates and policy functions. Explicit pricing formulas are presented along with two versions of Thiele's differential equation.

- The main part of our contribution starts in Chapter 6. Here, we specify our framework for modelling insurance policies with both stochastic policy functions and interest rates under HJM. We then prove two explicit pricing formulas analogous to results covered in Chapter 5. We then apply these equations in an example utilizing both our framework and the pricing formulas we developed in Chapter 4.

- The aforementioned example is further analysed in Chapter 7 where we apply numerical and statistical methods to demonstrate how the theory can be used. We here specify an exact interest model with stochastic policy functions in order to compute current and future reserve prices through simulation.

- In Chapter 8 a brief summary is given along with a lamentation of work left undone.

- Appendix A stores the relevant code used for exact computation and simulation in Chapter 7.

Chapter 2

Theoretical Background

2.1 Measure Theory

Nearly every part of the upcoming modelling work will include either integration, probability theory, or both. The underpinning of both fields is measure theory, the mathematical approach to sizes of sets.

The upcoming subsections will be mainly based on theory covered in [9], which we refer to for further reading.

2.1.1 Properties of σ-algebras

The basic building blocks of measure theory are subsets, but in order to ensure we can perform the necessary operations on these sets we will need some structure on the sets we will measure. In probability theory for instance, if a set consists of all possible outcomes of an experiment, then a σ-algebra would contain all events that can be assigned a probability.

Definition 2.1. *A σ-algebra on a set Ω is a collection $\mathcal{A}$ of subsets of Ω with the following properties*

- $\Omega \in \mathcal{A}$.
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.
- $\{A_n\}_{n\geq 1} \subset \mathcal{A} \Rightarrow \bigcup_{n\geq 1} A_n \in \mathcal{A}$.

Here we have used A^c to denote the complement of A, i.e. all elements in Ω that are not in A, and $\{A_n\}_{n\geq 1}$ to denote a countable collection of sets. We may note that these properties also imply

- $\emptyset \in \mathcal{A}$.
- $\{A_n\}_{n\geq 1} \subset \mathcal{A} \Rightarrow \bigcap_{n\geq 1} A_n \in \mathcal{A}$.

We will call all sets $A \in \mathcal{A}$ $\mathcal{A}$-measurable, or simply measurable if there is no ambiguity.

In many cases we are interested in the smallest possible σ-algebra containing a set $E \in \Omega$, in this case we let Σ be the collection of all σ-algebras that contain E. Note that this collection is non-empty since 2^Ω, the collection of all subsets of Ω, is trivially a σ-algebra. We then have the following result.

Proposition 2.1. *The smallest σ-algebra containing E exists and is on the following form.*

$$\sigma(E) = \bigcap_{\mathcal{A} \in \Sigma} \mathcal{A}.$$

Proof. We will demonstrate each defining property separately and in order.

- Since both Ω and E is in $\mathcal{A}$ for all $\mathcal{A} \in \Sigma$ they are also contained in $\sigma(E)$.
- If $A \in E$ then $A \in \mathcal{A}$ for all $\mathcal{A} \in \Sigma$ then by the second defining property of σ-algebras A^c must be in all $\mathcal{A} \in \Sigma$ and thus in E.
- If $\{A_n\}_{n \geq 1} \subset E$ then $\{A_n\}_{n \geq 1} \subset \mathcal{A}$ for all $\mathcal{A} \in \Sigma$ then by the third defining property of σ-algebras $\bigcap_{n \geq 1} A_n$ must be in all $\mathcal{A} \in \Sigma$ and thus in E.

This proves that $\sigma(E)$ is a σ-algebra containing E, but since $\sigma(E)$ is the intersection of all such σ-algebras we must have $\sigma(E) \subset \mathcal{A}$ for all $\mathcal{A} \in \Sigma$, and conclude that $\sigma(E)$ must be the smallest σ-algebra containing E. □

The σ-algebras generated by other collections of sets can often be complicated and may contain intricate and exotic sets. In order to actually work with these generated σ-algebras we will need the monotone class theorem, which requires a few more definitions and results.

Definition 2.2. *An algebra on a set Ω is a collection $\mathcal{A}$ of subsets of Ω with the following properties*

- $\Omega \in \mathcal{A}$.
- $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$.
- $\{A_i\}_{i=1}^n \subset \mathcal{A} \Rightarrow \bigcup_{i=1}^n A_i \in \mathcal{A}$.

This is the exact same requirements as a σ-algebra, but with the requirement on countable unions weakened to finite unions.

Definition 2.3. *A monotone class $\mathcal{M}$ is a collection of subsets of Ω with the following properties*

- *If we have an increasing sequence of sets in $\mathcal{M}$, i.e. $\{A_n\}_{n \geq 1} \subset \mathcal{M}$ and $n \leq m \Rightarrow A_n \subset A_m$ we have $\bigcup_{n \geq 1} A_n \in \mathcal{M}$.*
- *If we have a decreasing sequence of sets in $\mathcal{M}$, i.e. $\{A_n\}_{n \geq 1} \subset \mathcal{M}$ and $n \leq m \Rightarrow A_m \subset A_n$ we have $\bigcap_{n \geq 1} A_n \in \mathcal{M}$.*

Another way to put it is that monotone classes are closed under monotone limits of sets.

In a nearly identical way to when we showed the existence of $\sigma(E)$ there exists a smallest monotone class containing E denoted by $\mathcal{M}(E)$. These two constructions are closely related by the monotone class theorem

Theorem 2.1 (The Monotone Class Theorem for Sets)**.** *For an algebra $\mathcal{A}$ we have that $\sigma(\mathcal{A}) = \mathcal{M}(\mathcal{A})$. A proof of this statement can be found in [9, Theorem 1.1].*

2.1.2 Measure Spaces

Now that our building blocks are in place we can begin to define the namesake of measure theory, the measure.

Definition 2.4. *A measurable space is a pair $(\Omega, \mathcal{A})$ where $\mathcal{A}$ is a σ-algebra on Ω.*

Definition 2.5. *A measure on a measurable space $(\Omega, \mathcal{A})$ is a function $\mu : \mathcal{A} \to [0, \infty]$ with the following properties.*

- $\mu(\emptyset) = 0$.
- *Countable additivity: If $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ and $n \neq m \Rightarrow A_n \cap A_m = \emptyset$, we have*

$$\mu\left(\bigcup_{n\geq 1} A_n\right) = \sum_{n\geq 1} \mu(A_n).$$

This definition implies a few more useful properties.

- *Monotonicity: Let A, B be $\mathcal{A}$-measurable sets, then $A \subset B \Rightarrow \mu(A) \leq \mu(B)$.*
- *Continuity of measures:*
  - *if $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ is an increasing collection of sets, i.e. $n \leq m \Rightarrow A_n \subset A_m$, we have*

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

  - *if $\{A_n\}_{n\geq 1} \subset \mathcal{A}$ is a decreasing collection of sets, i.e. $n \leq m \Rightarrow A_m \subset A_n$, and $\mu(A_1) < \infty$ we have*

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

If $\mu(\Omega) = 1$ we often call the measure μ a probability measure.

Definition 2.6. *A measure space $(\Omega, \mathcal{A}, \mu)$ is a measurable space equipped with a measure. In the cases where $\mu(\Omega) = 1$ we will often call the measure space $(\Omega, \mathcal{A}, \mu)$ a probability space.*

We say that a measure space is finite if $\mu(\Omega) < \infty$. If the space is not finite we have a weaker condition that often serves the same purpose, namely σ-finiteness.

Definition 2.7. *A measure μ on a measurable space $(\Omega, \mathcal{A})$ is called σ-finite if there is a sequence of measurable sets $\{A_n\}_{n\geq 1}$ such that $\cup_n A_n = \Omega$ and $\mu(A_n) < \infty$ for all n.*

If we have a measure space $(\Omega, \mathcal{A}, \mu)$ such that μ is σ-finite we will say that the measure space $(\Omega, \mathcal{A}, \mu)$ is σ-finite.

When dealing with measure spaces we can often ignore events that occur on small sets. For any property that holds except on a set of measure zero we say that it occurs almost everywhere or a.e. for short. We will specify that we are referring to a measure μ in such a context by saying the event happens μ-a.e. if there is ambiguity.

If the measure space is also a probability space we often instead say that the event occurs almost surely or a.s. instead.

We will now consider some common constructions and specific examples of measure spaces, the first of which is perhaps the simplest.

Example 2.1. *Let Ω be any space and let 2^Ω be its power set, the collection of all subsets of Ω. We can then define the counting measure $\#$ on $(\Omega, 2^\Omega)$ as*

$$\#(A) = \{\textit{The number of elements in } A\}$$

For any $A \subset \Omega$. Note that this measure will never be σ-finite if Ω is not countable

To properly justify some of constructions in the next examples we will need to use the monotone class theorem for sets to guarantee uniqueness in certain cases. We will not prove the existence, but the construction of these extensions can be found in [9, Chapter 6].

Lemma 2.1 (Uniqueness of Measure Extensions)**.** *Assume we have a measurable space $(\Omega, \mathcal{G})$ such that $\mathcal{G} = \sigma(\mathcal{A})$ for some algebra $\mathcal{A}$. If we have two σ-finite measures μ, υ such that $\mu(A) = \upsilon(A)$ for all $A \in \mathcal{A}$ then $\mu = \upsilon$, i.e. $\mu(G) = \upsilon(G)$ for all $G \in \mathcal{G}$.*

Proof. Let $\mathcal{M}$ be the collection of all sets E such that $\mu(E) = \upsilon(E)$. We then have $\mathcal{A} \in \mathcal{M}$ and the following properties by continuity of measures.

If $\{E_n\}_{n\geq 1} \subset \mathcal{M}$ is a monotone increasing sequence then

$$\mu\left(\bigcup_n E_n\right) = \lim_{n\to\infty} \mu(E_n) = \lim_{n\to\infty} \upsilon(E_n) = \upsilon\left(\bigcup_n E_n\right),$$

which implies that $\cup_n E_n$ is in $\mathcal{M}$

For the next step we need to use σ-finiteness. Let $\{B_n\}_{n\geq 1}$ be a measurable partition of Ω such that $\mu(B_n)$ is finite for all n and let $\{C_n\}_{n\geq 1}$ be a partition of Ω such that $\upsilon(C_n)$ is finite for all n. We can then define $\{D_n\}$ to be the countable collection of all intersections on the form $B_i \cap C_j$ for $i, j \in \mathbb{N}$ and note that both $\mu(D_n)$ and $\upsilon(D_n)$ is finite for all n.

If $\{E_n\}_{n\geq 1}$ is a monotone decreasing sequence then we define $E_n^m = E_n \cap D_m$ for all $n, m \in \mathbb{N}$ and note that E_n^m is of finite measure. This allows us to apply continuity of measures and countable additivity to get

$$\begin{aligned}
\mu\left(\bigcap_n E_n\right) &= \mu\left(\bigcup_m \bigcap_n E_n^m\right) = \sum_m \mu\left(\bigcap_n E_n^m\right) \\
&= \sum_m \lim_{n\to\infty} \mu(E_n^m) = \sum_m \lim_{n\to\infty} \upsilon(E_n^m) \\
= \sum_m \upsilon\left(\bigcap_n E_n^m\right) &= \upsilon\left(\bigcup_m \bigcap_n E_n^m\right) = \upsilon\left(\bigcap_n E_n\right).
\end{aligned}$$

This means that $\cap_n E_n$ is in Ω which proves that $\mathcal{M}$ is a monotone class. Since $\mathcal{M}$ also contains $\mathcal{A}$ it must therefore contain the monotone class generated by $\mathcal{A}$. This means that $\mathcal{M}$ must contain the σ-algebra generated by $\mathcal{A}$ which, by assumption, equals $\mathcal{G}$. In other words $\mu(G) = \upsilon(G)$ for all $G \in \mathcal{G}$. □

This lemma allows us to consider many new constructions, the first of which is known as the completion of a σ-algebra. Whenever a set N is contained within a measurable set A such that $\mu(A) = 0$ we will call N a null set with respect to μ, or simply a null set if there is no ambiguity. A null set N is not necessarily measurable, but intuitively, we should be able to assign N a measure of zero.

Definition 2.8. *Let $(\Omega, \mathcal{A}, \mu)$ be a measure space and denote the collection of all null sets $\mathcal{N}$. We then define the completion of $\mathcal{A}$ with respect to μ, usually referred to as simply the completion of $\mathcal{A}$, as*

$$\sigma\left(\mathcal{N} \cup \mathcal{A}\right).$$

We could define μ on $\mathcal{N} \cup \mathcal{A}$ by $\mu(N) = 0$ for all $N \in \mathcal{N}$. Then, by Lemma 2.1, we can uniquely extend μ to the completion of $\mathcal{A}$.

If $\mathcal{A}$ already contains $\mathcal{N}$ we will say that $(\Omega, \mathcal{A}, \mu)$ is complete.

Example 2.2. *Once we have two σ-finite measure spaces $(\Gamma, \mathcal{G}, \mu)$ and $(\Lambda, \mathcal{H}, \upsilon)$ we can define the product measure space $(\Gamma \times \Lambda, \mathcal{G} \otimes \mathcal{H}, \mu \times \upsilon)$. The construction defines $\mathcal{G} \otimes \mathcal{H}$, the product σ-algebra on $\Gamma \times \Lambda$, as the completion of the σ-algebra generated by all measurable squares, i.e. sets on the form $E \times A$ such that $E \in \mathcal{G}, A \in \mathcal{H}$. The so-called product measure $\mu \times \upsilon$ is then defined on these measurable rectangles by $(\mu \times \upsilon)(E \times A) = \mu(E)\upsilon(A)$. Note that the collection of all finite disjoint unions of measurable rectangles forms an algebra which means that by extending $\mu \times \upsilon$ by finite additivity to that algebra we can apply our uniqueness result to ensure the measure has a unique extension to all sets in $\mathcal{G} \otimes \mathcal{H}$.*

Example 2.3. *Sometimes when considering a measure space $(\Omega, \mathcal{A}, \mu)$ we would like to work with only a measurable subset $\Gamma \subset \Omega$. In this case we can construct a σ-algebra on Γ by defining $\mathcal{H} = \{E \cap \Gamma : E \in \mathcal{A}\}$. By restricting μ to $\mathcal{H}$ we get a new measure space $(\Gamma, \mathcal{H}, \mu_{\mathcal{H}})$. We will refer to this construction as a restricted measure space.*

Example 2.4. *It is also worth mentioning the canonical measure space on the real line $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ where $\mathcal{B}(\mathbb{R})$ is the completion of $\sigma(\{U : U \text{ open in } \mathbb{R}\})$. This could be equivalently constructed by taking the completion of the σ-algebra generated by the algebra of finite disjoint unions of closed and bounded intervals, this means that the topology of $\mathbb{R}$ would be contained in $\mathcal{B}(\mathbb{R})$. We then define λ by $\lambda([a, b]) = b - a$ and then extend it uniquely to all of $\mathcal{B}(\mathbb{R})$. In other words, this measure generalizes the concept of length on $\mathbb{R}$. We will refer to this measure as the Lebesgue measure on the real line. For a more detailed construction we refer to [9].*

By Example 2.2 we can extend this to $\mathbb{R}^n$ for any n, in this case the measure will be an extension of n-dimensional volume.

Similarly, by Example 2.3, we can consider the restricted Borel σ-algebra on some measurable subset of $\mathbb{R}$, usually on the form $[0, T]$ for some $T > 0$.

We will refer to these σ-algebras as $\mathcal{B}(\mathbb{R}^n)$ or $\mathcal{B}([0, T])$ respectively, or simply $\mathcal{B}$ if there is no ambiguity.

2.1.3 Measurable Functions

We will usually deal with functions to analyse in the context of measure theory, however there may exist functions that do not play well with our measurable sets. We will therefore define the following class of functions.

Definition 2.9. *For a measure space $(\Omega, \mathcal{A}, \mu)$ a function $X : \Omega \to \mathbb{R}^n$ is called measurable if for all $B \in \mathcal{B}$ we have that $X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\}$ is $\mathcal{A}$-measurable.*

If μ is a probability measure we will call X a random variable, or an n-dimensional random variable if the dimension is important.

This definition can be extended to functions between any two measurable spaces, but this extension will not be necessary in this text.

It is also worth noting that measurability of functions are preserved under many operations, such as those described below.

Theorem 2.2. *Consider the measure space $(\Omega, \mathcal{A}, \mu)$. Let f, g be $\mathcal{A}$-measurable functions, let $a \in \mathbb{R}$ and let $\{h_n\}_{n\geq 0}$ be a sequence of $\mathcal{A}$-measurable functions*

- *The collection of measurable functions form an algebra i.e. the following conditions hold.*
  - *$f + g$ is $\mathcal{A}$-measurable.*
  - *af is $\mathcal{A}$-measurable.*
  - *$f \cdot g$ is $\mathcal{A}$-measurable.*
- *If $(\Omega, \mathcal{A}, \mu)$ is complete and some function k is such that $k = f$ μ-a.e. then k is $\mathcal{A}$-measurable.*
- *The functions $\max(f, g)$ and $\min(f, g)$ are $\mathcal{A}$-measurable.*
- *If it exists, the pointwise limit $\lim_{n\to\infty} h_n$ is $\mathcal{A}$-measurable.*
- *If $X : \mathbb{R}^n \to \mathbb{R}^m$ and $Y : \Omega \to \mathbb{R}^n$ are measurable functions such that the image of Y is contained in the range of X then their composition, $X(Y)$, is measurable as function $X(Y) : \Omega \to \mathbb{R}^m$.*

Similarly to working with σ-algebras we can struggle to prove things for large classes of functions. We will mention several results that will help us in later sections, but we will also occasionally have use for the function version of the monotone class theorem. This theorem allows us, under certain conditions, to extend properties of indicator functions to larger classes of functions. Note that the indicator function $\mathbb{1}_E$ of a set E is defined by $\mathbb{1}_E(\omega) = 1$ if $\omega \in E$ and 0 otherwise. This result is a slight simplification of [3, Theorem 6.3].

Theorem 2.3 (Monotone Class Theorem for Functions)**.** *Assume that Γ is any collection of subsets of a set Ω such that Γ is closed under finite intersections. Let $\mathcal{H}$ be any collection of real valued functions on Ω satisfying the following conditions.*

1. *$\mathcal{H}$ is a vector space.*
2. *For all $G \in \Gamma$ the indicator function $\mathbb{1}_G$ is in $\mathcal{H}$.*

3. *Let $\{f_n\}_{n\geq 1} \subset \mathcal{H}$ be a monotone increasing sequence of functions. If $f_n \to f$ pointwise for some bounded function f then $f \in \mathcal{H}$.*

In this case we have that $\mathcal{H}$ contains all bounded $\sigma(\Gamma)$-measurable functions.

While we mostly deal with real valued function we will occasionally encounter functions on the form $f : \Omega \to \mathbb{R}^n$, in dealing with these we will first clarify some standard notation. For two n-dimensional vectors $x, y \in \mathbb{R}^n$ we denote the following.

- The i'th component of x: $x^{(i)}$.
- The inner product on $\mathbb{R}^n$: $x \cdot y = \sum_{i=1}^n x^{(i)} y^{(i)}$.
- The absolute value of x: $|x| = x \cdot x = \sqrt{\sum_{i=1}^n x^{(i)2}}$.

We will use the similar notation for functions, let $f, g : \Omega \to \mathbb{R}^n$, we then define.

- The i'th component of $f(\omega)$: $f^{(i)}(\omega) = (f(\omega))^{(i)}$.
- Dot product of f and g: $(f \cdot g)(\omega) = f(\omega)g(\omega) = \sum_{i=1}^n f^{(i)}(\omega)g^{(i)}(\omega)$.

2.2 Integration Theory

We are now ready to define integration in the context of measure theory, this will be an extension of the classical Riemann integral in the case of the Lebesgue measure on the real line, but will also apply to many new measure spaces.

The following subsections will mainly be based on theory covered in [9], which we refer to for further reading.

2.2.1 Lebesgue Integration

Once we have defined a measure space $(\Omega, \mathcal{A}, \mu)$ we can begin defining integrals under our measure. We will start by considering our integrands to be measurable one-dimensional functions, but will extend the definition by carrying out our integration componentwise.

To construct this integral we first need a simple class of functions where we can intuitively define our integral. The most basic of which is the indicator functions defined by $\mathbb{1}_E(\omega) = 1$ if and only if $\omega \in E$ and 0 otherwise. We then define, for all $E \in \mathcal{A}$,

$$\int_\Omega \mathbb{1}_E(\omega) d\mu(\omega) = \mu(E).$$

We will often suppress the notation of some or all of the integrating variables, in this case ω. With this convention our definition is written as

$$\int_\Omega \mathbb{1}_E d\mu = \mu(E).$$

We then extend this by linearity to simple functions, which are functions on the form $\sum_{i=1}^n a_n \mathbb{1}_{E_n}$ where the E_n's are disjoint, i.e. $n \neq m \Rightarrow E_n \cap E_m = \emptyset$. For this class we define

$$\int_\Omega \sum_{i=1}^{n} a_n \mathbb{1}_{E_n} d\mu = \sum_{i=1}^{n} a_n \int_\Omega \mathbb{1}_{E_n} d\mu = \sum_{i=1}^{n} a_n \mu(E_n).$$

It turns out that for all nonnegative measurable functions f there exists an increasing sequence $\{s_n\}_{n\geq 1}$ of nonnegative simple functions such that $s_n \to f$ pointwise. With this we can extend our definition of the integral to any nonnegative measurable function by the following method. Since there exists simple functions approximating f the intuitive argument is that the integrals of these simple functions should approximate the integral of f. We state this mathematically by the following definition.

$$\int_\Omega f d\mu = \sup_s \int_\Omega s d\mu,$$

where the supremum is taken over all simple functions s such that $s \leq f$, we can also easily define

$$\int_E f d\mu = \int_\Omega f \mathbb{1}_E d\mu.$$

Now that the integral is defined for nonnegative functions we have some useful properties.

Proposition 2.2. *Let f and g be nonnegative measurable functions, E be a measurable set and let $a, b \in \mathbb{R}_+$. We then have the following.*

- *Linearity:*
$$\int_E (af + bg) d\mu = a \int_E f d\mu + b \int_E g d\mu.$$
- *Monotonicity: if $f \leq g$ on E then $\int_E f d\mu \leq \int_E g d\mu$.*
- *If $\mu(E) = 0$ then for any measurable function h we have $\int_E h d\mu = 0$.*
- *When A and B are measurable sets with $A \cap B = \emptyset$ we also have*
$$\int_{A\cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$

When we wish to apply these definitions we will often run into functions that are measurable, but with an infinite integral. This is still permitted within the definition, but can be impractical to work with. As such we usually say that a function f is Lebesgue integrable (or just integrable) if the integral of f is finite. We also say that a function is integrable over E whenever $f\mathbb{1}_E$ is integrable.

The final step is to include measurable functions with real, not just nonnegative, values. By simply using that $f = f^+ - f^-$ where $f^+(x) = max(0, f(x))$ and $f^-(x) = min(0, f(x))$, and that both f^+ and f^- are nonnegative, we get the following. Note that this definition does not make sense if f^+ and f^- are both non-integrable.

$$\int_\Omega f d\mu = \int_\Omega f^+ d\mu - \int_\Omega f^- d\mu$$

Now that we have introduced potentially negative functions we will say that a measurable function f is integrable if $|f|$ is integrable. Note that for $f = f^+ - f^-$ we have $|f| = f^+ + f^-$ and integrability of $|f|$ is therefore equivalent to the integrability of both f^+ and f^-.

In the case where μ is a probability measure we will often write $\int_\Omega X d\mu$ as $E_\mu[X]$ or simply $E[X]$ if there is no ambiguity.

Now that we have finally extended our definition to real functions we can restate our properties, note that we now require integrability in order to ensure our definitions hold.

Theorem 2.4 (Properties of the Lebesgue integral)**.** *Let f and g be integrable over E, and let $a, b \in \mathbb{R}$. We then have the following.*

- *Linearity: $af + bg$ is integrable over E and*
$$\int_E (af + bg)d\mu = a\int_E f d\mu + b\int_E g d\mu.$$
- *Monotonicity: if $f \leq g$ on E then $\int_E f d\mu \leq \int_E g d\mu$.*
- *$|\int_E f d\mu| \leq \int_E |f| d\mu$.*
- *When A and B are measurable sets with $A \cap B = \emptyset$ we also have*
$$\int_{A\cup B} f d\mu = \int_A f d\mu + \int_B f d\mu.$$
- *If $\mu(E) = 0$ then $\int_E f d\mu = 0$.*

Sometimes the condition of integrability is insufficient for certain applications. In these cases we will need an extension of integrability in the form of $\mathcal{L}^p$-spaces.

Definition 2.10. *For any measure space $(\Omega, \mathcal{A}, \mu)$ and $p > 0$ we define*
$$\mathcal{L}^p(\Omega, \mathcal{A}, \mu) = \{X : X \ \mathcal{A}\text{-measurable and } \int_\Omega |X|^p d\mu < \infty\}.$$
We will also denote the value $(\int_\Omega |X|^p d\mu)^{\frac{1}{p}}$ as $\|X\|_p$. Furthermore, if there is no ambiguity, $\mathcal{L}^p(\Omega, \mathcal{A}, \mu)$ will be denoted by $\mathcal{L}^p(\mu)$ or simply $\mathcal{L}^p$.

Note that for $p = 1$ the definition coincides with our previous definition of being integrable, we will similarly say that X is $\mathcal{L}^p$-integrable if $X \in \mathcal{L}^p$. Furthermore since $\|X\|_p$ is by definition finite on $\mathcal{L}^p$ we can use it to measure the size of our functions. In fact $\|X\|_p$ constitutes a pseudo-norm on $\mathcal{L}^p$. If we identify functions with all functions that are a.e. equal then $\|X\|_p$ will be a proper complete norm on this new space henceforth denoted by L^p, making them Banach spaces. In practice we will work with $\mathcal{L}^p$ as if they were Banach spaces and keep in mind that convergence results are only valid almost everywhere.

Special attention will be given to the L^2-spaces as the function $\langle X, Y\rangle = \int_\Omega XY d\mu$ forms an inner product, making L^2 a Hilbert space.

Lastly, we will say that an n-dimensional function f is $\mathcal{L}^p$ integrable if $f^{(i)} \in \mathcal{L}^p$ for all $1 \leq i \leq n$.

2.2.2 Multidimensional Integration

Since we can extended measures to several dimensions by Example 2.2 we can define our integral on the product space in the same we way define it on any other measure space. This however raises questions of measurability and whether we can interchange the order of integration, these are answered by the following theorems. For a proof of these results we refer to [9, Theorem 6.6,Theorem 6.7].

Theorem 2.5 (Tonelli's Theorem)**.** *Let $(\Gamma, \mathcal{G}, \mu)$ and $(\Lambda, \mathcal{H}, \upsilon)$ be σ-finite measure spaces. Then the following hold for a nonnegative extended real-valued $\mathcal{G} \otimes \mathcal{H}$-measurable function $f(x, y)$. For clarity we will have $x \in \Gamma, y \in \Lambda$.*

1. *The function $x \mapsto f(x, y)$ is $\mathcal{G}$-measurable for all $y \in \Lambda$.*
2. *The function $y \mapsto f(x, y)$ is $\mathcal{H}$-measurable for all $x \in \Gamma$.*
3. *The function $x \mapsto \int_E f(x, y) d\upsilon(y)$ is $\mathcal{G}$-measurable for all $y \in \Lambda, E \in \mathcal{H}$.*
4. *The function $y \mapsto \int_E f(x, y) d\mu(x)$ is $\mathcal{H}$-measurable for all $x \in \Gamma, E \in \mathcal{G}$.*
5. *The following equalities hold.*

$$\begin{aligned}
&\int_{\Gamma \times \Lambda} f(x, y) d(\mu \times \upsilon)(x, y) \\
&= \int_{\Gamma} \left(\int_{\Lambda} f(x, y) d\upsilon(y) \right) d\mu(x) \\
&= \int_{\Lambda} \left(\int_{\Gamma} f(x, y) d\mu(x) \right) d\upsilon(y).
\end{aligned}$$

This result holds generally for nonnegative functions, but we can extend it to real functions at the cost of some light restrictions.

Theorem 2.6 (Fubini's Theorem)**.** *Let $(\Gamma, \mathcal{G}, \mu)$ and $(\Lambda, \mathcal{H}, \upsilon)$ be σ-finite measure spaces. Let f be a real valued $\mathcal{G} \times \mathcal{H}$-measurable function such that for the nonnegative function $|f|$ we have that any of the integrals in (5) from the Tonelli theorem is finite. The following then holds. For clarity we will again have $x \in \Gamma, y \in \Lambda$.*

1. *The function $x \mapsto f(x, y)$ is in $\mathcal{L}^1(\mu)$ for υ-almost all $y \in \Lambda$.*
2. *The function $y \mapsto f(x, y)$ is in $\mathcal{L}^1(\upsilon)$ for μ-almost all $x \in \Gamma$.*
3. *The function $x \mapsto \int_E f(x, y) d\upsilon(y)$ is defined and in $\mathcal{L}^1(\mu)$ for υ-almost all $y \in \Lambda$.*
4. *The function $y \mapsto \int_E f(x, y) d\mu(x)$ is defined and in $\mathcal{L}^1(\upsilon)$ for μ-almost all $x \in \Gamma$.*
5. *The following equalities hold.*

$$\begin{aligned}
&\int_{\Gamma \times \Lambda} f(x, y) d(\mu \times \upsilon)(x, y) \\
&= \int_{\Gamma} \left(\int_{\Lambda} f(x, y) d\upsilon(y) \right) d\mu(x) \\
&= \int_{\Lambda} \left(\int_{\Gamma} f(x, y) d\mu(x) \right) d\upsilon(y).
\end{aligned}$$

2.2.3 Limit Theorems

Now that our definitions are finished we will cover the main limit theorems for the Lebesgue integral. These are the monotone and dominated convergence theorem. These are the main theorems for pulling limits inside and outside the Lebesgue integral. Proofs for these results can be found in [9, Theorem 5.6,Theorem 5.9].

In our original construction of the nonnegative integral we used a monotone increasing sequence of simple functions to approximate measurable functions and used that to justify the definition of our integral. This idea leads to the first limit theorem that extends this method to approximating integrals by monotone sequences of general nonnegative functions.

Theorem 2.7 (MCT, The Monotone Convergence Theorem)**.** *If $\{f_n\}_{n\geq 1}$ is a monotone increasing sequence of nonnegative functions converging pointwise to a function f we have for any measurable set E that*

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

Furthermore, if $f \in \mathcal{L}^1$ then

$$\lim_{n\to\infty} \|f - f_n\|_1 = \lim_{n\to\infty}\int_\Omega |f - f_n| d\mu = 0.$$

The second result is the dominated convergence theorem and covers more than just nonnegative functions.

Theorem 2.8 (DCT, The Dominated Convergence Theorem)**.** *Let $\{f_n\}_{n\geq 1}$ be a sequence of measurable functions converging pointwise to a function f. If there exists some integrable g such that $|f_n| \leq g$ for all n then we have that*

$$\lim_{n\to\infty} \|f - f_n\|_1 = \lim_{n\to\infty}\int_\Omega |f - f_n| d\mu = 0.$$

We note that this implies for any measurable set E that we again have

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

The monotone convergence theorem in particular implies the following well known corollary, which provides both a numeric method for calculating Lebesgue integrals as well as providing a connection to the classical Riemann integral. We first recall that λ denotes the canonical Lebesgue measure on the real line

Theorem 2.9. *Let f be a continuous function and let $a < b \in \mathbb{R}$. We then pick any sequence of partitions $\pi_n = \{t_i^{(n)}\}_{i=0}^n$ with $t_i^{(n)} < t_{i+1}^{(n)}$ for all i, n, as well as $t_0^{(n)} = a$ and $t_n^{(n)} = b$. Further assume that $mesh(\pi_n) = \max_i\{|t_{i+1}^{(n)} - t_i^{(n)}|\} \to 0$ as $n \to \infty$. We then have*

$$\int_{[a,b]} f(x) d\lambda(x) = \lim_{n\to\infty} \sum_{\pi_n} f\left(t_i^{(n)}\right)\left(t_{i+1}^{(n)} - t_i^{(n)}\right).$$

Proof. First we define

$$f_n(t) = \sum_{i=0}^{n-1} f(t_i) \mathbb{1}_{\left(t_{i+1}^{(n)} - t_i^{(n)}\right]}(t),$$

and note that $f_n(t) \to f(t)$ as $n \to \infty$. We can then compute from our definition of Lebesgue integrals that

$$\int_{[a,b]} f_n(t) = \sum_{i=0}^{n-1} f\left(t_i^{(n)}\right)\left(t_{i+1}^{(n)} - t_i^{(n)}\right).$$

Since f is continuous it is bounded on $[a, b]$ by some value M, by our definition of f_n we must also have $f_n(t) \leq M$ for all t, n.

This means we can apply the dominated convergence theorem which yields the desired formula.

□

2.2.4 Distribution Functions

A common trick in classic calculus is substitution, in particular for a monotone differentiable f and $g(x) = h(f(x))$ we have the useful equality

$$\int_a^b g(x)\frac{df(x)}{dx}dx = \int_a^b g(x)df(x) = \int_a^b h(f(x))df(x) = \int_{f(a)}^{f(b)} h(u)du.$$

This idea highlights the possibility of integrating with respect to functions, but we are limited by the requirement of monotonicity and differentiability. We are therefore going to define

$$\int_E X(s)dA(s)$$

for all $E \in \mathcal{B}$ and any measurable function A with right-continuous paths and finite variation, the meaning of which will be defined later. This is often called the Riemann-Stieltjes integral and will be done by integrating with respect to certain measures on the real line. When this approach is used the integral is usually referred to as the Lebesgue-Stieltjes integral.

Note that finite variation is sufficient, but not required for this construction. In order to further extend this notion one could use another approach, e.g. Young integration.

In the upcoming sections we will work with measures on the measurable space $(\mathbb{R}, \mathcal{B})$, these will be referred to as Borel measures.

In order to approach this integral we will begin by exploring the following useful class of functions.

Definition 2.11. *If μ is a finite Borel measure we may define the distribution function of μ to be*

$$F_\mu(x) = \mu((-\infty, x])$$

This definition yields a few important properties, a proof of these can be found [9, Proposition 6.6].

Proposition 2.3. *If μ is a finite Borel measure then its distribution function F has the following properties*

1. *F is monotone non-decreasing*
2. *F is right continuous*
3. *F is bounded*
4. $\lim_{x\to-\infty} F(x) = 0$

It turns out that there is a one-to-one correspondence between functions satisfying (1-4), referred to as distribution functions, and finite Borel measures. For a proof of this statement we refer to [9, Theorem 6.3].

Theorem 2.10. *If a real-valued function F satisfies the properties in the previous proposition there exists a unique finite Borel measure with F as its distribution function.*

With this we can for any distribution distribution A with a corresponding finite Borel measure μ define the so-called Lebesgue-Stieltjes integral with respect to A. For any measurable function f and $E \in \mathcal{B}$ we define

$$\int_E f(u)dA(u) = \int_E f(u)d\mu(u).$$

Note that this implies that $\int_{(s,t]} 1dA = A(t) - A(s)$.

2.2.5 Signed Measures

In order to extend the notion of integrating with respect to functions past the nonnegative case we will need a few results regarding signed measures.

Definition 2.12. *A signed measure μ on a measurable space $(\Omega, \mathcal{A})$ is an extended real-valued function with the following properties.*

- $\mu(\emptyset) = 0$.
- *μ is countably additive, i.e. if $\{A_i\}_{i\geq 1}$ is a disjoint collection of sets in $\mathcal{A}$ then*

$$\mu\left(\bigcup_{i\geq 1} A_i\right) = \sum_{i\geq 1} \mu(A_i).$$

Note that the second requirement assumes that μ must be bounded either above or below in order for the sum to be defined.

The main way to work with signed measures is to consider two positive measures whose difference equal the original signed measure. These measures will come from the two following theorems, for proofs we refer to [9, Theorem 9.1, Theorem 9.2].

Theorem 2.11 (Hahn Decomposition Theorem)**.** *For any signed measure μ on a measurable space $(\Omega, \mathcal{A})$ there exists a set $D \in \mathcal{A}$ such that $\mu(E) \geq 0$ for all $E \subset D$ and $\mu(E) \leq 0$ for all $E \subset D^c$. We will call the pair (D, D^c) a Hahn decomposition for μ.*

Note that there may very well be several Hahn decompositions, but the difference is made up by null-sets of μ and will not matter in applications.

Theorem 2.12 (Jordan Decomposition Theorem)**.** *For any signed measure μ on a measurable space $(\Omega, \mathcal{A})$ there exists two unsigned measures, μ^+ and μ^-, with the following properties. If (D, D^c) is any Hahn decomposition for μ we have $\mu^-(E) = 0$ for all $E \in D$ and $\mu^+(E) = 0$ for all $E \in D^c$. Furthermore we have the representation $\mu = \mu^+ - \mu^-$. This representation is known as the Jordan decomposition of μ and is unique despite our choice of D.*

With the Jordan decomposition we may define the integral of a function over a set E with respect to a signed measure as

$$\int_E f(x)d\mu(x) = \int_E f(x)d\mu^+(x) - \int_E f(x)d\mu^-(x).$$

2.2.6 The Lebesgue-Stieltjes Integral

We are now ready to extend the Lebesgue-Stieltjes integral to real functions. We will start by defining the total, positive and negative variation of a function A over an interval $J = [a, b]$. Here Π is the collection of partitions $\pi = \{t_i\}_{i=0}^n$ on J with $a = t_0 < t_1 < \cdots < t_n = b$ and we wil again use $x^+ = \max(x, 0)$ and $x^- = -\min(x, 0)$.

1. $V(A, J) = \sup_{\pi\in\Pi} \widehat{V}(A, \pi)$ where $\widehat{V}(A, \pi) = \sum_{i=1}^n |A(t_i) - A(t_{i-1})|$.
2. $V^+(A, J) = \sup_{\pi\in\Pi} \widehat{V}^+(A, \pi)$ where $\widehat{V}^+(A, \pi) = \sum_{i=1}^n (A(t_i) - A(t_{i-1}))^+$.
3. $V^-(A, J) = \sup_{\pi\in\Pi} \widehat{V}^-(A, \pi)$ where $\widehat{V}^-(A, \pi) = \sum_{i=1}^n (A(t_i) - A(t_{i-1}))^-$.

If value (1) is finite on a set J we say that A is of bounded variation on J, note that, for any $x \in \mathbb{R}$, since $x^+ + x^- = |x|$ and both x^+ and x^- is nonnegative we have that (1) being finite is equivalent to (2) and (3) being finite

Mimicking the idea of the Jordan decomposition we will now show that a function of finite variation can be split in a similar manner and use this for our extension. This lemma and the following theorem are known results, but we will provide our own proof of the construction to hopefully more clearly highlight the role of variation.

Lemma 2.2. *Any function A of finite variation can be decomposed in the following ways $V(A, J) = V^+(A, J) + V^-(A, J)$ and $A(b) - A(a) = V^+(A, J) - V^-(A, J)$. For simplicity we will suppress the notation of A and J.*

Proof. We start by noting the following two facts

1. for any specific partition we have $\widehat{V}^+(\pi) - \widehat{V}^-(\pi) = A(b) - A(a)$ and $\hat{V}^+(\pi) + \widehat{V}^-(\pi) = \widehat{V}(\pi)$
2. if π' is a refinement of π we have $\widehat{V}(\pi) \leq \widehat{V}(\pi')$

Now, for any ϵ we may pick three refinements (π^t, π^+, π^-) such that

- $V - \widehat{V}(\pi^t) \leq \epsilon/3$
- $V^+ - \widehat{V}^+(\pi^+) \leq \epsilon/3$
- $V^- - \widehat{V}^-(\pi^-) \leq \epsilon/3$

By fact (2) we see that all these values must be nonnegative and that any refinement of our partitions will preserve these inequalities. We define $\pi = \pi^t \cup \pi^+ \cup \pi^-$ and note that by the triangle inequality we have

$$|V - (V^+ + V^-)| \leq 3\epsilon/3 + |\widehat{V}(\pi) - (\widehat{V}^+(\pi) + \widehat{V}^-(\pi))| = \epsilon$$
$$|A(b) - A(a) - (V^+ - V^-)| \leq 2\epsilon/3 + |A(b) - A(a) - (\widehat{V}^+(\pi) - \widehat{V}^-(\pi))| = 2\epsilon/3 < \epsilon.$$

Since ϵ was arbitrary we conclude that $V = V^+ + V^-$ and $A(b) - A(a) = V^+ - V^-$ □

This lemma allows us to define the variation processes for a function A of bounded variation on an interval $J = [a, b]$

- $\mathbb{V}(x) = V(A, [a, x])$
- $\mathbb{V}^+(x) = V^+(A, [a, x])$
- $\mathbb{V}^-(x) = V^-(A, [a, x])$

We let $\mathbb{V}(x) = \mathbb{V}^+(x) = \mathbb{V}^-(x) = 0$ whenever $x \leq a$ and set them equal to their value at b for all $x > b$. Note that these functions are bounded monotone non-decreasing functions with a limit of zero when $x \to -\infty$, we also see that if A is right-continuous so are the variation processes, which would make them distribution functions.

Theorem 2.13. *For any right-continuous function of bounded variation on an interval J there exists a unique finite signed measure μ on the measurable space $([a, b], \mathcal{B}([a, b]))$ with the property that $\mu((x, y]) = A(y) - A(x)$. Here $\mathcal{B}([a, b]) = \{B \cap [a, b], B \in \mathcal{B}\}$ is the Borel σ-algebra on J.*

Proof. We apply Theorem 2.10 to W^+ to generate a measure μ^+ and do the same to W^- to generate μ^-. We define $\mu = \mu^+ - \mu^-$ and note that

$$\mu((x, y]) = \mu^+((x, y]) - \mu^-((x, y])$$
$$= \mathbb{V}^+(y) - \mathbb{V}^+(x) - (\mathbb{V}^-(y) - \mathbb{V}^-(x))$$
$$= A(y) - A(x)$$

as desired. Uniqueness can be proved by the monotone class theorem in a similar fashion to Lemma 2.1, which would also allow us to extend μ uniquely to all of $\mathbb{R}$. However, these technicalities will be skipped. □

Note that the measures, μ^+, μ^-, we produce are the same as those who appear in the Jordan decomposition of μ.

It is common to define for any measure υ the total variation measure $|\upsilon|$ as

$$|\upsilon|(E) = \sup_{\pi \in \Pi} \sum_{i=1}^{n} |\upsilon(E_n)|$$

Where Π is any finite measurable partition of E i.e the E_n's are disjoint and $\cup E_n = E$. For our newly constructed measure μ this definition coincides with the measure generated by $\mathbb{V}$ and equals $\mu^+ + \mu^-$.

This now allows us to define the Lebesgue-Stieltjes integral with respect to any right-continuous A of bounded variation by setting for any $E \in \mathcal{B}$

$$\int_E X(u) dA(u) = \int_E X(u) d\mu(u).$$

2.2.7 Independence and Distributions

One of the main properties in the realm of probability theory is the concept of independence, which will be essential for many upcoming results. We will here work within a probability space $(\Omega, \mathcal{A}, P)$.

Definition 2.13. *We have the following definitions for independence.*

- *Two events $A, B \in \mathcal{A}$ are independent if and only if*

$$P(A)P(B) = P(A \cap B).$$

- *A finite collection of one-dimensional random variables $\{X_i\}_{i=1}^n$ is independent of another finite collection of one-dimensional random variables $\{Y_j\}_{j=1}^m$ if for any two collections $\{A_i\}_{i=0}^n, \{B_i\}_{i=0}^n$ with $A_i, B_i \in \mathcal{B}$ for all i we have*

$$\begin{aligned} &P(X_i \in A_i \text{ for all } i, Y_j \in B_j \text{ for all } j) \\ &= P(X_i \in A_i \text{ for all } i)P(Y_j \in B_j \text{ for all } j) \end{aligned}$$

  Note that, as proven in [9, Theorem 7.5], it suffices to require

$$\begin{aligned} &P(X_i \leq x_i \text{ for all } i, Y_j \leq y_j \text{ for all } j) \\ &= P(X_i \leq x_i \text{ for all } i)P(Y_j \leq y_j \text{ for all } j) \end{aligned}$$

  for all $x_1, \cdots, x_n, y_1, \cdots, y_m \in \mathbb{R}$.

  In the case where we have collections of multidimensional random variables $\{X_i\}_{i=1}^n$ and $\{Y_j\}_{j=1}^m$ we say that these collections are independent if and only if

$$\bigcup_k \{X_i^{(k)}\}_i \text{ is independent from } \bigcup_l \{Y_j^{(l)}\}_{j,l}.$$

- *A collection of random variables $\{X_i\}_{i \in \mathscr{I}}$ is independent of another collection of random variables $\{Y_j\}_{j \in \mathscr{J}}$ if for all $I \subset \mathscr{I}, J \subset \mathscr{J}$ such that I and J are finite we have that $\{X_i\}_{i \in I}$ and $\{Y_j\}_{j \in J}$ are independent.*

- *A collection of random variables $\{X_i\}_{i \in \mathscr{I}}$ is said to be pair-wise independent if X_i is independent from X_j for any $i, j \in \mathscr{I}$ such that $i \neq j$.*

- *A collection of random variables $\{X_i\}_{i \in \mathscr{I}}$ is said to be mutually independent if and only if*

$$X_j \text{ is independent of } \{X_i\}_{i \in \mathscr{I}, i \neq j} \text{ for all } j \in \mathscr{I}$$

  In the one-dimensional case it suffices to require

$$P(X_i \leq x_i \text{ for all } i \in I) = \prod_{i \in I} P(X_i \leq x_i)$$

  for all finite $I \subset \mathscr{I}$ and any $x_1, \cdots, x_n \in \mathbb{R}$.

- *We say that a random variable X is independent of a σ-algebra $\mathcal{G}$ if and only if X is independent of $\mathbb{1}_G$ for any $G \in \mathcal{G}$.*

In order to properly state the consequences of independence we will need the concept of a random variable's law.

Definition 2.14 (Law of a Random Variable)**.** *Let $X_1, \cdots, X_n$ be random variables, we define the law of these variables, denoted $\mathscr{L}_{X_1,\cdots,X_n}$, as the set function on $\mathcal{B}(\mathbb{R}^n)$ defined by*

$$\mathscr{L}_{X_1,\cdots,X_n}(B) = P((X_1, \cdots, X_n) \in B)$$

for any $B \in \mathcal{B}(\mathbb{R}^n)$. By [9, Proposition 7.5] we have that the law is a measure on $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. Lastly, if $n = 1$ then we see that the law of X_1 equals

$$\mathscr{L}_{X_1}(B) = P(X_1 \in B).$$

This definition is very useful in conjunction with the following result .

Theorem 2.14. *Let $X_1, \cdots, X_n$ be random variables, then for any $\mathcal{B}(\mathbb{R}^n)$-measurable function g we have*

$$E[g(X_1, \cdots, X_n)] = \int_{\mathbb{R}^n} g(x_1, \cdots, x_n) d\mathscr{L}_{X_1,\cdots,X_n}(x_1, \cdots, x_n) \tag{2.1}$$

whenever either side of the equation exists and is finite.

The law also allows us to define an equivalent condition for mutual independence of a finite collection of random variables, in the case where it holds we can also simplify the calculation of equation (2.1). A proof of this result can be found in [9, Theorem 7.4].

Theorem 2.15. *If $X_1, \cdots, X_n$ are random variables then they are mutually independent if and only if*

$$\mathscr{L}_{X_1,\cdots,X_n}(x_1, \cdots, x_n) = \mathscr{L}_{X_1}(x_1) \times \cdots \times \mathscr{L}_{X_n}(x_n).$$

We then see that when $X_1, \cdots, X_n$ are mutually independent we can rewrite equation (2.1) as

$$E[g(X_1, \cdots, X_n)] = \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} g(x_1, \cdots, x_n) d\mathscr{L}_{X_1}(x_1) \cdots d\mathscr{L}_{X_n}(x_n).$$

As a corollary of this we get the following famous result.

Corollary 2.1. *If X, Y are independent random variables with finite expectations then XY has finite expectation as well. Lastly, we have*

$$E[XY] = E[X]E[Y].$$

Remark 1. *In the one-dimensional case we have for a random variable X that for any measurable function g that*

$$E[g(X)] = \int_{\mathbb{R}} g(x) d\mathscr{L}_X(x).$$

If we denote the distribution function of $\mathscr{L}_X$ as F_X we could alternatively write this as

$$E[g(X)] = \int_{\mathbb{R}} g(x) dF_X(x).$$

Furthermore, if F_X is differentiable with derivative f_X we have from classical calculus that

$$E[g(X)] = \int_{\mathbb{R}} g(x) \frac{dF_X(x)}{dx} dx = \int_{\mathbb{R}} g(x) f_X(x) dx.$$

The function f_X is called the probability density function and is widely used in the field of statistics. This notion could easily be extended to several dimensions, but would require distribution functions for multidimensional random variables which are not needed in this text and therefore omitted. Lastly we could extend this notion to a discontinuous f_X whenever $\mathscr{L}_X$ is absolutely continuous, a property which will be defined in the upcoming chapter.

2.3 Stochastic Analysis

In this section we will cover the study of stochastic processes, an essential tool for modelling many aspects of both finance and insurance.

The following subsections will be based on theory covered in [13],[3], as well as [9] and [2]. We refer to these sources for further reading.

2.3.1 Conditional Expectations

In the upcoming sections we will occasionally have several measures on the same measure space, in this case we often want some way of relating these measures.

Definition 2.15. *Let μ and υ be measures on the measurable space $(\Omega, \mathcal{A})$. If we have that $\upsilon(E) = 0$ whenever $\mu(E) = 0$ we say that υ is absolutely continuous with respect to μ, denoted as $\upsilon \ll \mu$.*

Starting with one measure μ it is possible to define a host of other measures by defining $\upsilon(E) = \int_E f d\mu$ for all E using any nonnegative measurable extended real-valued function f. Note that if $\mu(E) = 0$ we must have $\upsilon(E) = \int_E f d\mu = 0$, implying that $\upsilon \ll \mu$. Furthermore, it turns out that absolute continuity is also a sufficient condition on a measure in order for it to be on the form $\upsilon(E) = \int_E f d\mu$.

Theorem 2.16 (The Radon-Nikodym Theorem)**.** *If we have two σ-finite measures μ and υ with $\upsilon \ll \mu$ there exists a nonnegative measurable extended real valued function, denoted by $\frac{d\upsilon}{d\mu}$, such that $\upsilon(E) = \int_E \frac{d\upsilon}{d\mu} d\mu$ for any $E \in \mathcal{A}$. Furthermore $\frac{d\upsilon}{d\mu}$ is unique μ-almost everywhere. Lastly, we will refer to $\frac{d\upsilon}{d\mu}$ as the Radon-Nikodym derivative.*

A common occurrence in stochastic analysis is updating distributions of random variables given new information, this is usually done through conditional expectations.

We will now present the two main ways of considering conditional expectations.

Definition 2.16 (Conditional Probability and Expectation with respect to an Event)**.** *For a random variable X and any two events $E, A \in \mathcal{A}$ with $P(E) > 0$ we define*

$$P(A|E) = P(A \cap E)/P(E),$$
$$E[X|E] = E[X\mathbb{1}_E]/P(E).$$

We may note that $P(\cdot|E)$ constitutes a measure on the restricted probability space $(E, \{E \cap C\}, C \in \mathcal{A})$ as explained in Example 2.3. We further have that the conditional probability and expectation are just the canonical probability and expectation, respectively, in the restricted probability space.

This definition allows us to formulate a common result of basic statistics.

Theorem 2.17 (Law of Total Probability)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $\{E_i\}_{i\in I}$ be a countable or finite collection of disjoint sets such that $\bigcup_{i\in I} E_i = \Omega$ and $P(E_i) > 0$ for all $i \in I$. We then have the following for any events $A, B \in \mathcal{A}$ such that $P(B) > 0$.*

1. $$P(A) = \sum_{i\in I} P(A|E_i)P(E_i)$$

2. $$P(A|B) = \sum_{i\in I} P(A|B, E_i)P(E_i|B)$$

Proof. We note that $\{A \cup E_i\}_{i\in I}$ is a disjoint partition of A which means that (1) follows from the countable additivity of P along with $P(A \cup E_i) = P(A|E_i)P(E_i)$.

Repeating this argument for the measure space restricted to B yields (2). □

The other main way of conditioning on information is conditioning with respect to a collection of information instead of a single set.

Definition 2.17 (Conditional Expectation with respect to a σ-algebra)**.** *Let $(\Omega, \mathcal{A}, P)$ be a probability space, and let $X : \Omega \to \mathbb{R}$ be $\mathcal{L}^1(\Omega, \mathcal{A}, P)$-integrable. The conditional expectation of X with respect to a σ-algebra $\mathcal{G} \subset \mathcal{A}$, denoted $E[X|\mathcal{G}]$, is a $\mathcal{G}$-measurable function such that for all $G \in \mathcal{G}$ we have*

$$\int_G X dP = \int_G E[X|\mathcal{G}]dP.$$

Example 2.5. *If $\mathcal{G}$ is the σ-algebra generated by a countable partition $\{E_i\}_{i\in I}$ of Ω such that $P(E_i) > 0$ for all i we have for any $X \in \mathcal{L}^1$ that*

$$E[X|\mathcal{G}] = \sum_{i\in I} \mathbb{1}_{E_i} E[X|E_i].$$

Proof. Any $\mathcal{G}$-measurable function must be constant on E_i for all i, we also have for any i that

$$\int_{E_i} E[X|\mathcal{G}]dP = \int_{E_i} X dP = E[X|E_i]P(E_i).$$

This means that $E[X|\mathcal{G}]$ must have the value $E[X|E_i]$ a.s. for all $\omega \in E_i$ for all i. □

With this definition in place we would like to guarantee existence of the conditional expectation for general σ-algebras, which can be done via the Radon-Nikodym theorem.

Proposition 2.4. *Let $(\Omega, \mathcal{A}, P)$ be a probability space and let $X \in \mathcal{L}^1(P)$. If we have a σ-algebra $\mathcal{G} \subset \mathcal{A}$ then the conditional expectation $E[X|\mathcal{G}]$ exists and is P-a.s unique.*

Proof. Without loss of generality we consider a nonnegative random variable X, the general case is achieved by noting the decomposition $X = X^+ - X^-$. We will also consider the slightly different probability space $(\Omega, \mathcal{G}, P)$ and define a measure μ by $\mu(A) = \int_A X dP$ for all $A \in \mathcal{G}$. Note that $\mu \ll P$ which means we can apply the Radon-Nikodym theorem and get a $\mathcal{G}$-measurable Radon-Nikodym derivative $\frac{d\mu}{dP}$ such that for all $E \in \mathcal{G}$

$$\int_E X dP = \mu(E) = \int_E \frac{d\mu}{dP} dP.$$

We conclude by noting that the Radon-Nikodym theorem also guarantee uniqueness P-almost surely. □

The conditional expectation could easily be defined for any measure space, and will exist under the same conditions. For our purposes, however, the definition for probability spaces will suffice.

Theorem 2.18 (Properties of Conditional Expectations)**.** *Assume the following. $(\Omega, \mathcal{A}, P)$ is a probability space, $X, Y \in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, $a \in \mathbb{R}$ and $\mathcal{G}, \mathcal{H} \subset \mathcal{A}$ are σ-algebras. We then have the following properties for the conditional expectation.*

- *Linearity: $E[aX + Y|\mathcal{G}] = aE[X|\mathcal{G}] + E[Y|\mathcal{G}]$.*
- *Tower property: if $\mathcal{H}, \mathcal{G}$ are σ-algebras with $\mathcal{G} \subset \mathcal{H}$ we have $E[E[X|\mathcal{H}]|\mathcal{G}] = E[X|\mathcal{G}]$.*
- *Expectation: $E[E[X|\mathcal{G}]] = E[X]$.*
- *If X is independent of $\mathcal{G}$ then $E[X|\mathcal{G}] = E[X]$.*
- *If Y is $\mathcal{G}$-measurable then $E[YX|\mathcal{G}] = YE[X|\mathcal{G}]$.*
- *If $X_n \to X$ in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$ then $E[X_n|\mathcal{G}] \to E[X|\mathcal{G}]$ in $\mathcal{L}^2(\Omega, \mathcal{A}, P)$.*

We also have versions of the monotone and dominated convergence theorems for conditional expectations.

Theorem 2.19 (Dominated/Monotone Convergence)**.** *We again assume that $(\Omega, \mathcal{A}, P)$ is a probability space with $\mathcal{H} \subset \mathcal{A}$ as a σ-algebra. Let $\{X_n\}_{n\geq 1}$ be a sequence of integrable random variables converging a.s. to an integrable random variable X. We also assume that any of the following hold.*

- *There exists an integrable Y such that $|X_n| \leq Y$.*
- *The sequence X_n is monotone increasing or decreasing.*

Then we have that

$$E[X|\mathcal{G}](\omega) = \lim_{n\to\infty} E[X_n|\mathcal{G}]\omega \ P\text{-a.s.}$$

2.3.2 Stochastic Processes

Some of the main objects of study in this text are functions that develop randomly over time. This concept is expressed mathematically through stochastic processes.

Definition 2.18. *For a probability space $(\Omega, \mathcal{A}, P)$ a stochastic process is a function $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}^n$ where for every $t \in \mathbb{R}_+, \omega \in \Omega$ we have that $\omega \mapsto X(\omega, t)$ is a random variable. We will often suppress the notation of ω and denote $X(\omega, t)$ as X_t. Furthermore we will also consider the function X to be the collection $\{\omega \mapsto X(\omega, t), t \in \mathbb{R}_+\}$ for purposes of e.g. independence.*

Now that we have two parameters without assuming joint $\mathcal{A}\otimes\mathcal{B}$-measurability we will need a new way of saying that two processes are similar. Note that if we were to have joint measurability this will be a slightly stronger condition than being equal $P \times dt$-almost everywhere.

Definition 2.19. *For two stochastic processes X and Y we say that Y is a modification of X if and only if $X_t = Y_t$ almost surely for all t.*

An important aspect of having our processes depend on time is the flow of information available. We will represent this flow by the following definition.

Definition 2.20. *A filtration $\mathcal{F}$ is an increasing sequence of σ-algebras $\{\mathcal{F}_t\}_{t\geq 0}$ i.e. $s < t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t$.*

Definition 2.21. *A filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ is a probability space also equipped with a filtration $\mathcal{F}$. We will make the following assumptions, which are know as the usual conditions.*

- $\mathcal{A} = \sigma(\bigcup_{t\geq 0} \mathcal{F}_t)$.
- *$\mathcal{F}_t$ is complete, moreover, $\mathcal{F}_t$ contains all null-sets.*
- *$\mathcal{F}$ is right-continuous i.e. $\mathcal{F}_t = \bigcap_{s>t} \mathcal{F}_s$.*

Note that we may often fix a finite time horizon, in this case we will have $\mathcal{A} = \mathcal{F}_T$ for some T.

We will often work with filtrations of the following form.

Definition 2.22 (Filtration Generated by a Process)**.** *For any process X we may define the filtration $\widehat{\mathcal{F}}^X$ by*

$$\widehat{\mathcal{F}}_t^X = \sigma\left(\bigcup_{s\leq t}\bigcup_{E\in\mathcal{B}} \{X_s^{-1}(E)\}\right).$$

This filtration contains all inverse images under X of measurable sets and therefore aims to contain all available information about X up to a certain time. We call this the filtration generated by X. We would like the filtrations we work with to be complete, as such we will define the filtration generated by X as

$$\mathcal{F}_t^X = \sigma(\widehat{\mathcal{F}}_t^X \cup \mathcal{N}),$$

where $\mathcal{N}$ is the collection of all null-sets on $(\Omega, \mathcal{A}, P)$.

We may also want to have a collection $\mathcal{C} = \{X^{(i)}\}_{i\in\mathscr{I}}$ of processes generate a filtration, in this case we define

$$\mathcal{H}_t^{\mathcal{C}} = \sigma\left(\bigcup_{i\in\mathscr{I}} \mathcal{G}_t^{X^{(i)}}\right).$$

Definition 2.23 (Augmented Filtrations)**.** *We are sometimes interested in forcing a filtration $\mathcal{F}$ to be complete and right-continuous. In this case we define*

$$\mathcal{G}_t = \bigcap_{s>t} (\mathcal{F}_s \cup \mathcal{N}),$$

where $\mathcal{N}$ is the collection of all null-sets, and note that $\mathcal{G}$ is automatically complete and right-continuous. We will refer to $\mathcal{G}$ as the augmentation of $\mathcal{F}$.

Once we have a filtration we will need to specify how it relates to our processes.

Definition 2.24. *Assume we have a process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}^n$, we will then define the following properties.*

- *We say that X is adapted if $\omega \mapsto X(\omega, t)$ is $\mathcal{F}_t$-measurable for any t.*
- *We say that X is progressively measurable (or just progressive) if when we restrict $X : \Omega \times [0,t] \to \mathbb{R}^n$ we have $(\omega, s) \mapsto X(\omega, s)$ $\mathcal{F}_t \otimes \mathcal{B}[0,t]$-measurable for any t.*

We will define $Prog$ to be the σ-algebra generated by all progressive processes, it can be shown that being progressive and being $Prog$-measurable is equivalent. We will also, for any $S > 0$, denote $Prog$ restricted to $\Omega \times [0, S]$ as $Prog_S$.

Note that this definition implies that any stochastic process X is adapted with respect to the filtration generated by itself, in fact, this is the smallest such filtration.

However, being adapted is often insufficient for our purposes. With this in mind we present a sufficient condition for an adapted process to be progressive.

Proposition 2.5. *If a stochastic process $X(\omega, t)$ is right- or left-continuous in t and adapted to a filtration $\mathcal{F}$ it is also progressive with respect to the same filtration.*

Proof. We start by proving this for a left-continuous X. For any interval $[0,t]$ we may pick a sequence of partitions $\pi_n = \{t_i^{(n)}\}_{i\geq 0}^n$ with $t_0^{(n)} = 0$ and $t_n^{(n)} = t$ such that $mesh(\pi_n) = \max_i(|t_{i+1}^{(n)} - t_i^{(n)}|) \to 0$ as $n \to \infty$.

We define

$$X_n(\omega, s) = \sum_{i=0}^{n-1} X(\omega, t_i^{(n)}) \mathbb{1}_{[t_i^{(n)}, t_{i+1}^{(n)})}(s)$$

and note the following.

- $(\omega, s) \mapsto X(\omega, t_i^{(n)}) \mathbb{1}_{[t_i^{(n)}, t_{i+1}^{(n)})}(s)$ is $\mathcal{F}_t \otimes \mathcal{B}[0,t]$-measurable.
- $X_n \to X$ pointwise as a left limit.

Since measurability is preserved by both sums and pointwise limits X is $\mathcal{F}_t \otimes \mathcal{B}[0,t]$-measurable for all s and thus progressive.

For a right-continuous X we define X_n by the right limit of the interval, i.e.

$$X_n(\omega, s) = \sum_{i=0}^{n-1} X(\omega, t_{i+1}^{(n)}) \mathbb{1}_{[t_i^{(n)}, t_{i+1}^{(n)})}(s).$$

We repeat the argument as before but now have that $X_n \to X$ pointwise as right limits, the result follows. □

An important class of adapted processes are the martingales and the local martingales, we will start by defining the former.

Definition 2.25. *Let X be a stochastic process adapted to a filtration $\mathcal{F}$, X is then called a martingale if $E[|X_t|] < \infty$ and*

$$E[X_t|\mathcal{F}_s] = X_s$$

for all $s \leq t$.

To properly state the definition of local martingales we need the concept of a stopping time.

Definition 2.26. *Let $\mathcal{F}$ be a filtration and let $\mathcal{A} = \sigma(\bigcup_{t\geq 0} \mathcal{F}_t)$. A stopping time T with respect to $\mathcal{F}$ is a $\mathcal{A}$-measurable random variable such that $\{T \leq t\} \in \mathcal{F}_t$ for all t.*

The intuition behind this definition is that at time t we should now whether or not we are past the stopping time, hence the event $\{T \leq t\}$ should be $\mathcal{F}_t$ measurable.

With stopping times defined we can move on to martingales.

Definition 2.27. *Let X be a stochastic process adapted to a filtration $\mathcal{F}$. We say that X is a local martingale if there exists a sequence of stopping times $\{T_n\}$ such that the following hold.*

- *for $i < j$ we have $P(T_i < T_j) = 1$.*
- *$P(\lim_{n\to\infty} T_n = \infty) = 1$.*
- *$t \mapsto X_{\min(t,T_n)}$ is a martingale for all n.*

We may note by considering the sequence of deterministic stopping times $\{T_n\}$ with $T_n = n$ that all martingales are also local martingales.

The most famous martingale is the Wiener process, a central part of defining the upcoming Itô integral.

Definition 2.28. *The Wiener process, also known as Brownian motion, is a stochastic process $W_t = W(\omega, t)$ with the following properties.*

- *$W_0 = 0$ almost surely.*
- *W has continuous paths, i.e. $t \mapsto W_t$ is continuous almost surely.*

- *W_t has independent increments. This means that for any $\tau \in \mathbb{R}_+$ the collection of random variables on the form $W_t - W_s$ with $\tau \leq s \leq t$ is independent of the collection of random variables on the form $W_u - W_v$ with $v \leq u \leq \tau$.*
- *For $s \leq t$ we have that $W_t - W_s$ is normally distributed with mean zero and variance $t - s$.*

We will also need an n-dimensional Wiener process which will be defined as $W = (W^{(1)}, \cdots W^{(n)})$ where any component $W^{(i)}$ is a one-dimensional Wiener process. We will also assume that the components are independent.

We will now prove that the Wiener process W is a martingale, but to that we will first need a small lemma. Note that while the result of this lemma is widely known, we will provide our own proof for the sake of clarity.

Lemma 2.3. *Let $\mathcal{F}$ be the σ-algebra generated by some process $Y : \Omega \times \mathbb{R}_+ \to \mathbb{R}^n$ and let $\mathcal{G}$ be the σ-algebra generated by a process $X : \Omega \times \mathbb{R}_+ \to \mathbb{R}^m$. We then have that X independent of Y is equivalent to $\mathcal{G}$ independent of $\mathcal{F}$. Note that this implies specifically that X is independent of $\mathcal{F}$.*

Proof. We start by setting some notation.

$$X^{-1}(\mathcal{B}) = \{\alpha : \alpha = X_t^{-1}(B) \text{ for some } t \in \mathbb{R}_+, B \in \mathcal{B}\}$$
$$Y^{-1}(\mathcal{B}) = \{\beta : \beta = Y_t^{-1}(B) \text{ for some } t \in \mathbb{R}_+, B \in \mathcal{B}\}.$$

Firstly, we show that $\mathcal{G}$ independent of $\mathcal{F}$ implies X independent of Y. For any finite indicator sets I, J we define the following.

$$A_X = \bigcap_{i \in I} \alpha_i, \text{ where } \alpha_i \in X^{-1}(\mathcal{B}) \subset \mathcal{G},$$

$$A_Y = \bigcap_{j \in J} \beta_j, \text{ where } \beta_j \in Y^{-1}(\mathcal{B}) \subset \mathcal{F}.$$

We note that since $A_X \subset \mathcal{F}$ and $A_Y \subset \mathcal{G}$ we have that A_X is independent of A_Y. Since I and J were arbitrary we conclude by definition that X is independent of Y.

For the converse result we will use Theorem 2.1, the monotone class theorem for sets. We assume that X is independent of Y and aim first to prove that X is independent of $\mathcal{F}$.

We then define

$$\mathcal{C} = \{F \in \mathcal{A} : F \text{ is on the form } \bigcup_{i \in I} \bigcap_{j \in J} \beta_{i,j}\}$$

where I, J finite and $\beta_{i,j}$ is in $Y^{-1}(\mathcal{B})$ for all $i \in I, j \in J$. Note that $\mathcal{C}$ by definition becomes an algebra by the following arguments.

- $\mathcal{C}$ is closed under finite unions by

$$\bigcup_{k \in K} \bigcup_{i \in I} \bigcap_{j \in J} \beta_{i,j}$$

$$= \bigcup_{I\times K} \bigcap_{j\in J} \beta_{(i,k),j}.$$

Note here that K is assumed to be finite, which means $I \times K$ is finite.

- Since $\beta_{i,j} \in Y^{-1}(\mathcal{B})$ means $\beta_{i,j}^c \in Y^{-1}(\mathcal{B})$ we have that $\mathcal{C}$ is closed under complements by

$$\begin{aligned} F^c &= \Big(\bigcup_{i\in I}\bigcap_{j\in J} \beta_{i,j}\Big)^c \\ &= \bigcap_{i\in I}\bigcup_{j\in J} \beta_{i,j}^c \\ &= \bigcup_{j\in J}\bigcap_{i\in I} \beta_{i,j}^c. \end{aligned}$$

Then, we choose an arbitrary $\alpha = \bigcap_{l\in L} \alpha_l,$ with $\alpha_l \in X^{-1}(\mathcal{B})$ and L finite. For this α we define the class

$$\mathcal{H} = \{H \in \mathcal{F} : H \text{ is independent of } \alpha\}.$$

The aim is show that $\mathcal{F} \subset \mathcal{H}$ regardless of our choice of α by way of the monotone class theorem, which would show that $X^{-1}(\mathcal{B})$ is independent of $\mathcal{F}$.

In order to apply the monotone class theorem we need $\mathcal{C} \subset \mathcal{H}$ and that $\mathcal{H}$ is a monotone class.

We start by proving that $\mathcal{C} \subset \mathcal{H}$, it suffices to show that $\bigcap_{i=1}^n \bigcup_{j\in J} \beta_{i,j} \in \mathcal{H}$ for any n, any J finite and any $\beta_{i,j} \in Y^{-1}(\mathcal{B})$ for all i, j. However, to prove this we will prove a slightly stronger condition, namely that for any n and any K, J finite we have that

$$\bigcup_{i=1}^{n} \bigcap_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k$$

is in $\mathcal{H}$. To show this we will use induction on n. Note that for $n = 0$ we have the statement trivially from our assumption of independence of X from Y. We now assume that our statement holds for n and prove it for $n + 1$ by using that $P(A \cap B) = P(A) - P(A \cap B^c)$ for any events A and B.

$$= P\left(\alpha \cap \bigcap_{i=1}^{n+1} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right)$$

$$= P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k \cap \bigcup_{j\in J_{n+1}} \beta_{n+1,j}\right)$$

$$= P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right) - P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k \cap \Big(\bigcup_{j\in J_{n+1}} \beta_{n+1,j}\Big)^c\right)$$

$$= P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right) - P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k \cap \Big(\bigcap_{j\in J_{n+1}} \beta_{n+1,j}^c\Big)\right)$$

$$= P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right) - P\left(\alpha \cap \bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k \cap \left(\bigcap_{j\in J_{n+1}} \beta^c_{n+1,j}\right)\right)$$

By applying our induction hypothesis we then get

$$= P(\alpha)P\left(\bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right) - P(\alpha)P\left(\bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k \cap \left(\bigcap_{j\in J_{n+1}} \beta^c_{n+1,j}\right)\right)$$

$$= P(\alpha)P\left(\bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right) - P(\alpha)P\left(\bigcap_{i=1}^{n} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k \cap \left(\bigcup_{j\in J_{n+1}} \beta_{n+1,j}\right)^c\right)$$

$$= P(\alpha)P\left(\bigcap_{i=1}^{n+1} \bigcup_{j\in J} \beta_{i,j} \cap \bigcap_{k\in K} \beta_k\right)$$

This proves that $\mathcal{C} \subset \mathcal{H}$ which puts us one step closer to applying the monotone class theorem. The final step is proving that $\mathcal{H}$ is a monotone class.

- Let $\{H_n\}_{n\in\mathbb{N}}$ be a monotone increasing sequence of sets in $\mathcal{H}$. We then have by continuity of measures that

$$\begin{aligned} &P\left(\alpha \cap \bigcup_n H_n\right) \\ &= \lim_{n\to\infty} P(\alpha \cap H_n) \\ &= \lim_{n\to\infty} P(\alpha)P(H_n) \\ &= P(\alpha)P\left(\bigcup_n H_n\right). \end{aligned}$$

  Which implies $\bigcup_n H_n \in \mathcal{H}$

- Let $\{H_n\}_{n\in\mathbb{N}}$ be a monotone decreasing sequence of sets in $\mathcal{H}$. We then have by continuity of measures that

$$\begin{aligned} &P\left(\alpha \cap \bigcap_n H_n\right) \\ &= \lim_{n\to\infty} P(\alpha \cap H_n) \\ &= \lim_{n\to\infty} P(\alpha)P(H_n) \\ &= P(\alpha)P\left(\bigcap_n H_n\right). \end{aligned}$$

  Which implies $\bigcap_n H_n \in \mathcal{H}$

This means that $\mathcal{H}$ is a monotone class, since $\mathcal{H}$ also contains an algebra containing $Y^{-1}(\mathcal{B})$ it must be a σ-algebra containing $\sigma(Y^{-1}(\mathcal{B}))$. Furthermore, because null sets are independent from any set, we have that all null sets of $\sigma(Y^{-1}(\mathcal{B}))$ are in $\mathcal{H}$. Finally, we note that $\mathcal{H}$ is a σ-algebra containing $\sigma(Y^{-1}(\mathcal{B}))$ and all its null sets. From this we know that $\mathcal{H}$ must contain the completion of $\sigma(Y^{-1}(\mathcal{B}))$, which is our definition of $\mathcal{F}$.

We now have $\mathcal{F} \subset \mathcal{H}$, but used no properties of α, implying that $\mathcal{F} \subset \mathcal{H}$ for any choice of α. This means that any $\alpha = \bigcap_{l \in L} \alpha_l$ would be independent of any set in $\mathcal{F}$, but since $\mathcal{F}$ is closed under intersections we have that α is independent of $\bigcap_{l' \in L'} F_{l'}$ with $F_{l'} \in \mathcal{F}$ and L' finite. This implies by definition that $X^{-1}(\mathcal{B})$ is independent of $\mathcal{F}$.

Lastly we note that if we repeated this argument with the role of $X^{-1}(\mathcal{B})$ replaced by $\mathcal{F}$ and the role of $\mathcal{F}$ replaced by $\mathcal{G}$ we would demonstrate that $\mathcal{F}$ is independent of $\mathcal{G}$.

□

Proposition 2.6. *The Wiener process W is a martingale with respect to the filtration $\mathcal{F}$ generated by W.*

Proof. Since W has independent increments we have for any $u \leq s \leq t$ that $W_t - W_s$ is independent from $u \mapsto W_u$, $u \leq s$. By Lemma 2.3 we see that $W_t - W_s$ is independent from $\mathcal{F}_s$ which yields

$$\begin{aligned} E[W_t|\mathcal{F}_s] &= E[W_t - W_s + W_s|\mathcal{F}_s] \\ &= E[W_t - W_s|\mathcal{F}_s] + E[W_s|\mathcal{F}_s] \\ &= E[W_t - W_s] + W_s \\ &= W_s \end{aligned}$$

.

□

The Wiener process also has the following useful property, see [7, Theorem 4] for a proof.

Theorem 2.20. *The filtration $\mathcal{F}^W$ generated by a Wiener process W is right-continuous*

Remark 2. *Note that this result relies on our definition of $\mathcal{F}^W$, the filtration generated by W. We assumed in Definition 2.22 that all null-sets were included in $\mathcal{F}_t$ for all t, another option was to instead simply take the completion of $\mathcal{F}_t$ as part of the definition. In this case however we would lose right-continuity with*

$$\bigcap_{n \in \mathbb{N}} \{W_{t+h} > W_t \text{ for some } h \in (0, n^{-1})\}$$

as a counterexample. This set is the complement of a null-set and is not included in $\mathcal{F}_t$ by independent increments of W. Since there is no real drawback in including null-sets this example justifies the inclusion of $\mathcal{N}$ in Definition 2.22.

2.3.3 The Itô Integral

Our construction of the Lebesgue-Stieltjes integral is limited by the constraint of bounded variation. This is problematic in the case of a Wiener process, which is famously a.s. of unbounded variation. To solve this we will need another approach, specifically the Itô integral.

During this construction we will work within the filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$. This space is assumed to be equipped with an n-dimensional Wiener process W, and we assume that W is a martingale with respect to $\mathcal{F}$. Note that We can pick a filtration larger than the one generated by W as long as our assumption of the martingale property is preserved as explained in [13, Chapter 3.3]. In any case we will make the usual assumptions of completeness and right-continuity of $\mathcal{F}$ along with $\mathcal{A}$ being the completion of $\sigma\left(\bigcup_t \mathcal{F}_t\right)$. Furthermore we will often use the following classical notation for integration whenever we are integrating over an interval on the real line.

$$\int_a^b f(t)dt = \int_{[a,b]} f(t)d\lambda(t),$$

where λ is the canonical Lebesgue measure on the real line. We will also often work with the following classes of integrands when constructing the Itô integral. Here we will not fix a time horizon and instead let $\mathscr{T}$ be our time horizon i.e. $\mathscr{T} = [0, T]$ or $\mathbb{R}_+$.

Definition 2.29.

$$\mathcal{L}^2(W) = \mathcal{L}^2(\Omega \times \mathscr{T}, Prog, P \times \lambda) = \left\{ X \in Prog : E\left[\int_{\mathscr{T}} |X(\omega, t)|^2 dt\right] < \infty \right\}.$$

We may recall the L^2-norm on this space.

$$\|X\|_2 = \left(E\left[\int_{\mathscr{T}} |X(\omega, t)|^2 dt\right]\right)^{1/2}$$

Definition 2.30. *We also define the slightly larger class*

$$\mathcal{L}(W) = \{X \in Prog : \int_0^t |X(\omega, s)|^2 ds < \infty \text{ a.s. for all } t \in \mathscr{T}\}.$$

Note here the inclusion $\mathcal{L}^2(W) \subset \mathcal{L}(W)$.

These definitions may apply to processes of any dimension, as such we will always assume that all processes in any of these spaces share the dimension of W.

Once our integrands are established the first step in our construction is to define the integral of one-dimensional processes with respect to a one-dimensional Wiener process. For notational clarity we will for the moment assume that W is one-dimensional.

We will first define the integral for elementary processes which are functions on the form

$$\phi(\omega, t) = \sum_{n \geq 0} a_n(\omega) \mathbb{1}_{[t_n, t_{n+1})}(t).$$

Where, for all n, a_n is some $\mathcal{F}_{t_n}$-measurable one-dimensional random variable. Note that this ensures all elementary processes being progressive. For these simple functions we can define the integral with respect to W for any $i \leq j \in \mathbb{N}$ as

$$\int_{t_i}^{t_j} \phi(\omega, u) dW_u = \sum_{n=i}^{j-1} a_n(\omega)(W_{t_{n+1}} - W_{t_n}).$$

If we wish to integrate over an interval $[s, t]$ for any $s \leq t \in \mathscr{T}$ we define

$$\int_s^t \phi(\omega, u) dW_u = \sum_n a_n(\omega)(W_{s_{n+1}} - W_{s_n}).$$

Here $s_n = min(max(t_n, s), t)$ to ensure we do not integrate outside $[s, t]$.

From here we can extend the definition to any $\phi \in \mathcal{L}^2$ by picking a sequence of elementary functions $\{\phi_n\}_{n\geq 1} \subset \mathcal{L}^2(W)$ converging in the L^2-norm of $\mathcal{L}^2(W)$ to ϕ and defining

$$\int_s^t \phi(u, \omega) dW_u = \lim_{n\to\infty} \int_s^t \phi_n(u, \omega) dW_u$$

where the limit of the integrals is in the $L^2(\Omega, \mathcal{A}, P)$-norm. This definition makes sense for all of $\mathcal{L}^2$ and is independent of our choice of $\{\phi_n\}$. The proofs for the construction will be omitted, but can be found in [13, Chapter 3.1].

The integral can similarly be extended to $\mathcal{L}(W)$ [13, Chapter 3.3]. In order to explain the idea of this extension we will need the concept of convergence in probability.

Definition 2.31. *We say that a sequence of random variables $\{X_n\}_{n\geq 1}$ converges to a random variable X in probability if and only if*

$$P(|X_n - X| > \epsilon) \to 0 \text{ for every } \epsilon > 0.$$

To construct our extension we use that there exists a sequence of elementary functions $\{\phi_n\}_{n\geq 1} \subset \mathcal{L}(W)$ such that $\int_{\mathscr{T}} |\phi(t) - \phi_n(t)|^2 dt \to 0$ as $n \to \infty$. It turns out that $\int_s^t \phi_n dW$ converges in probability to some random variable dependent only on ϕ. With this we define

$$\int_s^t \phi(u, \omega) dW_u = \lim_{n\to\infty} \int_s^t \phi_n(u, \omega) dW_u.$$

With the integral defined for one-dimensional processes we now define the integral of an n-dimensional stochastic process $\phi = (\phi^{(1)}, \cdots \phi^{(n)})$ with respect to an n-dimensional Wiener process. To do this we now assume that W is a general n-dimensional Wiener process denoted $W = (W^{(1)}, \cdots W^{(n)})$. With this we define

$$\int_s^t \phi_u dW_u = \sum_{i=1}^n \int_s^t \phi_u^{(i)} dW_u^{(i)}.$$

Note that this differs from our definition of the Lebesgue integral on multi-dimensional functions which simply integrates component-wise.

With our construction complete we can state some of the most important properties of the Itô integral.

Theorem 2.21 (The Itô Isometry)**.** *Let* $X_t \in \mathcal{L}^2(W)$

$$E\left[\left(\int_{\mathscr{T}} X_t dW_t\right)^2\right] = \left(E\left[\int_{\mathscr{T}} |X(\omega,t)|^2 dt\right]\right)$$

In other words, the Itô integral preserves the norm from $\mathcal{L}^2(W)$ *to* $\mathcal{L}^2(\Omega, \mathcal{A}, P)$

We also have these further properties of the Itô integral.

Theorem 2.22. *Let the processes* X, Y *be in* $\mathcal{L}(W)$ *and* $a, b \in \mathbb{R}$*. For ease of notation both here and later we will denote the process* $t \mapsto \int_0^t X_s dW_s$ *as* $(X \bullet W)_t$*, or as* $(X \bullet W)$ *to denote the whole process.*

1. *There exists a continuous modification of* $(X \bullet W)$*.*
2. $(X \bullet W)$ *is a local martingale with respect to the filtration generated by* W*, if* X *is in* $\mathcal{L}^2(W)$ *then* $(X \bullet W)_t$ *is a martingale.*
3. *The Itô integral is linear, i.e.*

$$\big((aX + bY) \bullet W\big)_t = a\,(X \bullet W)_t + b\,(Y \bullet W)_t\,.$$

4. $\int_a^b X_s dW_s + \int_b^c X_s dW_s = \int_a^c X_s dW_s$ *almost surely.*

Note that by (1),(2), Theorem 2.2 and Proposition 2.5 $(X \bullet W)$ *is progressive.*

Whenever the integrand is deterministic we get an explicit description of the integrals distribution. While this is a well known result it is often left as an exercise, as such we will present our own proof.

Theorem 2.23. *Let* f *be a deterministic and square integrable function, we then have that* $\int_0^t f(s)dW_s$ *is normally distributed with expectation* 0 *and variance* $\int_0^t f(s)^2 ds$*.*

Proof. We start by approximating f by deterministic step functions on the form

$$f_n(s) = \sum_{i=0}^{n-1} K_i \mathbb{1}_{\left[t_i^{(n)}, t_{i+1}^{(n)}\right)}(s).$$

Here we will assume a standard sequence of partitions $\pi_n = \{t_i^{(n)}\}_{i\geq 0}^n$ with $t_i^{(n)} < t_{i+1}^{(n)}$ for all i, n, as well as $t_0^{(n)} = 0$ and $t_n^{(n)} = t$. Further assume that $mesh(\pi_n) = \max_i\{|t_{i+1}^{(n)} - t_i^{(n)}|\} \to 0$ as $n \to \infty$ and that $f_n \to f$ in $\mathcal{L}^2(\mathbb{R})$-norm.

Since f_n is simple we can compute from the definition of Itô integrals that

$$\int_0^t f_n(s)dW_s = \sum_{i=0}^{n-1} K_i \left(W_{t_{i+1}^{(n)}} - W_{t_i^{(n)}}\right).$$

We note that this is a sum of independent normally distributed random variables (by the assumption that W has independent and normally distributed increments), which means it is itself normally distributed. Since, by definition, we have

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

in probability we must have that $\int_0^t f(s)dW_s$ is normally distributed as well.

For the expectation we note that since f is deterministic and square integrable we have $f \in \mathcal{L}^2(W)$. This means that $(f \bullet W)$ is a martingale which yields

$$\begin{aligned} E\left[(f \bullet W)_t\right] &= E\left[E\left[(f \bullet W)_t \,\middle|\mathcal{F}_0\right]\right] \\ &= E\left[(f \bullet W)_0\right] \\ &= 0. \end{aligned}$$

Finally, the variance is a direct consequence of Itô's isometry, Theorem 2.21. □

Another important property of the Itô integral is how it interacts with the Lebesgue integral.

Theorem 2.24 (Fubini's Theorem for Stochastic Integrals)**.** *[3, Thm 6.2] Let $X(\omega, s, t)$ be an n-dimensional $Prog_T \otimes \mathcal{B}[0,T]$-measurable stochastic process with two indices $s, t \in [0,T]$. We also require that $\sup_{s,t\leq T} |X(s,t)| < \infty$ for almost all ω. We then have the following.*

- *The process $s \mapsto \int_0^T X(s,t)dt$ is in $\mathcal{L}(W)$*
- *There exists a progressive modification $Y(t)$ of $t \mapsto \int_0^T X(s,t)dW_s$ such that $\int_0^T Y(t)^2 dt < \infty$.*
- *The order of integration can be changed, i.e.*

$$\int_0^T \left(\int_0^T X(s,t)dW_s\right) dt = \int_0^T \left(\int_0^T X(s,t)dt\right) dW_s.$$

As a corollary we can note that the function $t \mapsto \int_0^T \mathbb{1}_{[0,t]}(s)X(\omega,s,t)ds = \int_0^t X(\omega,s,t)ds$ is $Prog \otimes \mathcal{B}$ measurable. More specifically we also have $\int_0^t X(\omega,s)ds$ progressive. Both of these facts will be useful to guarantee progressiveness of processes in upcoming sections.

2.3.4 Itô Processes

Definition 2.32. *A one-dimensional Itô process is a stochastic process X on the following form.*

$$X_t = X_0 + \int_0^t a_s ds + \int_0^t b_s dW_s.$$

Here we must require $b \in \mathcal{L}(W)$ and $a : \Omega \times \mathbb{R}^+ \to \mathbb{R}$ to be progressive with $\int_0^t |a_s| ds < \infty$ for all t to ensure the integrals are both defined and that X is progressive. This is often written in a short-hand notation as

$$dX_t = a_t dt + b_t dW_t \text{ or even more compactly as } dX = adt + bdW$$

This notation will be referred to as the dynamics of X. Moreover, it can be shown that this representation is unique in the sense that if we have

$$X_t = X_0 + \int_0^t a'_s ds + \int_0^t b'_s dW_s$$

for some other processes a', b' then $a = a'$ and $b = b'$ $dP \otimes dt$-a.e.

We can define, for any Itô process X with dynamics $dX_t = a_t dt + b_t dW_t$ the following classes of integrands. Here we will assume that Y is a one-dimensional and adapted stochastic process in addition to the stated requirements.

$$\mathcal{L}^2(X) = \left\{ Y : E\left[\int_{\mathscr{T}} (Y_t a_t)^2 dt \right] < \infty \text{ and } Yb \in \mathcal{L}^2(W) \right\}$$

$$\mathcal{L}(X) = \left\{ Y : \int_0^t |Y_s a_s| ds < \infty \text{ a.s. for all } t \text{ and } Yb \in \mathcal{L}(W) \right\}$$

Note again the inclusion $\mathcal{L}^2(X) \subset \mathcal{L}(X)$. With these definitions we can define the integral of a one-dimensional process Y with respect to X when $Y \in \mathcal{L}(X)$ by setting

$$\int_0^t Y_s dX_s = \int_0^t Y_s a_s ds + \int_0^t Y_s b_s dW_s.$$

This definition can be extended to d-dimensional processes. Note here that d is not necessarily equal to n, the dimension of W. We call $X = (X^{(i)}, \cdots X^{(d)})$ a d-dimensional Itô process if all components $X^{(i)}$ are Itô processes. In this case we define

$$\mathcal{L}^2(X) = \left\{ Z : Z^{(i)} \in \mathcal{L}^2(X^{(i)}) \text{ for all } i \leq d \right\}$$

and

$$\mathcal{L}(X) = \left\{ Z : Z^{(i)} \in \mathcal{L}(X^{(i)}) \text{ for all } i \leq d \right\}.$$

With this we define the integral for any d-dimensional process $Y \in \mathcal{L}(X)$ as

$$\int_0^t Y_s dX_s = \sum_{i=1}^d \int_0^t Y_s^{(i)} dX_s^{(i)}$$

Note that these definitions coincides with our previous definition of $\mathcal{L}^2(W)$, $\mathcal{L}(W)$, and Itô integration when $X = W$.

Definition 2.33. *Let X and Y be one-dimensional Itô processes with dynamics $dX_t = a_t dt + b_t dW_t$ and $dY_t = u_t dt + v_t dW_t$. The covariation process of X and Y, denoted $[X, Y]_t$, is defined as*

$$\int_0^t b_s v_s ds.$$

It can also be equivalently defined as

$$\lim_{n \to \infty} \sum_{i=1}^n (X_{t_i} - X_{t_{i-1}})(Y_{t_i} - Y_{t_{i-1}}) \text{ in probability,}$$

defined for any partition of $[0, t]$ with a mesh tending to zero. The quadratic variation of a single process X is then defined as $[X]_t = [X, X]_t$.

As a side-note it is provable that being of bounded variation implies a quadratic variation of zero, and a non-zero quadratic variation implies infinite variation. The final part of the side-note is that if X is of bounded variation then $[X, Y]_t = 0$ for all Y and all t.

The most important example of this definition is the quadratic variation of the one-dimensional Wiener process. Since $W_t = \int_0^t 1 dW_s$ we see that W is trivially an Itô process with $[W, W]_t = t$. This property can also be used to define the Wiener process as it is the only martingale with a linear quadratic variation, see [3, Theorem 4.2].

A neat formal trick is formally treating the dynamics of the quadratic variation as a product of dynamics i.e.

$$d[X, Y] = dXdY = au(dt^2) + avdtdW + budWdt + bv(dW^2) = bvdt$$

where $dt^2 = dtdW = dWdt = 0$ and $dW_t^2 = d[W, W]_t = dt$. This trick is defined here as simply formal manipulation, but is justified by the finite variation of the function $t \mapsto t$ and is useful as an intuitive shorthand.

These definitions are useful for formulating what is probably the most important results when it comes to working with stochastic integration and stochastic differential equations. Because of this importance it will be presented as a theorem despite its name.

Theorem 2.25 (Itô's Lemma)**.** *Let X be an Itô process on the form $dX = adt + bdW$ and let $(t, x) \mapsto f(t, x) \in C^{(1,2)}([0, \infty) \times \mathbb{R})$ i.e. continuously differentiable in t and twice continuously differentiable in x. We can then construct a new process $Y_t = f(t, X_t)$ and have that Y_t is an Itô process with dynamics*

$$dY_t = \frac{\partial f}{\partial t}(t, X_t)dt + \frac{\partial f}{\partial x}(t, X_t)dX_t + \frac{1}{2}\frac{\partial^2 f}{\partial x^2}(t, X_t)d[X]_t.$$

This extends to processes of higher dimensions, let $X^{(i)}$ be an Itô process on the form $dX^{(i)} = a^{(i)}dt + b^{(i)}dW^{(i)}$ for $1 \leq i \leq n$ and let $f \in C^{(1,2)}([0, \infty) \times \mathbb{R}^n)$. We can now for the n-dimensional Itô process $X = (X^{(1)}, \cdots, X^{(n)})$ again define $Y_t = f(t, X_t)$ and get that Y is an Itô process with dynamics

$$dY_t = \frac{\partial f}{\partial t}(t, X_t)dt + \sum_i \frac{\partial f}{\partial x^{(i)}}(t, X_t)dX_t^{(i)} + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x^{(i)}\partial x^{(j)}}(t, X_t)d[X^{(i)}, X^{(j)}]_t.$$

This means that the class of Itô processes are closed under transformation by twice-differentiable functions

We immediately see the similarities to classic calculus, in particular, if the dependence on W is removed in the one-dimensional equation reduces to the fundamental theorem of calculus. The reason the corrective term $\frac{\partial^2 f}{\partial t^2}(t, X_t)d[X]_t$ appears is due to the Wiener process having non-zero second order variation.

2.3.5 Stochastic Differential Equations

We are now ready to define stochastic differential equations, our main tool for modelling stochastic processes.

Definition 2.34. *Let the processes $\mu(\omega,t,x), \sigma(\omega,t,x)$ be $Prog \otimes \mathbb{R}$-measurable functions and let $x_0 \in \mathbb{R}$. We then call X the solution of the SDE (stochastic differential equation)*

$$dX_t = \mu(\omega,t,X_t)dt + \sigma(\omega,t,X_t)dW_t \text{ and } X_0 = x_0$$

if X is an Itô process with dynamics satisfying the equation

We may ask when there exists solutions to a specific SDE, the following result shows that under certain conditions on μ and σ there will always exist solutions.

Theorem 2.26. *Let $\mu(t,x)$ and $\sigma(t,x)$ be deterministic functions satisfying Lipschitz continuity and linear growth uniformly in time, i.e.*

$$|\mu(t,x) - \mu(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq K|x-y|$$
$$|\mu(t,x)| + |\sigma(t,x)| \leq K(1+|x|),$$

for every $t \geq 0$ and all $x, y \in \mathbb{R}$. Then there exists, for every pair (t_0, x_0), an adapted solution X_t of the SDE such that

$$dX_t = \mu(t_0+t, X(t))dt + \sigma(t_0+t, X(t))dW_t$$
$$X_0 = x_0$$

Furthermore, if there exists some other X' that also solves the SDE then $X_t = X'_t$ for all t almost surely.

One notable example of a solution to a SDE, which we will have use for later, is the stochastic exponential of an Itô process X. We define

$$\mathcal{E}_t(X) = e^{X_t - X_0 - \frac{1}{2}[X]_t}.$$

This is also known as the Doléans-Dade exponential and has the following useful properties.

Lemma 2.4. *Let X be an Itô process, and $\mathcal{E}(X)$ its stochastic exponential. We then have*

- *$\mathcal{E}(X)$ is the unique solution to $dY = YdX$, $Y_0 = 1$.*
- *$\mathcal{E}(X)$ is positive almost surely.*
- *if X is a local martingale then $\mathcal{E}(X)$ is a local martingale as well.*
- $\mathcal{E}(X)\mathcal{E}(Y) = \mathcal{E}(X+Y)e^{[X,Y]}$.
- $\mathcal{E}(X)^{-1} = \mathcal{E}(-X)e^{[X]}$.

2.3.6 Markov Chains

An essential part of financial and insurance modelling is the concept of Markov processes, functions whose future is only dependent on the present and not the past. In dealing with these processes we will here assume the standard setup of a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ with $\mathcal{A}$ and $\mathcal{F}$ satisfying the *usual conditions*, i.e. completeness and right-continuity of $\mathcal{F}$ along with $\mathcal{A}$ being the completion of $\sigma\left(\bigcup_t \mathcal{F}_t\right)$.

To state the Markov property in the way it is often presented we will introduce the notation

$$P(A|\mathcal{G}) = E[\mathbb{1}_A|\mathcal{G}]$$

For any event $A \in \mathcal{A}$, and any σ-algebra $\mathcal{G} \subset \mathcal{A}$. Note that, since the indicator function is bounded, the conditional expectation will always be defined regardless of the properties of X. This notation is closely related to our definition of conditional expectations with respect to events, which will be useful in working with some of our upcoming definitions.

Definition 2.35. *A stochastic process $X : \Omega \times \mathbb{R} \to \mathbb{R}^n$ is said to be a Markov process with respect to $\mathcal{F}$ if*

$$P(X_t \in A|\mathcal{F}_s) = P(X_t \in A|\sigma(X_s))$$

for all $s \leq t$ and all $A \in \mathcal{B}(\mathbb{R}^n)$. Here $\sigma(X_s)$ is the sigma algebra generated by the inverse images of $\omega \mapsto X_s$ and the null sets. This could be equivalently defined by requiring

$$E[\phi(X)|\mathcal{F}_t] = E[\phi(X)|\sigma(X_t)]$$

for all bounded measurable functions $\phi : \mathbb{R}^n \to \mathbb{R}$.

A common example of a Markov chain is the Wiener process, but to show this we will need a short lemma.

Lemma 2.5. *Assume Γ is a σ-algebra and X, Y are two random variables such that X is Γ-measurable and Y is independent of Γ.*

Then for any bounded $\mathcal{B}(\mathbb{R}^2)$-measurable function $g : \mathbb{R}^2 \to \mathbb{R}$

$$E[g(X, Y)|\Gamma] = E[g(X, Y)|\sigma(X)] \tag{2.2}$$

Proof. For this result we provide our own proof, this will be done by applying Theorem 2.3, the monotone class theorem for functions. First, consider $\mathcal{B} \times \mathcal{B} = \{A \times E : A, E \in \mathcal{B}\}$, the set of all measurable rectangles on $\mathbb{R}^2$, along with $\mathcal{H} = \{g : \mathbb{R}^2 \to \mathbb{R}$ such that (2.2) holds$\}$. We then need to prove the following.

1. $\mathcal{H}$ is a vector space
2. For all $B \in \mathcal{B} \times \mathcal{B}$ the indicator function $\mathbb{1}_B$ is in $\mathcal{H}$
3. Let $\{f_n\}_{n\geq 1} \subset \mathcal{H}$ is a monotone increasing sequence of functions. If $f_n \to f$ pointwise for some bounded function f then $f \in \mathcal{H}$.

We will prove these in order.

1. The first property follows from linearity of conditional expectations, contained in Theorem 2.18.

2. Let $B = A \times E \in \mathcal{B} \times \mathcal{B}$ with $A, E \in \mathcal{B}$, this would imply that we have $\mathbb{1}_B(x,y) = \mathbb{1}_A(x)\mathbb{1}_E(y)$. Note that since Y is independent of Γ it is also independent of $\sigma(X) \subset \Gamma$, yielding

$$E[\mathbb{1}_E(Y)|\Gamma] = E[\mathbb{1}_E(Y)] = E[\mathbb{1}_E(Y)|\sigma(X)].$$

   This implies that

$$\begin{aligned} E[\mathbb{1}_B(X,Y)|\Gamma] &= E[\mathbb{1}_A(X)\mathbb{1}_E(Y)|\Gamma] \\ &= \mathbb{1}_A(X)E[\mathbb{1}_E(Y)|\Gamma] \\ &= \mathbb{1}_A(X)E[\mathbb{1}_E(Y)|\sigma(X)] \\ &= E[\mathbb{1}_A(X)\mathbb{1}_E(Y)|\sigma(X)] \\ &= E[\mathbb{1}_B(X,Y)|\sigma(X)]. \end{aligned}$$

   With this we conclude that $\mathbb{1}_B$ is in $\mathcal{H}$ for all $B \in \mathcal{B} \times \mathcal{B}$

3. The third and last property follows from Theorem 2.19, the monotone convergence theorem for conditional expectations.

This means that we can apply the monotone class theorem and conclude that equation (2.2) holds for any bounded $\sigma(\mathcal{B} \times \mathcal{B}) = \mathcal{B}(\mathbb{R}^2)$-measurable function. □

Example 2.6. *The Wiener process W is Markovian with respect to the filtration $\mathcal{F}^W$ generated by W.*

Proof. For simplicity we assume a one-dimensional Wiener process, furthermore we assume $s, t \in \mathbb{R}$, $s \leq t$ and that $\phi : \mathbb{R} \to \mathbb{R}$ is some bounded $\mathcal{B}$-measurable function.

We then define $X = W_s, Y = W_t - W_s$ and $g(x,y) = \phi(x+y)$ and note from Proposition 2.6 that X is W_s-measurable and that Y is independent of $\mathcal{F}_s$. By applying Lemma 2.5 we get that

$$E[g(X,Y)|\mathcal{F}_s^W] = E[g(X,Y)|\sigma(X)].$$

Filling in for X and Y we get

$$E[\phi(W_t)|\mathcal{F}_s^W] = E[\phi(W_t)|\sigma(W_s)].$$

Since ϕ was chosen to be any bounded function we conclude that W is a Markov process.

□

We will also make use of Markov processes with a finite image.

Definition 2.36. *A Markov chain will here be defined as a stochastic process $X : \Omega \times \mathbb{R}_+ \to \mathcal{S}$ where $\mathcal{S}$ is a finite space and X is Markovian. Furthermore, any Markov chain in this text will be assumed to be càdlàg (right-continuous and with left limits).*

We will refer to $\mathcal{S}$ as the state space of X.

This class of processes will be useful in insurance by modelling the state of the insured. A common choice in this regard is the simple state space $\{*, \diamond, \dagger\}$, where $*$ means healthy, $\diamond$ means disabled and $\dagger$ means dead.

In order to describe these Markov chains we will need the probability of being in some state at a certain time given the state at a prior time.

Definition 2.37. *We define the transition probabilities of a Markov chain X with state space $\mathcal{S}$ as functions on the following form.*

$$p_{i,j}(s,t) = P(X_t = j | X_s = i),$$

where we require $s \leq t \in \mathbb{R}_+$ and $i, j \in \mathcal{S}$.

To prove the most fundamental property of transition probabilities, the Chapman-Kolmogorov equation, we will first need a small lemma. Note that the following property of Markov chains is usually taken as the defining property in the case of discrete-time Markov chains.

Lemma 2.6. *Consider a finite sequence of time points $\{t_i\}_{i=1}^n$ such that $t_i \leq t_{i+1}$ for all $i < n$ with an associated sequence of sets $\{A_i\}_{i=1}^n$. If we have that $P(X_i \in A_i \text{ for all } i < n) > 0$ we get*

$$P(X_{t_n} \in A_n | X_{t_i} \in A_i \text{ for all } i < n) = P(X_{t_n} \in A_n | X_{t_{n-1}} \in A_{n-1}).$$

Proof. For this result we provide our own proof, we will here denote $X_i = X_{t_i}$ for readability. We start by noting the following two facts for any $A \in \mathcal{A}$.

1. $$P(X_n \in A | \sigma(X_i, i < n)) = P(X_n \in A | \sigma(X_{n-1})),$$

   where $\sigma(X_i, i < n)$ and $\sigma(X_{n-1})$ are the sigma algebras generated by inverse images of $\{X_i\}_{i=1}^{n-1}$ and X_{n-1}. As before we also include the null sets in both.

   This follows from the tower property of conditional expectations along with the Markov property, which yields

   $$\begin{aligned} P(X_n \in A | \sigma(X_i, i < n)) &= E[\mathbb{1}_{\{X_n \in A\}} | \sigma(X_i, i < n)] \\ &= E[E[\mathbb{1}_{\{X_n \in A\}} | \mathcal{F}_{t_{n-1}}] | \sigma(X_i, i < n)] \\ &= E[E[\mathbb{1}_{\{X_n \in A\}} | \sigma(X_{n-1})] | \sigma(X_i, i < n)] \\ &= E[\mathbb{1}_{\{X_n \in A\}} | \sigma(X_{n-1})] \\ &= P(X_n \in A | \sigma(X_{n-1})). \end{aligned}$$

2. We have the following equalities, note that we will both here and later only sum over the indices where $P(X_i = S_i \text{ for all } i < n) \neq 0$.

   $$\begin{aligned} &P(X_n \in A | \sigma(X_i, i < 1)) \\ &= \sum_{S_1 \in \mathcal{S}} \cdots \sum_{S_{n-2} \in \mathcal{S}} \sum_{S_{n-1} \in \mathcal{S}} \mathbb{1}_{[X_i = S_i, i < n]} P(X_n \in A | X_i = S_i, i < n), \end{aligned}$$

   and in particular

$$P(X_n \in A|\sigma(X_{n-1}))$$
$$= \sum_{S_{n-1}\in\mathcal{S}} \mathbb{1}_{\{X_{n-1}=S_{n-1}\}} P(X_n \in A|X_{n-1} = S_{n-1}).$$

This is a result of the fact that any $\sigma(X_i, i < 1)$-measurable function must be constant on the sets $X_i = S_i$ for all $i < n$, which are the smallest sets making up $\sigma(X_i, i < 1)$. The exact value of these functions follows from the definition of conditional expectations. Alternatively, we can note this as a consequence of Example 2.5.

Putting these facts together we get

$$\sum_{S_1\in\mathcal{S}} \cdots \sum_{S_{n-2}\in\mathcal{S}} \sum_{S_{n-1}\in\mathcal{S}} \mathbb{1}_{\{X_i=S_i, i<n\}} P(X_n \in A|X_i = S_i, i < n)$$
$$= \sum_{S_{n-1}\in\mathcal{S}} \mathbb{1}_{\{X_{n-1}=S_{n-1}\}} P(X_n \in A|X_{n-1} = S_{n-1}),$$

which implies, by multiplying both sides by $\mathbb{1}_{\{X_{n-1}=S_{n-1}\}}$ for some choice of $S_{n-1} \in \mathcal{S}$, that

$$\sum_{S_1\in\mathcal{S}} \cdots \sum_{S_{n-2}\in\mathcal{S}} \mathbb{1}_{\{X_i=S_i, i<n\}} P(X_n \in A|X_i = S_i, i < n)$$
$$= \mathbb{1}_{\{X_{n-1}=S_{n-1}\}} P(X_n \in A|X_{n-1} = S_{n-1}).$$

Furthermore, since $P(X_n \in A|X_{n-1} = S_{n-1})$ is constant on the set $\{X_{n-1} = S_{n-1}\}$, we have that

$$\sum_{S_1\in\mathcal{S}} \cdots \sum_{S_{n-2}\in\mathcal{S}} \mathbb{1}_{\{X_i=S_i, i<n\}} P(X_n \in A|X_i = S_i, i < n)$$

must be constant and equal to $P(X_n \in A|X_{n-1} = S_{n-1})$ on $\{X_{n-1} = S_{n-1}\}$. Since S_{n-1} was chosen arbitrarily the only way this can hold is if

$$P(X_n \in A|X_i = S_i, i < n) = P(X_n \in A|X_{n-1} = S_{n-1}),$$

which yields the desired result. □

We now move on to the Chapman-Kolmogorov equations.

Theorem 2.27. *For any $s, u, t \in \mathbb{R}_+$ with $s \leq u \leq t$ and any $i, j \in \mathcal{S}$ we have*

$$p_{i,j}(s,t) = \sum_{k\in\mathcal{S}} p_{i,k}(s,u)p_{k,j}(u,t)$$

Proof. By applying the law of total probability and Lemma 2.6 we get that

$$\begin{aligned} p_{i,j}(s,t) &= P(X_t = j|X_s = j) \\ &= \sum_{k\in\mathcal{S}} P(X_t = j|X_s = j, X_u = k)P(X_u = k|X_s = j) \\ &= \sum_{k\in\mathcal{S}} P(X_t = j|X_u = k)P(X_u = k|X_s = j) \\ &= \sum_{k\in\mathcal{S}} p_{i,k}(s,u)p_{k,j}(u,t) \end{aligned}$$

□

This result will be particularly useful in combination with the following definition.

Definition 2.38. *We define the following transition rates for a Markov chain X whenever the respective limits exist. We will here use the notation $\lim_{x\to 0^+}$ to mean that $x \to 0$ and $x > 0$. For any $i \in \mathcal{S}$ we then define*

$$\mu_i(t) = \lim_{h\to 0^+} \frac{1 - p_{i,i}(t, t+h)}{h}.$$

If $i, j \in \mathcal{S}$ and $i \neq j$ we define

$$\mu_{i,j}(t) = \lim_{h\to 0^+} \frac{p_{i,j}(t, t+h)}{h}.$$

By convention we also define $\mu_{i,i}(t) = -\mu_i(t)$, note that this implies that

$$\mu_{i,j}(t) = \lim_{h\to 0^+} \frac{p_{i,j}(t, t+h) - p_{i,i}(t, t)}{h} = \frac{\partial p_{i,j}(s,t)}{\partial t}\Big|_{s=t}$$

for all $i, j \in \mathcal{S}$ along with, for any $i \in \mathcal{S}$,

$$\sum_{j\in\mathcal{S}} \mu_{i,j}(s,t) = 0.$$

Definition 2.39. *A Markov chain is said to be regular if for all $i, j \in \mathcal{S}$ we have that $t \mapsto \mu_{i,j}(t)$ is defined and continuous.*

This leads us to the famous Kolmogorov equations, which are useful as they provide a way to recover transition probabilities from the transition rates. In practical applications it is generally the rates that are observable, which means the following result is needed to compute the probabilities.

Theorem 2.28 (Kolmogorov's Equations)**.** *Assume X is a regular Markov chain, we then have the following equations.*

- *The backwards Kolmogorov equation:*

$$\frac{\partial p_{i,j}(s,t)}{\partial s} = -\sum_{k\in\mathcal{S}} \mu_{i,k}(s) p_{k,j}(s,t)$$

- *The forwards Kolmogorov equation:*

$$\frac{\partial p_{i,j}(s,t)}{\partial t} = \sum_{k\in\mathcal{S}} p_{i,k}(s,t)\mu_{k,j}(t)$$

We might also be interested in knowing the chance of X staying in a certain state $i \in \mathcal{S}$ for an interval of time $[s,t] \subset \mathbb{R}_+$. We then denote

$$\bar{p}_{i,i}(s,t) = P\Big(\bigcap_{u\in[s,t]} \{X_u = i\} \Big| X_s = i \Big).$$

Note that $\bigcap_{u\in[s,t]}\{X_u = i\}$ is an uncountable intersection and thus not automatically measurable. However, by our assumption of right-continuity we may, for any $u < t$, consider a rational decreasing sequence of numbers $\{q_n\}_{n\in\mathbb{N}}$ converging to u. We then have that if $X_{q_n} = i$ for all n then $X_u = \lim_{n\to\infty} X_{q_n} = i$. This means that if $X_q = i$ for all rational numbers $q \in [s,t] \cup \mathbb{Q}$ then $X_u = i$ for all $u \in [s,t)$, which yields

$$\bigcap_{u\in[s,t]} \{X_u = i\} = \bigcap_{u\in[s,t]\cap\mathbb{Q}} \{X_u = i\} \cap \{X_t = i\}.$$

Since the right-hand side of the equation is a countable intersection it is measurable, which guarantees us that our definition is well defined.

Example 2.7. *Let us assume a simple insurance model for modelling the state of the insured. We will here consider a regular Markov chain with state-space $\mathcal{S} = \{*, \dagger\}$ where $*$ means that the insured is alive and $\dagger$ means that they are dead. Furthermore we assume that the probability of returning from the dead state is 0, which means that*

$$\bar{p}_{*,*}(s,t) = p_{*,*}(s,t) = 1 - p_{*,\dagger}(s,t).$$

By applying the forward Kolmogorov equation we get

$$\begin{aligned}&\frac{\partial p_{*,*}(s,t)}{\partial t}\\ &= p_{*,*}(s,t)\mu_{*,*}(t) + p_{*,\dagger}(s,t)\mu_{\dagger,*}(t)\\ &= p_{*,*}(s,t)\mu_{*,*}(t).\end{aligned}$$

Using that $p_{*,*}(s,s) = 1$ we have that the solution to this differential equation is

$$p_{*,*}(s,t) = \bar{p}_{*,*}(s,t) = exp\left(\int_s^t \mu_{*,*}(u)du\right).$$

The argument given generalizes to general Markov chains, yielding the following result.

Theorem 2.29. *Let X be a regular Markov chain, we then have for any $i \in \mathcal{S}$ that*

$$\bar{p}_{i,i}(s,t) = exp\left(\int_s^t \mu_{i,i}(u)du\right).$$

Note that we could equivalently write

$$\bar{p}_{i,i}(s,t) = exp\left(-\sum_{\substack{j\in\mathcal{S}\\ j\neq i}} \int_s^t \mu_{i,j}(u)du\right).$$

Chapter 3

Financial Markets and Arbitrage Theory

3.1 Financial Markets

Financial markets is an important area of study, not only in its own right, but also for the applications it offers. In particular, the financial market allows for the hedging of risk which affects pricing in other fields, such as insurance. We will cover the specific effects on pricing in the next section on arbitrage, but first we need a few basic definitions and results.

For these definitions we will be working within a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$. We assume the filtration $\mathcal{F}$ is right-continuous and complete and that $\mathcal{A}$ is the completion of $\sigma\left(\bigcup_{t\in\mathbb{R}_+} \mathcal{F}_t\right)$. Lastly we assume the existence of a Wiener process W that is a martingale with respect to $\mathcal{F}$.

Note that in this chapter we will be covering theory mainly based on what is covered in [3, Chapter 4], which we refer to for further reading.

Definition 3.1. *A zero-coupon bond is a financial agreement that the holder of this bond is guaranteed one unit of currency at a time of maturity T. The value of a zero-coupon bond at time t, with maturity T, also referred to as a T-bond is often denoted $P(t,T)$. We will make the following assumptions.*

1. *There exists an unrestrained market for T-bonds for any maturity $T > 0$*
2. *$P(T,T) = 1$*
3. *$P(t,T)$ is differentiable in T*

Remark 3. *Property (1) is never satisfied in reality. T-bonds are usually only traded for specific maturities, e.g. monthly, and selling large amounts of these T-bonds is restricted due to the risk of the issuer defaulting. This same risk means Property (2) could fail as well if the issuer is again at risk of defaulting. Despite the lack of perfect realism these conditions will be useful for developing our mathematical framework. It is hard to comment on the realism of property (3) as $P(t,T)$ is only given for discrete values in reality, so we consider it a purely technical condition.*

Definition 3.2. *We define the forward rate $f(t,T)$ and the short-rate r_t as*

$$f(t,T) = -\frac{\partial P(t,T)}{\partial T},$$
$$r_t = f(t,t).$$

Note that by the definition of $f(t,T)$, along with our assumption that $P(T,T) = 1$, we have that

$$P(t,T) = e^{-\int_t^T f(t,s)ds}.$$

We will often refer to the function $T \mapsto f(t,T)$ as the forward curve.

With this we need one final definition before properly starting on financial markets.

Definition 3.3. *We define the bank account B, often referred to as the risk-free asset by*

$$B_t = e^{\int_0^t r_s ds}.$$

With this we can move on to financial markets, in the rest of this section we will assume that we have a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ equipped with an d-dimensional Wiener process. We will further assume as we did when defining the Itô integral that W is a martingale with respect to $\mathcal{F}$. Lastly we require that $\mathcal{F}$ is right continuous, complete and that $\mathcal{A}$ is the completion of $\sigma(\bigcup_{t\geq 0} \mathcal{F}_t)$.

Definition 3.4. *A financial market S is a collection of assets represented by the $n+1$-dimensional stochastic process $(S^{(0)}, \cdots, S^{(n)})$ Where $S^{(0)} = B$ is the bank account discussed earlier and $S^{(i)}$ for $i \geq 1$ is a financial asset on the form*

$$dS^{(i)} = S^{(i)} dX^{(i)}$$

where $X^{(i)}$ is some Itô process. We must therefore require $S^{(i)} \in \mathcal{L}(X^{(i)})$ and could equivalently define

$$S_t^{(i)} = S_0^{(i)} \mathcal{E}_t(X)$$

A common choice for some of the assets will be zero-coupon bonds for different time horizons, this choice will be essential in developing the HJM framework. Note that S by definition becomes a $n+1$-dimensional Itô process.

We will later go on to price claims according to the zero arbitrage principle, but before we can discuss arbitrage we need a concept of possible financial trading strategies.

A portfolio or trading strategy $\phi = (\phi^{(0)}, \cdots \phi^{(n)})$ is an $n+1$-dimensional progressive process and its value process V is defined by

$$V_t = \phi_t \cdot S_t = \sum_{i=0}^{n} \phi_t^{(i)} S_t^{(i)}.$$

Such a portfolio is called self-financing if $\phi \in \mathcal{L}(S)$ and that all change in the value process comes changes in the value of our assets. To state the latter mathematically we require $dV_t = \phi_t dS_t$.

We will often consider normalized asset values, which is done by scaling the value of our assets by certain processes known as nemeraires. A numeraire then becomes the base value of the market and is usually chosen to be U.S. dollars. There are many ways of normalizing, such as scaling both assets with the conversion rate between dollars and Norwegian kroner to get our prices in NOK. We will here consider $S^{(0)} = B$ as our numeraire. We will use calligraphic letters for the discounted price vector and value process, and with this we get

$$\mathcal{S} = S/B = (1, S^{(1)}/B, \cdots, S^{(n)}/B) = (1, \mathcal{S}^{(1)}, \cdots, \mathcal{S}^{(n)}).$$

We will use this following lemma to ensure this scaling does not interfere with our other definitions.

Lemma 3.1. *If $\phi \in \mathcal{L}(S) \cap \mathcal{L}(\mathcal{S})$ then ϕ is self-financing for S if and only it is self-financing for $\mathcal{S}$.*

We may also calculate the dynamics of $\mathcal{S}$ by an application of the two-dimensional version of Itô's lemma. This will be calculated for any component $\mathcal{S}^{(i)}, i \neq 0$.

We start by noting that $B^{-1} = e^{-\int_0^t r_s ds}$ implying $d(B^{-1}) = -rB^{-1}dt$. Applying Itô's lemma to $f(x, y) = xy$ yields

$$\begin{aligned} d\mathcal{S}^{(i)} &= d(S^{(i)}B^{-1}) = d(f(S^{(i)}, B^{-1})) \\ &= S^{(i)}d(B^{-1}) + B^{-1}dS^{(i)} \\ &= -rS^{(i)}B^{-1}dt + S^{(i)}B^{-1}(\mu^{(i)}dt + \sigma^{(i)}dW) \\ &= \mathcal{S}^{(i)}\big((\mu^{(i)} - r)dt + \sigma^{(i)}dW\big). \end{aligned} \tag{3.1}$$

With this we see that the discounted asset price $\mathcal{S}^{(i)}$ is still a stochastic exponential, but with a discounted drift.

These definitions now mimic real-life trading strategies, but we still might have arbitrage opportunities. An arbitrage portfolio is usually defined as a portfolio ϕ with a value process V satisfying

$$V(0) = 0 \ , \ V(T) \geq 0 \ , \ P[V(T) > 0] > 0 \text{ for some } T$$

This is problematic because it allows for strategies that guarantee no loss while offering a chance for profit. If this model was representative of reality everyone could simply scale this strategy in order to eventually make riskless profit. In a stable market there should not exist such opportunities. In order to avoid this we need further conditions on which strategies are admissible, along with conditions to ensure our market can be free of arbitrage.

3.2 Arbitrage Theory

In order to avoid arbitrage we need the concept of an equivalent (local) martingale measure, and in order to construct such a measure we need a few theoretical results.

Definition 3.5. *An equivalent measure Q of a measure P, denoted $Q \sim P$ is a measure s.t. $Q \ll P$ and $P \ll Q$.*

An equivalent martingale measure (EMM) is a probability measure $Q \sim P$ such that $\mathcal{S}^{(i)}$ is a martingale under Q for every i.

An equivalent local martingale measure (ELMM) is a probability measure $Q \sim P$ such that $\mathcal{S}^{(i)}$ is a local martingale under Q for all i.

Note that every EMM is also an ELMM

When such a measure exists we will often be interested to see how the dynamics of our Wiener process behaves under Q, to see this we will use Girsanov's theorem. In order to properly state this theorem we will need the technical condition of uniform integrability.

Definition 3.6. *A collection of functions $\{f_i\}_{i \in \mathscr{I}}$ on a measure space $(\Omega, \mathcal{A}, \mu)$ is called uniformly integrable if for every δ there exists an ϵ such that for every i and every $E \in \mathcal{A}$ with $\mu(E) \leq \epsilon$ we have*

$$\int_E f_i d\mu \leq \delta$$

This now allows us to properly state Girsanov's theorem. Note that there are many versions of this result, but we here present the same version as stated in [3, Theorem 4.6].

Theorem 3.1 (Girsanov's Theorem)**.** *Let $\gamma \in \mathcal{L}(W)$ be such that $\mathcal{E}(\gamma \bullet W)$ is a uniformly integrable martingale. We also require that $\mathcal{E}_T(\gamma \bullet W) > 0$ for some time horizon T and get that*

$$\frac{dQ}{dP} = \mathcal{E}_T(\gamma \bullet W) \tag{3.2}$$

defines an equivalent probability measure $Q \sim P$. We also have for any equivalent measure given by equation (3.2) that the process W^* defined by

$$W_t^* = W_t - \int_0^t \gamma_s ds$$

is a Wiener process under Q.

Lastly by exchanging T for ∞ we have the same results for an infinite time horizon. In this case we would define

$$\mathcal{E}_\infty(\gamma \bullet W) = \lim_{u \to \infty} \mathcal{E}_u(\gamma \bullet W).$$

The requirements for $\mathcal{E}(\gamma \bullet W)$ can be troublesome. Luckily there exists a sufficient, but not necessary, condition for these properties to hold.

Theorem 3.2 (Novikov's condition)**.** *[3, Thm 4.7] If*

$$E\left[e^{\frac{1}{2}\int_0^T |\gamma_t|^2 dt}\right] < \infty \tag{3.3}$$

then $\mathcal{E}(\gamma)$ is a uniformly integrable martingale on $[0, T]$ and $\mathcal{E}_T(\Gamma) > 0$.

Again, exchanging T for ∞ gives the same result for an infinite time horizon.

If an equivalent measure on the form in Girsanov's theorem, equation (3.2), that is also an E(L)MM exists we have the following Q-dynamics for $\mathcal{S}^{(i)}$,$i \neq 1$, by using $dW = dW^* + \gamma dt$.

$$\begin{aligned} d\mathcal{S}^{(i)}/\mathcal{S}^{(i)} &= (\mu^{(i)} - r)dt + \sigma^{(i)}dW \\ &= (\mu^{(i)} - r)dt + \sigma^{(i)}(dW^* + dt) \\ &= (\mu^{(i)} - r + \sigma^{(i)}\gamma)dt + \sigma^{(i)}dW^*. \end{aligned}$$

In order for $\mathcal{S}$ to be a local martingale we cannot have non-zero drift term and must require

$$-\gamma = (\mu^{(i)} - r)/\sigma^{(i)} \text{ for all } i \geq 1. \tag{3.4}$$

This is often referred to as the market price of risk. Note that if the converse holds and there exists a process γ solving (3.4) in such a way that γ also meets the conditions for Girsanov (3.2) (or Novikov (3.3)), then there exists an ELMM for our market.

With this established we can move on to finally define the class of trading strategies we will consider. Here we will use the definition of Filipović

Definition 3.7. *An admissible trading strategy ϕ is a self-financing strategy such that the discounted value process $\mathcal{V}$ is a Q-martingale for some ELMM Q.*

Another common way of defining admissible strategies is instead requiring them to be bounded from below. As noted in [3, p. 70], the next result would still hold under these conditions. This is however not needed for our purposes, and we will stick to this more manageable definition.

We are now finally ready to present the main reason why the existence of an ELMM is important

Theorem 3.3 (The First Fundamental Theorem of Asset Pricing)**.** *If there exists an ELMM the market has no admissible arbitrage portfolios*

Proof. Let $\mathcal{V}$ be the discounted value process of some admissible trading strategy ϕ such that $\mathcal{V}(0) = 0$ and $\mathcal{V}(T) \geq 0$. We then have by definition of admissibility that $\mathcal{V}$ is a martingale for some $Q \sim P$ and get

$$0 \leq E_Q[\mathcal{V}_T] = \mathcal{V}_0 = 0.$$

This means that $Q(\mathcal{V}_T > 0) = 0$, which by equivalence to P means that $P(\mathcal{V}_T > 0) = 0$. This means that ϕ cannot be an arbitrage opportunity. □

The converse is not exactly true, it is possible to have an arbitrage free market without an ELMM. There is however a slightly stronger condition than the absence of arbitrage, known as "no free lunch without risk", which is equivalent to the existence of an ELMM. Regardless we shall assume both the absence of arbitrage and then existence of an ELMM of the Girsanov form given in equation (3.2) for our market.

Now that we have established the absence of arbitrage we will move on to find a way of assigning prices to our options.

Definition 3.8. *A contingent claim due at time T or T-claim is some payoff due at time T and is defined to be any $\mathcal{F}_T$-measurable random variable X.*

Notice that if a T-claim X is equal to the payout of some portfolio V i.e. $X = V_T$ then the value of X at time t given all information up to t should be V_t. If this was not the case we could obviously buy the cheaper one and short the other to generate profit while paying nothing at time T. Furthermore we have that $\mathcal{V} = V/B$ and that $\mathcal{V}$ is a martingale under Q, yielding

$$\begin{aligned} B_t E[X/B_T|\mathcal{F}_t] &= B_t E[V_T/B_T|\mathcal{F}_t] \\ &= B_t E[\mathcal{V}_T|\mathcal{F}_t] = B_t \mathcal{V}_t = V_t. \end{aligned}$$

So we have another way of calculating an arbitrage free price, we will use this pricing method on what will be defined as attainable claims.

Definition 3.9. *An attainable claim is a T-claim X with $X = V_T$ for the value function of some portfolio at time T. We will refer to such a portfolio as a replicating portfolio.*

In order to work with attainability we will need some way to represent our T-claims. To do this we will use the following representation theorem.

Theorem 3.4. *Assume that our filtration $\mathcal{F}$ is the filtration generated by our Wiener process W. Then for every P-local martingale M with respect to $\mathcal{F}$ there exists some $\psi \in \mathcal{L}^2$ such that*

$$M_t = M_0 + \int_0^t \psi_s dWs.$$

This will be necessary for the second fundamental theorem of asset pricing, which we will use to define arbitrage-free prices for our claims. First, however, we have a short corollary showing that when $\mathcal{F}$ is generated by W all ELMMs are on the Girsanov form(3.2).

Corollary 3.1. *[3, Thm 4.8] Let Q be an ELMM, and let $\frac{dQ}{dP}$ be the Radon-Nikodym derivative with respect to P. Then there exists some process γ such that.*

$$\frac{dQ}{dP} = \mathcal{E}_T(\gamma \bullet W).$$

Proof. We define $D_t = E\left[\frac{dQ}{dP}|\mathcal{F}_t\right]$ and note that D_t is a martingale by the tower property of conditional expectations. By our representation theorem we then have for some ψ

$$D_t = D_0 + \int_0^t \psi dW.$$

Since $\frac{dQ}{dP}$ is positive so is D_t for all t and we get

$$dD_t = \psi_t dW_t = D_t(\psi_t/D_t)dW$$

We then define $\gamma = \psi/D$ and note that $d(\gamma \bullet W)_t = \gamma_t dW_t$ to see that

$$dD_t = D_t d(\gamma \bullet W).$$

Which shows that $D_t = \mathcal{E}_t(\gamma \bullet W)$ implying $\frac{dQ}{dP} = D_T = \mathcal{E}_T(\gamma \bullet W)$ □

This corollary implies that if $\mathcal{F}$ is generated by W then any ELMM must be of the Girsanov form given in equation (3.2). In this case we could simplify our assumption about the Girsanov form of our ELMM and just equivalently assume that some ELMM exists.

Lastly note that this result also applies in the case where our time horizon is infinite, in which case we would get

$$\frac{dQ}{dP} = \mathcal{E}_\infty (\gamma \bullet W) .$$

Now that this is established we return to pricing. We know how to price attainable claims, but we would like to know which T-claims are attainable.

Definition 3.10. *A market is complete if for all T any T-claim with bounded discounted payoff X/B_T is attainable*

We can now present conditions for when our market is complete.

Theorem 3.5 (The Second Fundamental Theorem of Asset Pricing)**.** *[3, Theorem 4.9] Assume that $\mathcal{F}$ is generated by W and that there exists an ELMM for our market. The following are then equivalent*

- *The market is complete*
- *Our ELMM is unique*
- *The market price of risk γ is unique $dP \otimes dt$-a.e.*

Furthermore, if these hold then any claim X with

$$E_Q \left[\frac{|X|}{B_T} \right] < \infty \tag{3.5}$$

is attainable.

We know from (3.4) that γ is uniquely defined by our assets which means that our market is complete as long as there exists some ELMM on the Girsanov form (3.2). We can then define the arbitrage free price of any T-claim satisfying (3.5) as

$$\pi(X, t) = B_t \mathcal{E}[X/B_T | \mathcal{F}_t].$$

Chapter 4

Interest Rate Modelling

We will now cover the class of interest models we are going to apply in this text. Like in Chapter 3 on finance we will work within a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$. We assume the filtration $\mathcal{F}$ is right-continuous and complete and that $\mathcal{A}$ is the completion of $\sigma\left(\bigcup_{t\in\mathbb{R}_+} \mathcal{F}_t\right)$. We also assume the existence of a Wiener process W that is a martingale with respect to $\mathcal{F}$ and lastly that we have a market as defined in Chapter 3 with a unique ELMM.

In this chapter we will mainly present theory covered in [3, Chapter 6,Chapter 7], which we refer to for further reading.

4.1 The HJM Framework

One of the issues with many short-rate models is calibrating them to the initial term structure (i.e. the function $T \mapsto P(0,T)$). The HJM framework (named after Heath Jarrow and Morton) was proposed in the late 80's which models the entire forward curve directly.

We make the same assumptions as stated earlier, but we will also assume a model for the forward rate $f(t,T)$ from Definition 3.2 based on two stochastic processes, α and σ. While these processes are stochastic we will as usual suppress the notation of ω and denote $\alpha(t,T) = \alpha(\omega,t,T), \sigma(t,T) = \sigma(\omega,t,T)$. The assumptions we will make on these processes are the following.

- HJM1: α and σ are $Prog \otimes \mathcal{B}$-measurable.
- HJM2: $\int_0^T \int_0^T |\alpha(s,t)| ds dt < \infty$.
- HJM3: $\sup_{s,t\leq T} |\sigma(s,t)| < \infty$ for all T and all ω.

We will also assume that we have given some integrable initial forward curve $(T \mapsto f(0,T))$. With this we can specify our model for $f(t,T)$ for any T as

$$f(t,T) = f(0,T) + \int_0^t \alpha(s,T) ds + \int_0^t \sigma(s,T) dW_s.$$

Note that HJM2 guarantees the existence of the first integral and HJM1 along with the non-stochastic Fubini theorem (Theorem 2.6) makes it progres-

sive. We also see that HJM3 implies that

$$\int_0^t |\sigma(s,T)|^2 ds \leq t \sup_{u,v\leq T} |\sigma(u,v)|^2 < \infty,$$

for all t, ω. This along with HJM1 means we have $\sigma \in \mathcal{L}(W)$ which makes the second integral defined as well. Moreover, by continuity Proposition 2.5 ensures that $t \mapsto f(t,T)$ is progressive.

We also define the short-rate.

$$r_t = f(t,t) = f(0,t) + \int_0^t \alpha(s,t)ds + \int_0^t \sigma(s,t)dW_s.$$

It is worth noting that r is no longer necessarily continuous due to the dependence on t inside the integrals. To ensure progressiveness we need the following result.

Proposition 4.1. *Assume $X(\omega,s,t)$ is a stochastic process satisfying the conditions in the stochastic Fubini theorem (Theorem 2.24), i.e. X is $Prog \otimes \mathcal{B}$-measurable and bounded jointly in s,t for all ω. Then $s \mapsto \int_0^t X(\omega,s,t)dW_s$ has a progressive modification J_t such that $\int_0^T J_t^2 dt < \infty$ almost surely.*

Since $\sigma(s,t)$ satisfies the given conditions we see that $t \mapsto \int_0^t \sigma(s,t)dW_s$ has a progressive modification. By the non-stochastic version of Fubini's theorem (Theorem 2.6) we also have that $\int_0^t \alpha(s,t)ds$ is progressive. Putting these together we get that r_t is progressive.

Moving on to our zero-coupon bonds we have by definition that

$$P(t,T) = e^{\int_t^T f(t,s)ds}.$$

We can also specify the dynamics of $P(t,T)$ in t.

Lemma 4.1. *For every $T > 0$ we have that $t \mapsto P(t,T)$ is an Itô process with the following dynamics.*

$$\frac{dP(t,T)}{P(t,T)} = (r_t + b(t,T))dt + v(t,T)dW_t.$$

For $t \leq T$ where we have defined the following.

$$v(t,T) = -\int_t^T \sigma(t,u)du$$

$$b(t,T) = -\int_t^T \alpha(t,u)du + \frac{1}{2}\left|\int_t^T \sigma(t,u)du\right|^2$$

The proof of this statement will be omitted, but it will be a direct consequence of the Leibniz integral rule for Itô processes, Lemma 6.1, which will be proved in Chapter 6.

As a corollary we can also consider the discounted zero-coupon bond price.

Corollary 4.1. *For all $t \leq T$ we have*

$$d\left(\frac{P(t,T)}{B_t}\right) / \frac{P(t,T)}{B_t} = b(t,T)dt + v(t,T)dW_t.$$

Proof. The proof is a repetition of the arguments in equation (3.1). □

For Q to be an ELMM we need by definition that $\frac{P(t,T)}{B_t}$ is a local martingale under Q. We assume that $Q \sim P$ exists as an equivalent measure on the Girsanov form given by equation (3.2) and we let $W_t^* = W_t - \int_0^t \gamma_s ds$ be the Wiener process under Q from Girsanov's theorem.

Theorem 4.1 (HJM Drift Condition)**.** *We have that Q is an ELMM if and only if*

$$b(t,T) = -v(t,T)\gamma_t \; dP \times dt\text{-a.e. for all } T.$$

In this case we also have that the Q-dynamics of $f(t,T)$ for $t \leq T$ is

$$df(t,T) = \left(\sigma(t,T)\int_t^T \sigma(t,s)ds\right)dt + \sigma(t,T)dW_t^*.$$

We also have from Corollary 4.1 and Lemma 2.4 that

$$\frac{P(t,T)}{B_t} = P(0,T)\mathcal{E}_t(v(\cdot,T) \bullet W^*)$$

for $t \leq T$.

4.2 Forward Measures

Most of our financial theory has used the concept of an ELMM Q to develop arbitrage free pricing of claims and assets. In this chapter we extend this notion to include more equivalent measures to get several useful results. To do this we must assume that Q is an EMM. With this assumption we get that $P(t,T)/B_t$ is a martingale which yields the following fact

$$E_Q\left[\frac{P(T,T)}{P(0,T)B_T}\right] = E_Q\left[\frac{1}{P(0,T)B_T}\right] = \frac{P(0,T)}{P(0,T)} = 1.$$

This allows us to define an equivalent measure $Q_T \sim Q$ on $\mathcal{F}_T$ by

$$\frac{dQ^T}{dQ} = \frac{1}{P(0,T)B_T} = \frac{P(T,T)}{P(0,T)B_T},$$

for any T. These measure are known as forward measures, furthermore, since $P(t,T)/B_t$ is a martingale we can define

$$\left.\frac{dQ^T}{dQ}\right|_{\mathcal{F}_t} = E_Q\left[\frac{dQ^T}{dQ}\middle|\mathcal{F}_t\right] = \frac{P(t,T)}{P(0,T)B_t}.$$

Remark 4. *We could equivalently define $\left.\frac{dQ^T}{dQ}\right|_{\mathcal{F}_t}$ as the Radon-Nikodym derivative of Q^T with respect to Q when both measures are restricted to $\mathcal{F}_t$. This equivalence follows from*

$$\int_A X dQ^T = \int_A X\frac{dQ^T}{dQ}dQ = \int_A E_Q\left[X\frac{dQ^T}{dQ}\middle|\mathcal{F}_t\right]dQ = \int_A XE_Q\left[\frac{dQ^T}{dQ}\middle|\mathcal{F}_t\right]dQ$$

for any Q^T-integrable and $\mathcal{F}_t$-measurable X, and any $A \in \mathcal{F}_t$.

Regardless of definition we know that this process, by Theorem 4.1, has the representation.

$$\left.\frac{dQ^T}{dQ}\right|_{\mathcal{F}_t} = \mathcal{E}_t(v(\cdot, T) \bullet W^*).$$

Applying Girsanov's theorem we get that W^T defined by

$$dW_t^* = dW_t^T + v(t, T)dt$$

is a Wiener process under Q^T.

In order to properly utilize these forward measures we will need a small result that generalizes Bayes' rule to conditional expectations based on σ-algebras.

Lemma 4.2 (Bayes' Rule)**.** *We denote*

$$D_t = \left.\frac{d\mu}{d\upsilon}\right|_{\mathcal{F}_t}$$

for any two probability measures μ and υ. We then have for any μ-integrable and $\mathcal{F}_T$-measurable random variable X that

$$E_\mu[X|\mathcal{F}_t] = \frac{E_\upsilon[XD_T|\mathcal{F}_t]}{D_t}.$$

Since this is left as an exercise in [3] we will provide our own proof.

Proof. From the argument in Remark 4 we know that D_t is equal to the Radon-Nikodym derivative of μ with respect to υ when both measures are restricted to $\mathcal{F}_t$. This yields for any $K \in \mathcal{F}_t \subset \mathcal{F}_T$ that

$$\begin{aligned}\int_K E_\mu[X|\mathcal{F}_t] D_t d\upsilon &= \int_K E_\mu[X|\mathcal{F}_t] d\mu \\ &= \int_K X d\mu \\ &= \int_K X D_T d\upsilon \\ &= \int_K E_\upsilon[XD_T|\mathcal{F}_t] d\upsilon,\end{aligned}$$

which proves the result. □

Using this formula we get the following useful result.

Lemma 4.3. *For $t \leq min(S, T)$ we have that $\frac{P(t,S)}{P(t,T)}$ is a Q^T-martingale, we also have the representation*

$$\frac{P(t, S)}{P(t, T)} = \frac{P(0, S)}{P(0, T)} \mathcal{E}_t\left(v_{S,T} \bullet W^T\right),$$

where we define

$$v_{S,T}(t) = \int_S^T \sigma(t, u)du.$$

Proof. Let $s \leq t \leq min(S,T)$, Bayes' rule then yields

$$E_{Q^T}\left[\frac{P(t,S)}{P(t,T)}\middle|\mathcal{F}_s\right] = \frac{E_Q\left[\frac{P(t,T)}{P(0,T)B_t}\frac{P(t,S)}{P(t,T)}\middle|\mathcal{F}_s\right]}{\left(\frac{P(s,T)}{P(0,T)B_t}\right)} = \frac{\left(\frac{P(s,S)}{B_s}\right)}{\left(\frac{P(s,T)}{B_s}\right)} = \frac{P(s,S)}{P(s,T)}.$$

□

As a result of this lemma, we can relate Q^S and Q^T for $t \leq min(T,S)$ by

$$\left.\frac{dQ^S}{dQ^T}\right|_{F_t} = \left.\frac{dQ^S}{dQ}\right|_{F_t}\left.\frac{dQ}{dQ^T}\right|_{F_t} = \frac{P(t,S)P(0,T)}{P(0,S)P(t,T)} = \mathcal{E}_t\left(v_{S,T} \bullet W^T\right).$$

Remark 5. *When we defined an EMM we required that the discounted assets were martingales under Q. However, from this lemma we see that if we used $P(t,T)$ as our numeraire (the discount factor) we could use Q^T as an EMM instead. This means that we now have an EMM for every T, each with its own numeraire. Note that doing so would assume that all our assets, except the bank account, would be zero-coupon bonds.*

The main advantage with forward measures is option pricing, where we have the following result.

Proposition 4.2. *Let X be a T-claim with*

$$E_Q\left[\frac{|X|}{B_T}\right] < \infty.$$

We then have that

$$E_{Q^T}\left[|X|\right] < \infty,$$

and that

$$B_t E_Q\left[\frac{X}{B_T}\middle|\mathcal{F}_t\right] = P(t,T)E_{Q^T}\left[X\middle|\mathcal{F}_t\right].$$

Proof. The first statement follows from

$$E_{Q^T}\left[|X|\right] = E_Q\left[|X|\frac{dQ^T}{dQ}\right] = E_Q\left[\frac{|X|}{P(0,T)B_T}\right] < \infty.$$

For the second statement we have by Bayes' rule that

$$\begin{aligned} B_t E_Q\left[\frac{X}{B_T}\middle|\mathcal{F}_t\right] &= P(0,T)B_t E_Q\left[\frac{X}{B_T P(0,T)}\middle|\mathcal{F}_t\right] \\ &= P(0,T)B_t\frac{P(t,T)}{P(0,T)B_t}E_{Q^T}\left[X\frac{B_T P(0,T)}{B_T P(0,T)}\middle|\mathcal{F}_t\right] \\ &= P(t,T)E_{Q^T}\left[X\middle|\mathcal{F}_t\right]. \end{aligned}$$

□

In the rest of this section we will use this result to, under certain conditions, calculate the arbitrage free price of assets more explicitly. We will follow the arguments from [3, Chapter 7.2], but will present a more thorough proof, along with a generalization of some results. In [3] the price of a call option was calculated for $t = 0$, we will generalize this to any t, as well as calculating the price of digital options.

Consider a European call option with strike price K and expiry date T on a zero-coupon bond with maturity at time S. Such an option has a payout of $(P(T,S) - K)^+$, which means the unique arbitrage free price at time t must be

$$B_t E_Q \left[\frac{(P(T,S) - K)^+}{B_T} \middle| \mathcal{F}_t \right].$$

We have an explicit formula for this expression under certain conditions by the following result.

Theorem 4.2. *Assume a general financial setup as described in Section 3 and with interest rates determined by an HJM model as described in Section 4.1. Further assume that the ELMM Q is an EMM and that the volatility process $\sigma(t,T)$ of the forward rates $f(t,T)$ are deterministic.*

We then have that unique arbitrage-free price of a European call option with payout $(P(T,S) - K)^+$ is

$$P(t,S)\Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right) + \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}} \right)$$
$$- KP(t,T)\Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right) - \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}} \right),$$

where Φ is the distribution function of the standard normal distribution.

Proof. We start by splitting up the pricing formula.

$$B_t E_Q \left[\frac{(P(T,S) - K)^+}{B_T} \middle| \mathcal{F}_t \right]$$
$$= B_t E_Q \left[\frac{P(T,S)\mathbb{1}_{P(T,S)\geq K}}{B_T} \middle| \mathcal{F}_t \right] - K B_t E_Q \left[\frac{\mathbb{1}_{P(T,S)\geq K}}{B_T} \middle| \mathcal{F}_t \right].$$

Applying Proposition 4.2 to the first term we get

$$B_t E_Q \left[\frac{P(T,S)\mathbb{1}_{P(T,S)\geq K}}{B_T} \middle| \mathcal{F}_t \right] = P(t,S) E_{Q^T} \left[P(T,S)\mathbb{1}_{P(T,S)\geq K} \middle| \mathcal{F}_t \right],$$

and by applying Bayes' rule and Lemma 4.3 we get

$$
\begin{aligned}
&P(t,T)E_{Q^T}\left[P(T,S)\mathbb{1}_{P(T,S)\geq K}\middle|\mathcal{F}_t\right] \\
&= P(t,T)\frac{dQ^S}{dQ^T}\bigg|_{\mathcal{F}_t} E_{Q^S}\left[\frac{dQ^T}{dQ^S}\bigg|_{\mathcal{F}_T} P(T,S)\mathbb{1}_{P(T,S)\geq K}\middle|\mathcal{F}_t\right] \\
&= P(t,T)\frac{P(t,S)P(0,T)}{P(0,S)P(t,T)}E_{Q^S}\left[\frac{P(T,T)P(0,S)}{P(0,T)P(T,S)}P(T,S)\mathbb{1}_{P(T,S)\geq K}\middle|\mathcal{F}_t\right] \\
&= P(t,S)E_{Q^S}\left[\mathbb{1}_{P(T,S)\geq K}\middle|\mathcal{F}_t\right] \\
&= P(t,S)Q^S(P(T,S)\geq K|\mathcal{F}_t).
\end{aligned}
$$

The second term can be simplified in a similar manner, by Proposition 4.2 we have

$$
\begin{aligned}
-KB_tE_Q\left[\frac{\mathbb{1}_{P(T,S)\geq K}}{B_T}\middle|\mathcal{F}_t\right] &= -KP(t,S)E_{Q^T}\left[\mathbb{1}_{P(T,S)\geq K}\middle|\mathcal{F}_t\right] \\
&= -KP(t,S)Q^T(P(T,S)\geq K|\mathcal{F}_t).
\end{aligned}
$$

It now remains to calculate the probability of $P(T,S)\geq K$ under Q^S and Q^T, to do this we first note that

$$
\begin{aligned}
Q^S(P(T,S)\geq K|\mathcal{F}_t) &= Q^S\left(\frac{P(T,T)}{P(T,S)}\frac{P(t,S)}{P(t,T)}\leq\frac{1}{K}\frac{P(t,S)}{P(t,T)}\middle|\mathcal{F}_t\right), \\
Q^T(P(T,S)\geq K|\mathcal{F}_t) &= Q^T\left(\frac{P(T,S)}{P(T,T)}\frac{P(t,T)}{P(t,S)}\geq K\frac{P(t,T)}{P(t,S)}\middle|\mathcal{F}_t\right).
\end{aligned}
\tag{4.1}
$$

To simplify notation we will define

$$
\Gamma_u^{T,S} = \frac{P(u,T)}{P(u,S)},
$$

$$
\Gamma_u^{S,T} = \frac{P(u,S)}{P(u,T)}.
$$

From Lemma 4.3 we have the representations

$$
\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} = exp\left(\int_t^T v_{T,S}(u)dW_u^T - \frac{1}{2}\int_t^T |v_{T,S}(u)|^2du\right)
$$

and

$$
\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}} = exp\left(\int_t^T v_{S,T}(u)dW_u^S - \frac{1}{2}\int_t^T |v_{S,T}(u)|^2du\right).
$$

From this we see that if $v_{T,S}$ is deterministic then both of these are log-normally distributed by Theorem 2.23 under their respective measures. This motivates our assumption that $\sigma(t,T)$, and therefore $v_{T,S}$, is deterministic.

By the assumption of independent increments of the Wiener process we have that the process $(\omega,t)\mapsto W_u^T, u\leq t$ and the process $(\omega,t)\mapsto W_u^T, u\geq t$ are independent. By Lemma 2.3 we get that the σ-algebra generated by W_u^T for $u\geq t$, denoted $\mathcal{F}_t^*$, must then be independent of $\mathcal{F}_t$. Since $\int_t^T v_{T,S}(u)dW_u^T$ is

measurable with respect to $\mathcal{F}_t^*$ it must therefore be independent of $\mathcal{F}_t$. Since this argument holds for W^S as well we have that $\int_t^T v_{S,T}(u)dW_u^S$ is independent of $\mathcal{F}_t$. From this we have that $\Gamma_T^{T,S}/\Gamma_t^{T,S}$ and $\Gamma_T^{S,T}/\Gamma_t^{S,T}$ are independent of $\mathcal{F}_t$, by applying Lemma 2.5 we get

$$Q^S\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{K\Gamma_t^{T,S}}\middle|\mathcal{F}_t\right) = Q^S\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{K\Gamma_t^{T,S}}\middle|\sigma\left(\Gamma_t^{T,S}\right)\right),$$
$$Q^T\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}} \geq \frac{K}{\Gamma_t^{S,T}}\middle|\mathcal{F}_t\right) = Q^T\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}} \geq \frac{K}{\Gamma_t^{S,T}}\middle|\sigma\left(\Gamma_t^{S,T}\right)\right).$$

Furthermore, we have

$$\begin{aligned} Q^S\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{K\Gamma_t^{T,S}}\middle|\sigma\left(\Gamma_t^{T,S}\right)\right) &= Q^S\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{Kx}\right)\Big|_{x=\Gamma_t^{T,S}}, \\ Q^T\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}} \geq \frac{K}{\Gamma_t^{S,T}}\middle|\sigma\left(\Gamma_t^{T,S}\right)\right) &= Q^T\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}} \geq \frac{K}{x}\right)\Big|_{x=\Gamma_t^{S,T}}. \end{aligned} \tag{4.2}$$

We will here prove the first part of equation (4.2), to show this we need

$$\begin{aligned} &E_{Q^S}\left[\mathbb{1}(A)\,\mathbb{1}\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{K\Gamma_t^{T,S}}\right)\right] \\ &= E_{Q^S}\left[\mathbb{1}(A)\,E_{Q^S}\left[\mathbb{1}\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{Kx}\right)\right]\Big|_{x=\Gamma_t^{T,S}}\right], \end{aligned}$$

for any set $A \in \sigma\left(\Gamma_t^{T,S}\right)$. We can without loss of generality assume that $A = \{\omega \in \Omega : \Gamma_t^{T,S} \in B\}$ with $B \in \mathcal{B}(\mathbb{R})$ as $\sigma(\Gamma_t^{T,S})$ is generated by these sets and the null sets of $\mathcal{F}_\infty$.

To keep notation orderly we here denote the indicator function of an event E as $\mathbb{1}(E)$ instead of the usual $\mathbb{1}_E(\omega)$, and will denote the distribution functions of the laws of $\Gamma_t^{T,S}$ and $\Gamma_T^{T,S}/\Gamma_t^{T,S}$ under Q^S as F and G respectively. Since we are dealing with a function of two independent random variables we can apply Theorem 2.14 and 2.15 to get

$$\begin{aligned} &E_{Q^S}\left[\mathbb{1}\left(\Gamma_t^{T,S} \in B\right)\mathbb{1}\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{K\Gamma_t^{T,S}}\right)\right] \\ &= \int_{\mathbb{R}}\int_{\mathbb{R}} \mathbb{1}(x \in B)\,\mathbb{1}\left(y \leq \frac{1}{Kx}\right) dG(y)dF(x) \\ &= \int_{\mathbb{R}} E_{Q^S}\left[\mathbb{1}(x \in B)\,\mathbb{1}\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{Kx}\right)\right] dF(x) \\ &= \int_{\mathbb{R}} E_{Q^S}\left[\mathbb{1}(x \in B)\right] E_{Q^S}\left[\mathbb{1}\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{Kx}\right)\right] dF(x) \\ &= E_{Q^S}\left[\mathbb{1}\left(\Gamma_t^{T,S} \in B\right) E_{Q^S}\left[\mathbb{1}\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{Kx}\right)\right]\Big|_{x=\Gamma_t^{T,S}}\right], \end{aligned}$$

which proves the first part of (4.2), the second part follows from the same arguments.

With equation (4.2) we can easily calculate the probabilities from equation (4.1). The computation starts by noting that from Theorem 2.23 we have the following.

$$\frac{log\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}}\right) + \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}} = \frac{\int_t^T v_{T,S}(u) dW^S(u)}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}}$$

is standard normally distributed under Q^S, and

$$-\frac{log\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}}\right) + \frac{1}{2}\int_t^T |v_{S,T}(u)|^2 du}{\sqrt{\int_t^T |v_{S,T}(u)|^2 du}} = -\frac{\int_t^T v_{S,T}(u) dW^T(u)}{\sqrt{\int_t^T |v_{S,T}(u)|^2 du}}$$

is standard normally distributed under Q^T. From this we get

$$\begin{aligned}
&Q^S\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}} \leq \frac{1}{Kx}\right)\Bigg|_{x=\Gamma_t^{T,S}} \\
&= Q^S\left(\frac{log\left(\frac{\Gamma_T^{T,S}}{\Gamma_t^{T,S}}\right) + \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}} \leq \frac{log\left(\frac{1}{Kx}\right) + \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}}\right)\Bigg|_{x=\Gamma_t^{T,S}} \\
&= \Phi\left(\frac{log\left(\frac{1}{K\Gamma_t^{T,S}}\right) + \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}}\right) \\
&= \Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right) + \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}}\right),
\end{aligned} \tag{4.3}$$

Where Φ is the distribution function for the standard normal distribution. Similarly, by noting that $v_{S,T} = -v_{T,S}$ implies $|v_{S,T}|^2 = |v_{T,S}|^2$ we get

$$
\begin{aligned}
&Q^S\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}} \geq \frac{K}{x}\right)\Bigg|_{x=\Gamma_t^{S,T}} \\
&= Q^S\left(\frac{log\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}}\right)+\frac{1}{2}\int_t^T |v_{S,T}(u)|^2du}{\sqrt{\int_t^T |v_{S,T}(u)|^2du}} \geq \frac{log\left(\frac{K}{x}\right)+\frac{1}{2}\int_t^T |v_{S,T}(u)|^2du}{\sqrt{\int_t^T |v_{S,T}(u)|^2du}}\right)\Bigg|_{x=\Gamma_t^{S,T}} \\
&= Q^S\left(-\frac{log\left(\frac{\Gamma_T^{S,T}}{\Gamma_t^{S,T}}\right)+\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}} \leq \frac{-log\left(\frac{K}{x}\right)-\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}}\right)\Bigg|_{x=\Gamma_t^{S,T}} \\
&= \Phi\left(\frac{-log\left(\frac{K}{\Gamma_t^{S,T}}\right)-\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}}\right) \\
&= \Phi\left(\frac{log\left(\frac{\Gamma_t^{S,T}}{K}\right)-\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}}\right) \\
&= \Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right)-\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}}\right).
\end{aligned}
\tag{4.4}
$$

With these probabilities explicitly calculated we get the price at time t of a European call with payout $(P(T,S)-K)^+$ as

$$P(t,S)Q^S(P(T,S)\geq K) - KP(t,T)Q^T(P(T,S)\geq K),$$

where we may plug in the probabilities from equation (4.3) and (4.4) to get

$$
\begin{aligned}
&P(t,S)\Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right)+\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}}\right) \\
&-KP(t,T)\Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right)-\frac{1}{2}\int_t^T |v_{T,S}(u)|^2du}{\sqrt{\int_t^T |v_{T,S}(u)|^2du}}\right),
\end{aligned}
$$

as desired. □

As a corollary to our calculations we will also get the price of a digital option with payout $\mathbb{1}(P(T,S)\geq K)$. The unique arbitrage free price at time $t\leq T$ for such an option is

$$B_t E_Q\left[\frac{\mathbb{1}(P(T,S)\geq K)}{B_T}\middle|\mathcal{F}_t\right].$$

Applying Proposition 4.2 we see that this equals

$$P(t,T)Q^T\big(P(T,S)\geq K\big|\mathcal{F}_t\big),$$

which equals

$$P(t,T)\Phi\left(\frac{log\left(\frac{P(t,S)}{KP(t,T)}\right) - \frac{1}{2}\int_t^T |v_{T,S}(u)|^2 du}{\sqrt{\int_t^T |v_{T,S}(u)|^2 du}}\right)$$

by applying equation (4.4).

Chapter 5

Classical Life Insurance

Life insurance policies are agreements where the insurer takes on risk from the insured by agreeing to compensate in the case of e.g. death or injury. There are also pension schemes which guarantee the insured an agreed-upon payment stream after retiring. Finding fair prices for these types of policies is an important part of actuarial science which helps insurance companies set premiums for their contracts.

In this chapter we will be developing the basic theory of life insurance in the case of deterministic interest rates. We will work within the time horizon $\mathbb{R}_+$ along with a filtered probability space $(\Omega, \mathcal{A}, \mathcal{F}, P)$ where the filtration $\mathcal{F}$ satisfies *the usual conditions*. These conditions state that $\mathcal{F}$ is right-continuous, complete and that $\mathcal{A}$ is the completion of $\sigma\left(\bigcup_{t\in\mathbb{R}_+} \mathcal{F}_t\right)$. We will also assume the existence of a regular càdlàg Markov chain X that is progressive with respect to $\mathcal{F}$.

We here mainly cover theory discussed in [6], which we refer to for further reading.

5.1 Basic Definitions

We will need the following definitions.

Definition 5.1.

$$I_i(t) = \mathbb{1}_{\{X_t=i\}}(t),$$
$$N_{i,j}(t) = \#\{s \in (0,t] : X_{s^-} = i, X_s = j\},$$

where # is the counting measure and $\mathbb{1}$ is the indicator function. In other words, $N_{i,j}(t)$ counts the number of jumps from state i to state j. Also note that X_{s^-} is shorthand for $\lim_{u\to s^-} X_u$ where similarly to before $\lim_{u\to s^-}$ means that $u \to s$ and $u < s$.

In order to use $N_{i,j}$ we need to ensure certain properties. By our assumption that X is càdlàg we have that $N_{i,j}$ is well defined and right-continuous. Lastly, we need that $N_{i,j}$ is of finite variation which. Since $N_{i,j}$ is monotone non-decreasing this is equivalent to it being finite, which is ensured by the following result from [11, Example, page 454-455].

Lemma 5.1. *The process $N_{i,j}$ is bounded on any finite interval almost surely. Furthermore, the expectation of the Lebesgue-Stieltjes integral of a function $b(s) : \mathbb{R} \to \mathbb{R}$ with respect to $dN_{i,j}$ can be calculated by*

$$E\left[\int_t^T b(s)dN_{j,k}(s)\middle|X_t = i\right] = \int_t^T b(s)p_{i,j}(t,s)\mu_{j,k}(s)ds.$$

With this we are ready to construct our insurance model, we will start by defining some terminology.

Definition 5.2. *A stochastic cash flow is a stochastic process A such that the paths of A are a.s. right-continuous and of bounded variation.*

Definition 5.3. *Policy functions are stochastic processes $a_i, a_{i,j}$ $(i, j \in \mathcal{S})$ that model either of the following insurance quantities.*

$$a_i(t) = \text{the total payout the insured receives over a time period } [0,t], \\ \text{assuming the state of the insured } X_u = i \text{ for all } u \in [0,t].$$

$$a_{i,j} = \text{the payment the insured receives when the state of the insured } X \\ \text{has a jump from } i \text{ to } j \text{ at time } t.$$

In order to properly define policy cash flows later, we will require that the paths of the a_i's are almost surely right-continuous and of finite variation. Note that in this chapter we will assume that our policy functions are deterministic, the stochastic element will be considered in later chapters.

Combining these definitions along with our Markov chain X we may then define the following.

Definition 5.4. *Let $a_i, a_{i,j}$ be policy functions describing an insurance policy. We then define the policy cash flow as*

$$A(t) = \sum_{i\in\mathcal{S}} A_i(t) + \sum_{\substack{i,j\in\mathcal{S}\\ i\neq j}} A_{i,j}(t),$$

where

$$A_i(t) = \int_0^t I_i(s)da_i(s) \text{ and } A_{i,j}(t) = \int_0^t a_{i,j}(s)dN_{i,j}(s).$$

These cash flows represent the total payouts from all different sources given a Markov chain representing the state of the insured over time.

We reintroduce our bank account B_t, but will in this chapter consider B as deterministic. We recall the definition as

$$B_t = exp\left(\int_0^t r_u du\right),$$

where r_u is some deterministic integrable function modelling the short-rate, the stochastic case will be considered in later chapters. Following convention and for ease of notation we will introduce the discount factor

$$v_t = \frac{1}{B_t} = exp\left(-\int_0^t r_u du\right).$$

We will similarly to Chapter 3 use B to discount the value of future money, in this case it will give us the current value of a policy cash flow.

For ease of notation we will be denoting

$$\int_{t^+}^{\infty} f(s)da(s) = \int_{(t,\infty)} f(s)da(s).$$
$$\int_{t^+}^{T} f(s)da(s) = \int_{(t,T]} f(s)da(s),$$

for any function f. Note that this means that $\int_{t^+}^{T} da = a(T) - a(t)$.

Definition 5.5. *The prospective value of a cash flow A, also known as the stochastic prospective reserve at time t is the discounted future value of the total cash flow, mathematically defined as*

$$V^+(t,A) = \frac{1}{v_t}\int_{t^+}^{\infty} v_s dA(s).$$

We may also break this equation down as

$$V^+(t,A) = \frac{1}{v_t}\left(\sum_{i\in\mathcal{S}}\int_{t^+}^{\infty} v_s dA_i(s) + \sum_{\substack{i,j\in\mathcal{S}\\ i\neq j}}\int_{t^+}^{\infty} v_s dA_{i,j}(s)\right),$$

or expand it completely to

$$V^+(t,A) = \frac{1}{v_t}\left(\sum_{i\in\mathcal{S}}\int_{t^+}^{\infty} v_s I_i(s) da_i(s) + \sum_{\substack{i,j\in\mathcal{S}\\ i\neq j}}\int_{t^+}^{\infty} v_s a_{i,j}(s) dN_{i,j}(s)\right).$$

Note that this definition deals with future information and is therefore not adapted. As a result we will need a new definition that only relies on information up to the current time.

Definition 5.6. *The prospective reserve $V^+_{\mathcal{F}}(t,A)$ is then defined as*

$$V^+_{\mathcal{F}}(t,A) = E[V^+(t,A)|\mathcal{F}_t]$$

Note that we here assume that $V^+(t)$ is integrable with respect to P for any t.

Furthermore, we can note that since $V^+(t,A)$ is a function of X_s for $s \geq t$ we have by the Markov property of X that

$$E[V^+(t,A)|\mathcal{F}_t] = E[V^+(t,A)|\sigma(X_t)].$$

Since X_t is discrete we have that

$$E[V^+(t,A)|\sigma(X_t)] = \sum_{i\in\mathcal{S}} I_i(t)E[V^+(t,A)|X_t = i],$$

which leads to another definition of the prospective reserve. For any $i \in \mathcal{S}$ *we define*

$$V_i^+(t,A) = E[V^+(t,A)|X_t = i].$$

Remark 6. *Note that this is analogous to the pricing we developed in the chapter on arbitrage theory, in this case however, we have no way of hedging the risk coming from the uncertainty in the state of the insured. This means that we could, in theory, price our reserves as* $E_Q[V^+(t,A)|\mathcal{F}_t]$ *for any equivalent measure* Q*. This is only a theoretical curiosity as the ordinary expectation is the only reasonable option in this case.*

Another way to consider the choice to use an equivalent measure is in pricing a coin flip with equal probability of yielding 1 *and* -1*. The rational fair price would be the expectation, i.e.* $1 \cdot \frac{1}{2} - 1 \cdot \frac{1}{2} = 0$*. However, in the absence of hedging possibilities we would not introduce any arbitrage by pricing the coin flip as* $1 \cdot q + -1 \cdot (1-q)$ *for any probability* $q \in (0,1)$*.*

Returning to our chosen definition of the prospective reserve we note that by splitting up the policy cash flow into the smaller cash flows $A_i, A_{i,j}$, we get a way to break the prospective reserve down as well.

Definition 5.7. *The mathematical reserves for being in state* j *over a time interval* $J \subset [t,\infty)$ *given* $X_t = i$ *is*

$$V_i(t,A_j,J) = E\left[\frac{1}{v_t}\int_J v_s I_j(s)da_j(s)\middle| X_t = i\right].$$

Similarly we define the mathematical reserve for transitioning from state i *to state* j *over a time interval* $J \subset [t,\infty)$ *given* $X_t = i$ *as*

$$V_i(t,A_{j,k},J) = E\left[\frac{1}{v_t}\int_J v_s a_{j,k}(s)dN_{j,k}(s)\middle| X_t = i\right].$$

We will also denote $V_i(t,A_j,(t,\infty))$ *and* $V_i(t,A_{j,k},(t,\infty))$ *as* $V_i(t,A_j)$ *and* $V_i(t,A_{j,k})$ *respectively.*

5.2 Integral and Differential Equations

With this we will present an explicit formula for calculating our mathematical reserves and, by extension, the prospective reserve.

Theorem 5.1. *Let* $J \subset [t,\infty) \subset \mathbb{R}_+$ *be an interval and let* $i,j,k \in \mathcal{S}$*. Furthermore we assume that* b,c *are functions such that* b *is integrable and* c *is right-continuous and of bounded variation on* J*. We then have*

$$E\left[\int_J I_j(s)dc(s)\middle| X_t = i\right] = \int_J p_{i,j}(t,s)dc(s)$$

and

$$E\left[\int_J b(s)dN_{j,k}(s)\middle|X_t=i\right]=\int_J b(s)p_{i,j}(t,s)\mu_{j,k}(s)ds$$

Proof. The first formula follows from Theorem 2.6, Fubini's theorem, which yields

$$\begin{aligned}E\left[\int_J I_j(s)dc(s)\middle|X_t=i\right]&=E\left[I_i(t)\int_J I_j(s)dc(s)\right]/P(X_t=i)\\&=\int_J E\left[I_i(t)I_j(s)\right]dc(s)/P(X_t=i)\\&=\int_J E[I_j(s)|X_t=i]dc(s)\\&=\int_J p_{i,j}(t,s)dc(s).\end{aligned}$$

The second equation is simply restating Lemma 5.1. □

We may now apply these to the mathematical reserves.

Theorem 5.2. *From Theorem 5.1 we have the following explicit formulas.*

$$V_i(t,A_j)=E\left[\frac{1}{v_t}\int_{t^+}^{\infty}v_sI_j(s)da_j(s)\middle|X_t=i\right]=\frac{1}{v_t}\int_{t^+}^{\infty}v_sp_{i,j}(t,s)da_j(s),$$

$$V_i(t,A_{j,k})=E\left[\frac{1}{v_t}\int_{t^+}^{\infty}v_sa_{j,k}(s)dN_{j,k}(s)\middle|X_t=i\right]=\frac{1}{v_t}\int_{t^+}^{\infty}v_sp_{i,j}(t,s)a_{j,k}(s)\mu_{j,k}(s)ds.$$

Putting these together we get the explicit formula for the prospective reserve.

$$v_tV_i^+(t,A)=\sum_{j\in\mathcal{S}}\int_{t^+}^{\infty}v_sp_{i,j}(t,s)da_j(s)+\sum_{\substack{j,k\in\mathcal{S}\\j\neq k}}\int_{t^+}^{\infty}v_sp_{i,j}(t,s)a_{j,k}(s)\mu_{j,k}(s)ds.$$

This leads into Thiele's differential equation, a useful tool for both analysis of policies and an efficient numerical method for computation.

To simplify we will suppress notation of the cash flow A and denote

$$W_i^+(t)=v_tV_i^+(t).$$

We start by presenting a useful integral equation for W^+.

Lemma 5.2 (Integral Lemma for Thiele's Equation)**.**

$$\begin{aligned}W_i^+(t)=&\sum_{j\in\mathcal{S}}p_{i,j}(t,u)W_j^+(u)\\&+\sum_{j\in\mathcal{S}}\left(\int_{t^+}^{u}v_sp_{i,j}(t,s)da_j(s)+\sum_{\substack{k\in\mathcal{S}\\k\neq j}}\int_{t^+}^{u}v_sp_{i,j}(t,s)\mu_{j,k}(s)a_{j,k}(s)ds\right)\end{aligned}$$

Proof. The idea is to use Theorem 5.2 to consider $W_i^+(t)$ as an integral over (t, ∞), and then split the interval into $(t, u]$ and (u, ∞) to get the two sums. The technical details will be omitted, but can be found in [6, Lemma 4.7.2]. □

With this we can discuss Thiele's differential equation, we will first consider the version where da_i is absolutely continuous with respect to the Lebesgue measure λ, i.e. $a_i(t) = \int_0^t \dot{a}_i(s)ds$ where $\dot{a}_i$ is the Radon-Nikodym derivative $da_i/d\lambda$. We then have the following result.

Theorem 5.3 (Thiele's Differential Equation)**.**

$$\frac{dW_i(t)}{dt} = -v_t \left(\dot{a}_i(t) + \sum_{\substack{j \in \mathcal{S} \\ i \neq j}} \mu_{i,j}(t) a_{i,j}(t) \right) - \sum_{j \in \mathcal{S}} \mu_{i,j}(t) W_j(t).$$

Proof. Note that since da_i is absolutely continuous we can simplify the notation of our integrals and get from applying our integral lemma with $u = t + h, h > 0$ that

$$\begin{aligned} W_i^+(t) = &\sum_{j \in \mathcal{S}} p_{i,j}(t, t+h) W_j^+(t+h) \\ &+ \sum_{j \in \mathcal{S}} \left(\int_t^{t+h} v_s p_{i,j}(t,s) da_j(s) + \sum_{\substack{k \in \mathcal{S} \\ k \neq j}} \int_t^{t+h} v_s p_{i,j}(t,s) \mu_{j,k}(s) a_{j,k}(s) ds \right). \end{aligned}$$

By subtracting this from $W_i^+(t+h)$ we get that

$$\begin{aligned} W_i^+(t+h) - W_i^+(t) = &W_i^+(t+h) - \sum_{j \in \mathcal{S}} p_{i,j}(t, t+h) W_j^+(t+h) \\ &- \sum_{j \in \mathcal{S}} \left(\int_t^{t+h} v_s p_{i,j}(t,s) da_j(s) + \sum_{\substack{k \in \mathcal{S} \\ k \neq j}} \int_t^{t+h} v_s p_{i,j}(t,s) \mu_{j,k}(s) a_{j,k}(s) ds \right). \end{aligned}$$

We then proceed with dividing by h and note the following facts.

- For $i \neq j$ we have by definition
$$\lim_{h \to 0^+} p_{i,j}(t, t+h)/h = \mu_{i,j}(t).$$
- Similarly, by definition of $\mu_{i,i}(t)$ we have
$$\lim_{h \to 0^+} (1 - p_{i,i}(t, t+h))/h = -\mu_{i,i}(t).$$
- By the fundamental theorem of calculus along with our assumption that $(da_i/d\lambda)(s) = \dot{a}_i$ we have

$$\lim_{h\to 0^+}\int_t^{t+h} v_s p_{i,j}(t,s)da_j(s)/h$$
$$= \lim_{h\to 0^+}\int_t^{t+h} v_s p_{i,j}(t,s)\dot{a}_j(s)ds/h$$
$$= v_t p_{i,j}(t,t)\dot{a}_j(t).$$

- By the same argument we have

$$= \lim_{h\to 0^+}\int_t^{t+h} v_s p_{i,j}(t,s)\mu_{j,k}(s)a_{j,k}(s)ds/h$$
$$= v_t p_{i,j}(t,t)\mu_{j,k}(t)a_{j,k}(t).$$

- Lastly we note that $p_{i,j}(t,t) = 0$ for $i \neq j$ and that $p_{i,i}(t,t) = 1$.

Putting these together we see that

$$\lim_{h\to 0^+}(W_i^+(t+h) - W_i^+(t))/h$$
$$= -\sum_{j\in\mathcal{S}}\mu_{i,j}(t)W_j^+(t) - v_t\dot{a}_i(t) - \sum_{\substack{k\in\mathcal{S}\\k\neq i}} v_t\mu_{i,k}(t)a_{i,k}(t)$$
$$= -v_t\left(\dot{a}_i(t) - \sum_{\substack{j\in\mathcal{S}\\j\neq i}}\mu_{i,k}(t)a_{i,j}(t)\right) - \sum_{j\in\mathcal{S}}\mu_{i,j}(t)W_j^+(t),$$

as desired. This proves Thiele's equation for the right derivative of V_i^+, but by repeating the argument where we instead apply our integral lemma to $W_i^+(t-h)$ with $u = t$ we get the same result for the left derivative and conclude our argument.

□

We also have a version of Thiele's equation for V_i^+ where we can also allow $a_i(t)$ to have a single discontinuity at $t = T$, note that this only affects the boundary condition.

Theorem 5.4. *Assume that da_i is absolutely continuous with respect to λ on $[0,T)$ and assume that $a_i(t)$ has at most one discontinuity in $t = T$. For a reserve $V_i^+(t), t \in [0,T]$ we then have*

$$\frac{dV_i^+(t)}{dt} = r_t V_i^+(t) - \dot{a}_i(t) - \sum_{\substack{j\in\mathcal{S}\\i\neq j}}\mu_{i,j}(t)\Big(a_{i,j}(t) + V_j^+(t) - V_i^+(t)\Big),$$

with a border condition of $V_i^+(T) = a_i(T) - a_i(T^-)$. Note that if a_i has no discontinuities the differential equation would hold for $V_i^+(t), t \in \mathbb{R}_+$.

Proof. Since da_i is absolutely continuous on $[0,T)$ we will again use the Radon-Nikodym derivative $\dot{a}_i$. By also taking into account the single discontinuity of a_i at T we get that

$$\int_0^T f(s)da_i(s) = f(T)(a_i(T) - a_i(T^-)) + \int_0^T f(s)\dot{a}_i(s)ds$$

for any function f that is integrable with respect to da_i. Applying this to our explicit formula for V_i^+ (Theorem 5.2) we get

$$\begin{aligned} v_t V_i^+(t) = &\sum_{j\in\mathcal{S}} v_T p_{i,j}(t,T)(a_j(T) - a_j(T^-)) \\ &+ \sum_{j\in\mathcal{S}} \int_{t^+}^\infty v_s p_{i,j}(t,s)da_j(s) + \sum_{\substack{j,k\in\mathcal{S}\\ j\neq k}} \int_{t^+}^\infty v_s p_{i,j}(t,s)a_{j,k}(s)\mu_{j,k}(s)ds. \end{aligned}$$

We will simplify this expression by denoting the following.

$$G_i(t) = v_T \sum_{j\in\mathcal{S}} p_{i,j}(t,T)(a_j(T) - a_j(T^-))$$

$$F_i(t,s) = \sum_{j\in\mathcal{S}} v_s p_{i,j}(t,s) \left(\dot{a}_j(s) + \sum_{\substack{k\in\mathcal{S}\\ k\neq j}} a_{j,k}(s)\mu_{j,k}(s) \right) \tag{5.1}$$

Note that this implies that

$$v_t V_i^+(t) = G_i(t) + \int_t^T F_i(t,s)ds.$$

To differentiate V we will need the derivatives of G_i and F_i. We will here use the backwards Kolmogorov equation (theorem 2.28). Note that by

$$-\mu_{i,i}(t) = \sum_{\substack{k\in\mathcal{S}\\ k\neq i}} \mu_{i,k}(t)$$

we have

$$\begin{aligned} \frac{\partial p_{i,j}(t,T)}{\partial t} &= -\sum_{k\in\mathcal{S}} \mu_{i,k}(t)p_{k,j}(t,T) \\ &= -\mu_{i,i}(t)p_{i,j}(t,T) - \sum_{\substack{k\in\mathcal{S}\\ k\neq j}} \mu_{i,k}(t)p_{k,j}(t,T) \\ &= \sum_{\substack{k\in\mathcal{S}\\ k\neq j}} \mu_{i,k}(t)(p_{i,j}(t,T) - p_{k,j}(t,T)). \end{aligned}$$

By applying this we can calculate the derivatives of G and F. For G we have

$$\begin{aligned}\frac{dG_i(t)}{dt} &= v_T \sum_{j\in\mathcal{S}} \frac{\partial p_{i,j}(t,T)}{\partial t}(a_j(T) - a_j(T^-)) \\ &= v_T \sum_{j\in\mathcal{S}} \sum_{\substack{k\in\mathcal{S}\\k\neq j}} \mu_{i,k}(t)(p_{i,j}(t,T) - p_{k,j}(t,T))(a_j(T) - a_j(T^-)) \\ &= \sum_{\substack{k\in\mathcal{S}\\k\neq j}} \mu_{i,k}(t)(G_i(t) - G_k(t)),\end{aligned} \tag{5.2}$$

and by using the same method for F we get

$$\frac{\partial F_i(t,s)}{\partial t} = \sum_{\substack{k\in\mathcal{S}\\k\neq j}} \mu_{i,k}(t)(F_i(t,s) - F_k(t,s)). \tag{5.3}$$

The last thing we need is to differentiate

$$I(x,y) = \int_y^T F_i(x,s)ds$$

for $x = y = t$. We may calculate the partial derivatives to be the following.

- For the derivative with respect to y we have by the fundamental theorem of calculus that
$$\frac{\partial\left(\int_0^y F_i(x,s)ds\right)}{\partial y} = F_i(x,y).$$
However, since $\int_0^y F_i(x,s)ds + I(x,y) = \int_0^T F_i(x,s)ds$ is constant in y we have
$$\frac{\partial I(x,y)}{\partial y} = -F_i(x,y).$$

- In calculating the derivative with respect to x we start by denoting
$$f_j(s) = v_s\dot{a}_j(s) + v_s \sum_{\substack{k\in\mathcal{S}\\k\neq j}} a_{j,k}(s)\mu_{j,k}(s),$$
and note that f_j does not rely on x. We then define μ_j by
$$\frac{d\mu_j}{d\lambda}(s) = f_j(s),$$
and get
$$\lim_{h\to 0}\frac{I(x+h,y) - I(x,y)}{h} = \lim_{h\to 0}\sum_{j\in\mathcal{S}}\int_y^T \frac{p_{i,j}(x+h,s) - p_{i,j}(x,s)}{h}d\mu_j(s).$$
By the mean-value theorem we have

$$\frac{p_{i,j}(x+h,s)-p_{i,j}(x,s)}{h}=\frac{\partial p_{i,j}(\hat{x},s)}{\partial x}$$

for some $\hat{x}$ with $|\hat{x}-x|\leq h$. Since $\partial p_{i,j}(x,s)/\partial x$ is continuous we see that $\partial p_{i,j}(\hat{x},s)/\partial x$ is bounded for $s\in[y,T]$ uniformly in h, and we can apply the dominated convergence theorem to get

$$\lim_{h\to 0}\int_y^T \frac{\partial p_{i,j}(\hat{x},s)}{\partial x}d\mu_j(s)=\int_y^T \frac{\partial p_{i,j}(x,s)}{\partial x}d\mu_j(s).$$

With this we conclude that

$$\frac{\partial I(x,y)}{\partial x}=\int_y^T \frac{\partial F_i(x,s)}{\partial x}ds.$$

With the partial derivatives calculated we can move on to $I(t,t)$, the classical chain rule then yields that

$$\begin{aligned}\frac{dI(t,t)}{dt}&=\left.\frac{\partial I(x,y)}{\partial y}\right|_{(x,y)=(t,t)}+\left.\frac{\partial I(x,y)}{\partial x}\right|_{(x,y)=(t,t)}\\&=-F_i(t,t)+\int_t^T\frac{\partial F(t,s)}{\partial t}ds\end{aligned}$$

To calculate the integral term we may note that by combining equation (5.1) and (5.3) we get

$$\begin{aligned}\int_t^T\frac{\partial F_i(t,s)}{\partial t}ds&=\int_t^T\sum_{\substack{k\in\mathcal{S}\\k\neq j}}\mu_{i,k}(F_i(t,s)-F_k(t,s))ds\\&=\sum_{\substack{k\in\mathcal{S}\\k\neq j}}\mu_{i,k}\int_t^T(F_i(t,s)-F_k(t,s))ds\\&=v_t\sum_{\substack{k\in\mathcal{S}\\k\neq j}}\mu_{i,k}(V_i^+(t)-V_k^+(t))-\sum_{\substack{k\in\mathcal{S}\\k\neq j}}\mu_{i,k}(G_i(t)-G_k(t)).\end{aligned}$$

By recognizing the derivative of G from equation (5.2) we get

$$\int_t^T\frac{\partial F_i(t,s)}{\partial t}ds=v_t\sum_{\substack{j\in\mathcal{S}\\j\neq i}}\mu_{i,j}(V_i^+(t)-V_j^+(t))-\frac{dG_i(t)}{dt}. \tag{5.4}$$

Lastly, we have that

$$\begin{aligned}F_i(t,t)&=\sum_{j\in\mathcal{S}}v_tp_{i,j}(t,t)\left(\dot{a}_j(t)+\sum_{\substack{k\in\mathcal{S}\\k\neq j}}a_{j,k}(t)\mu_{j,k}(t)\right)\\&=v_t\dot{a}_i(t)+v_t\sum_{\substack{j\in\mathcal{S}\\j\neq i}}a_{i,k}(t)\mu_{i,k}(t)\end{aligned}$$

With this we are ready to differentiate V_i^+ directly by

$$\frac{d(v_t V_i^+(t))}{dt} = -r_t v_t V_i^+(t) + v_t \frac{dV_i^+(t)}{dt}.$$

We also have that

$$\begin{aligned}
\frac{d(v_t V_i^+(t))}{dt} &= \frac{dG_i(t)}{dt} + \frac{dI(t,t)}{dt} \\
&= \frac{dG_i(t)}{dt} + -F_i(t,t) + \int_t^T \frac{\partial F(t,s)}{\partial t} ds \\
&= -v_t \dot{a}_i(t) - v_t \sum_{\substack{j \in \mathcal{S} \\ j \neq i}} a_{i,k}(t)\mu_{i,k}(t) + v_t \sum_{\substack{k \in \mathcal{S} \\ k \neq j}} \mu_{i,k}(V_i^+(t) - V_j^+(t)) \\
&= -v_t \dot{a}_i(t) - v_t \mu_{i,k}(t) \sum_{\substack{j \in \mathcal{S} \\ j \neq i}} \left(a_{i,k}(t) + V_j^+(t) - V_i^+(t)\right).
\end{aligned}$$

Combining these formulas for $d(v_t V_i^+(t))/dt$ and dividing by v_t we get

$$\frac{dV_i^+(t)}{dt} = r_t V_i^+(t) - \dot{a}_i(t) - \sum_{\substack{j \in \mathcal{S} \\ i \neq j}} \mu_{i,j}(t) \Big(a_{i,j}(t) + V_j^+(t) - V_i^+(t) \Big)$$

as desired.

□

Chapter 6

Insurance Policies with Stochastic Interest Rate under HJM

6.1 Setup

We are now finally ready to tackle insurance policies in the case where both the interest rate and the policy functions are stochastic.

In this section we will consider two filtered probability spaces, the first of which represents the world of insurance, denoted by $(\Omega^X, \mathcal{F}^X_\infty, \mathcal{F}^X, P^X)$. We will here assume the following.

- The filtration $\mathcal{F}^X$ is the augmented filtration generated by a right-continuous and regular Markov chain X, making $\mathcal{F}^X$ complete and right-continuous by Definition 2.23.
- We also assume $\mathcal{F}^X_\infty = \sigma\left(\bigcup_{t\in\mathbb{R}_+} \mathcal{F}^X_t\right)$, which means the last of the regular conditions are satisfied.

The forward rate f and policy functions $a_i, a_{i,j}$ will be assumed to exist in the financial world denoted by $(\Omega^W, \mathcal{F}^W_\infty, \mathcal{F}^W, P^W)$. We will here assume the following.

- The filtration $\mathcal{F}^W$ is generated by a Wiener process W, note that this makes $\mathcal{F}^W$ right-continuous and complete by Theorem 2.20.
- For the last of the regular conditions we require $\mathcal{F}^W_\infty = \sigma\left(\bigcup_{t\in\mathbb{R}_+} \mathcal{F}^W_t\right)$.
- We assume a market as discussed in Chapter 3.
  - The market is assumed to have a unique ELMM as in Definition 3.5. Note that this makes the market arbitrage free and complete by Theorem 3.3 and 3.5.
  - We assume an interest rate model within the HJM framework as discussed in Chapter 4.1.

- Lastly, we require policy functions as in Definition 5.3, but will now assume them stochastic and adapted to $\mathcal{F}^W$, making them progressive by Proposition 2.5.

We will combine these spaces by considering the product space, denoted as

$$(\Omega, \mathcal{F}_\infty, P) = (\Omega^W \times \Omega^X, \mathcal{F}_\infty^W \otimes \mathcal{F}_\infty^X, P^W \times P^X).$$

To justify the notation of $\mathcal{F}_\infty$ we define the filtration $\mathcal{F}$ by

$$\mathcal{F}_t = \sigma\left(\mathcal{F}_t^W \otimes \mathcal{F}_t^X \cup \mathcal{N}\right),$$

where $\mathcal{N}$ is the collection of all null-sets on $\mathcal{F}_\infty$. Note that $\mathcal{F}$ is complete by definition and inherits right-continuity from $\mathcal{F}^W$ and $\mathcal{F}^X$.

The first thing we will need is to find a way to price our reserves within the structure of financial markets. Similarly to Chapter 5 we will define the fair price of any payout based purely on insurance information (i.e. $\mathcal{F}_\infty^X$-measurable random variables) to be it's ordinary expectation as mentioned in Remark 6. In the financial world however, we have from Theorem 3.5 that any $\mathcal{F}_\infty^W$-measurable T-claim Y with $E_{Q^W}[|Y|/B_T] < \infty$ has a unique arbitrage-free price at time t given by $E_{Q^W}[|Y|/B_T|\mathcal{F}_t^X]$. Combining these we get that the fair price of any $\mathcal{F}_\infty$-measurable T-claim Y must be

$$E_{P^X}\left[E_{Q^W}\left[\frac{Y}{B_T}\middle|\mathcal{F}_t^W\right]\middle|\mathcal{F}_t^X\right]. \tag{6.1}$$

We may also write this more compactly as

$$E_Q\left[\frac{Y}{B_T}\middle|\mathcal{F}_t\right],$$

where $Q = Q^W \times P^X$ by the following result.

Proposition 6.1. *For any integrable $\mathcal{F}_\infty$-measurable variable Z we have*

$$E_{P^X}\left[E_{Q^W}\left[Z|\mathcal{F}_t^W\right]|\mathcal{F}_t^X\right] = E_Q\left[Z|\mathcal{F}_t\right]$$

Proof. By the definition of conditional expectations (Definition 2.17) the result would follow from showing that the left-hand side is $\mathcal{F}_t$-measurable with

$$\int_G E_{P^X}\left[E_{Q^W}\left[Z|\mathcal{F}_t^W\right]|\mathcal{F}_t^X\right] dQ = \int_G Z dQ$$

for any $G \in \mathcal{F}_\infty$. By Theorem 2.3, the monotone class theorem for functions, it suffices to prove the result for $Z = \mathbb{1}_A\mathbb{1}_B$ with $A \in \mathcal{F}_\infty^W, B \in \mathcal{F}_\infty^X$. Furthermore, by Theorem 2.1, the monotone class theorem for sets, it suffices the result for $G = C \times D$ with $C \in \mathcal{F}_\infty^W, D \in \mathcal{F}_\infty^X$. By noting

$$E_{P^X}\left[E_{Q^W}\left[\mathbb{1}_A\mathbb{1}_B|\mathcal{F}_t^W\right]|\mathcal{F}_t^X\right] = E_{Q^W}\left[\mathbb{1}_A|\mathcal{F}_t^W\right]E_{P^X}\left[\mathbb{1}_B|\mathcal{F}_t^X\right]$$

we see that the double conditional expectation is $\mathcal{F}_t$-measurable and that

$$\begin{aligned}\int_G E_{P^X}\left[E_{Q^W}\left[Z|\mathcal{F}_t^W\right]|\mathcal{F}_t^X\right]dQ &= \int_C\int_D E_{P^X}\left[E_{Q^W}\left[\mathbb{1}_A\mathbb{1}_B|\mathcal{F}_t^W\right]|\mathcal{F}_t^X\right]dP^X dQ^W \\ &= \int_C E_{Q^W}\left[\mathbb{1}_A|\mathcal{F}_t^W\right]dQ^W \int_D E_{P^X}\left[\mathbb{1}_B|\mathcal{F}_t^X\right]dP^X \\ &= \int_C \mathbb{1}_A dQ^W \int_D \mathbb{1}_B dP^X \\ &= \int_G Z dQ\end{aligned}$$

□

With this we can price insurance policies by using equation (6.1). To do this we first recall the following functions from Definition 5.1.

$$I_i(t) = \mathbb{1}_{\{X_t=i\}},$$
$$N_{i,j}(t) = \#\{s \in (0,t] : X_{s^-} = i, X_s = j\}.$$

With these we define the policy cash flow as in Definition 5.4 to be

$$A(t) = \sum_{i\in\mathcal{S}} A_i(t) + \sum_{\substack{i,j\in\mathcal{S}\\ i\neq j}} A_{i,j}(t),$$

where

$$A_i(t) = \int_0^t I_i(s)da_i(s) \text{ and } A_{i,j}(t) = \int_0^t a_{i,j}(s)dN_{i,j}(s).$$

In order to price the cash flow we need to calculate the discounted value of the future cash flow at time t. This is the stochastic prospective reserve from Definition 5.5, defined as

$$V^+(t,A) = \frac{1}{v_t}\int_{t^+}^\infty v_s dA(s),$$

where we recall $v_t = B_t^{-1}$ with B_t as the bank account. We may also break this equation down as

$$V^+(t,A) = \sum_{i\in\mathcal{S}} V(t,A_i) + \sum_{\substack{i,j\in\mathcal{S}\\ i\neq j}} V(t,A_{i,j}),$$

where

$$V(t,A_i) = \frac{1}{v_t}\int_0^t v_s I_i(s)da_i(s)$$

and

$$V(t,A_{i,j}) = \frac{1}{v_t}\int_0^t v_s a_{i,j}(s)dN_{i,j}(s)$$

are the stochastic mathematical reserves.

equation (6.1) now allows us to calculate the fair value of a policy at time t by means of the conditional expectation.

Definition 6.1. *We define the prospective reserve in the presence of a financial market analogously to the case in Definition 5.6, but with r and A no longer deterministic. We first define*

$$V_{\mathcal{F}}^{+}(t) = E_Q\left[V^{+}(t)\middle|\mathcal{F}_t\right]. \tag{6.2}$$

By expanding the expectation via Proposition 6.1 and noting that $V_{\mathcal{F}}^{+}(t)$ is a function of X_s for $s \geq t$ we have from the Markov property of X that

$$\begin{aligned} V_{\mathcal{F}}^{+}(t) &= E_{P^X}\left[E_{Q^W}\left[V^{+}(t)\middle|\mathcal{F}_t^W\right]\middle|\mathcal{F}_t^X\right] \\ &= \sum_{i\in\mathcal{S}} I_i(t) E_{P^X}\left[E_{Q^W}\left[V^{+}(t)\middle|\mathcal{F}_t^W\right]\middle|X_t = i\right]. \end{aligned}$$

Which, like before, justifies the definition of

$$V_i^{+}(t) = E_{P^X}\left[E_{Q^W}\left[V^{+}(t)\middle|\mathcal{F}_t^W\right]\middle|X_t = i\right]. \tag{6.3}$$

Like before we may split the prospective reserve into the mathematical reserves.

Definition 6.2. *The mathematical reserves for being in state j over a time interval $J \subset [t,\infty)$ given $X_t = i$ is*

$$V_i(t, A_j, J) = \frac{1}{v_t} E_{P^X}\left[E_{Q^W}\left[\int_J v_s I_j(s) da_j(s)\middle|\mathcal{F}_t^W\right]\middle|X_t = i\right].$$

Similarly we define the mathematical reserve for transitioning from state i to state j over a time interval $J \subset [t,\infty)$ given $X_t = i$ as

$$V_i(t, A_{j,k}, J) = \frac{1}{v_t} E_{P^X}\left[E_{Q^W}\left[\int_J v_s a_{j,k}(s) dN_{j,k}\middle|\mathcal{F}_t^W\right]\middle|X_t = i\right].$$

We will also denote $V_i(t, A_j, (t,\infty))$ and $V_i(t, A_{j,k}, (t,\infty))$ as $V_i(t, A_j)$ and $V_i(t, A_{j,k})$ respectively.

6.2 Explicit Formulas

In order to more explicitly state these values we will need analogues of the formulas in Theorem 5.1 and Theorem 5.2. To develop these results we will have to assume that the a_i's have absolutely continuous paths, we then define $t \mapsto \dot{a}_i(\omega, t)$ path-wise as $t \mapsto (da_i(\omega)/d\lambda)(t)$ where λ is the canonical Lebesgue measure on the real line.

Theorem 6.1. *Let $J \subset [t,\infty) \subset \mathbb{R}_+$ be an interval and let $i, j, k \in \mathcal{S}$. Furthermore we assume that $b, c : \mathbb{R}_+ \times \mathcal{F}_\infty^W \to \mathbb{R}$ are stochastic processes such that b and c are integrable on J. We then have*

$$E_{P^X}\left[E_{Q^W}\left[\int_J I_j(s) c(s) ds\middle|\mathcal{F}_t^W\right]\middle|X_t = i\right] = \int_J p_{i,j}(t,s) E_{Q^W}\left[c(s)|\mathcal{F}_T^W\right] ds,$$

and

$$E_{P^X}\left[E_{Q^W}\left[\int_J b(s)dN_{i,j}(s)\middle|\mathcal{F}_t^W\right]\middle|X_t=i\right]=\int_J E_{Q^W}[b(s)]p_{i,j}(t,s)\mu_{j,k}(s)ds$$

Proof. This proof follows the same structure as the proof of Theorem 5.1 with some minor modifications.

The first equation follows from Fubini's theorem, yielding

$$\begin{aligned}
E_{P^X}\left[E_{Q^W}\left[\int_J I_j(s)c(s)ds\middle|\mathcal{F}_t^W\right]\middle|X_t=i\right] &= E_{P^X}\left[\int_J E_{Q^W}\left[I_j(s)c(s)\middle|\mathcal{F}_t^W\right]ds\middle|X_t=i\right]\\
&= E_{P^X}\left[\int_J I_j(s)E_{Q^W}\left[c(s)\middle|\mathcal{F}_t^W\right]ds\middle|X_t=i\right]\\
&= \int_J E_{P^X}\left[I_j(s)E_{Q^W}\left[c(s)|\mathcal{F}_T^W\right]\middle|X_t=i\right]ds\\
&= \int_J E_{P^X}\left[I_j(s)\middle|X_t=i\right]E_{Q^W}\left[c(s)\middle|\mathcal{F}_T^W\right]ds\\
&= \int_J p_{i,j}(t,s)E_{Q^W}\left[c(s)\middle|\mathcal{F}_T^W\right]ds
\end{aligned}$$

For the second equation we will for convenience denote

$$d(s)=E_{Q^W}[b(s)|\mathcal{F}_t^W].$$

With this we note that for a.s. every path of X we have that $N_{i,j}$ provides a deterministic measure on J allowing us to apply Fubini's theorem which yields

$$E_{Q^W}\left[\int_J b(s)dN_{i,j}(s)\middle|\mathcal{F}_t^W\right]=\int_J E_{Q^W}\left[b(s)\middle|\mathcal{F}_t^W\right]dN_{i,j}(s)=\int_J d(s)dN_{i,j}(s).$$

It then remains to calculate

$$E_{P^X}\left[\int_J d(s)dN_{i,j}(s)\middle|X_t=i\right],$$

where we note that by the monotone class theorem for functions it suffices to calculate the expectation for $d(s)=\mathbb{1}_F(\omega)\mathbb{1}_B(s)$ where $F\subset\mathcal{F}_t^W$ and $B\subset J$.

We then get by Lemma 5.1

$$\begin{aligned}
E_{P^X}\left[\int_J d(s)dN_{i,j}(s)\middle|X_t=i\right] &= E_{P^X}\left[\int_J \mathbb{1}_F(\omega)\mathbb{1}_B(s)dN_{i,j}(s)\middle|X_t=i\right]\\
&= \mathbb{1}_F(\omega)E_{P^X}\left[\int_J \mathbb{1}_B(s)dN_{i,j}(s)\middle|X_t=i\right]\\
&= \mathbb{1}_F(\omega)\int_J \mathbb{1}_B(s)p_{i,j}(t,s)\mu_{j,k}(s)\\
&= \int_J \mathbb{1}_F(\omega)\mathbb{1}_B(s)p_{i,j}(t,s)\mu_{j,k}(s)\\
&= \int_J d(s)p_{i,j}(t,s)\mu_{j,k}(s)
\end{aligned}$$

which proves our result. □

With this we may calculate more explicit formulas for our reserve analogously to Theorem 5.2.

Corollary 6.1. *We have the following equations for our reserves, note here that since the measure of a single point is* 0 *we can drop the* $+$ *in the limit of the integrals.*

$$\begin{aligned} V_i(t, A_j) &= \frac{1}{v_t} E_{P^x}\left[E_{Q^W}\left[\int_{t^+}^{\infty} v_s I_j(s)\dot{a}_j ds \middle| \mathcal{F}_t^W\right] \middle| X_t = i\right] \\ &= \frac{1}{v_t}\int_t^{\infty} p_{i,j}(t,s) E_{Q^W}\left[v_s \dot{a}_j(s) | \mathcal{F}_T^W\right] ds, \end{aligned}$$

and

$$\begin{aligned} V_i(t, A_{j,k}) &= \frac{1}{v_t} E_{P^x}\left[E_{Q^W}\left[\int_{t^+}^{\infty} v_s a_{j,k}(s) dN_{i,j}(s) \middle| \mathcal{F}_t^W\right] \middle| X_t = i\right] \\ &= \frac{1}{v_t}\int_t^{\infty} E_{Q^W}[v_s a_{j,k}(s) | \mathcal{F}_t^W] p_{i,j}(t,s)\mu_{j,k}(s) ds. \end{aligned}$$

By combining these we have the following formula for the prospective reserve.

$$\begin{aligned} V_i^+(t, A) &= \sum_{j\in\mathcal{S}} V_i(t, A_j) + \sum_{\substack{j,k\in\mathcal{S}\\ j\neq k}} V_i(t, A_{j,k}) \\ &= \frac{1}{v_t}\sum_{j\in\mathcal{S}}\int_t^{\infty} p_{i,j}(t,s) E_{Q^W}\left[v_s \dot{a}_j(s) | \mathcal{F}_t^W\right] ds \\ &+ \frac{1}{v_t}\sum_{\substack{j,k\in\mathcal{S}\\ j\neq k}}\int_t^{\infty} E_{Q^W}[v_s a_{j,k}(s) | \mathcal{F}_t^W] p_{i,j}(t,s)\mu_{j,k}(s) ds. \end{aligned}$$

The most problematic term is the conditional expectations $E_{Q^W}\left[v_s \dot{a}_j(s) | \mathcal{F}_t^W\right]$ and $E_{Q^W}[v_s a_{j,k}(s) | \mathcal{F}_t^W]$. In many cases, some which will be covered in upcoming examples, these can be written explicitly as Itô process. From this we refine the formulas in Corollary 6.1 by the following results.

Lemma 6.1 (Leibniz Integral Rule)**.** *Let* $Y(t,s)$ *be an Itô process on the form*

$$Y(t,s) = Y(0,s) + \int_0^t a(u,s)du + \int_0^t b(u,s)dW_u,$$

such that the volatility b *satisfies* $\sup_{u,s\leq S}|b(u,s)| < \infty$ *a.s. and the drift* a *satisfies* $\int_0^S\int_0^S |a(u,s)|duds < \infty$*. We then have that* $\mathcal{I}(t,S) = \int_t^S Y(t,s)ds$ *is an Itô process on the form*

$$\mathcal{I}(t,S) = \mathcal{I}(0,S) + \int_0^t \left(\int_u^S a(u,s)ds - Y(u,u)\right) du + \int_0^t \left(\int_u^S b(u,s)ds\right) dW_u.$$

Proof. By applying the deterministic and stochastic version of Fubini's theorem (Theorem 2.6 and Theorem 2.24) we get

$$\begin{aligned}\int_0^t \int_u^t a(u,s)dsdu &= \int_0^t \int_0^t \mathbb{1}_{u\leq s}(u,s)a(u,s)dsdu \\ &= \int_0^t \int_0^t \mathbb{1}_{u\leq s}(u,s)a(u,s)duds \\ &= \int_0^t \int_0^s a(u,s)duds,\end{aligned}$$

and similarly

$$\begin{aligned}\int_0^t \int_u^t b(u,s)dsdW_u &= \int_0^t \int_0^t \mathbb{1}_{u\leq s}(u,s)b(u,s)dsdW_u \\ &= \int_0^t \int_0^t \mathbb{1}_{u\leq s}(u,s)b(u,s)dW_uds \\ &= \int_0^t \int_0^s b(u,s)dW_uds.\end{aligned}$$

Applying these equations we get

$$\begin{aligned}&\int_t^S Y(t,s)ds \\ &= \int_t^S \left(Y(0,s) + \int_0^t a(u,s)du + \int_0^t b(u,s)dW_u\right)ds \\ &= \int_t^S Y(0,s)ds + \int_t^S \int_0^t a(u,s)duds + \int_t^S \int_0^t b(u,s)dW_uds \\ &= \int_t^S Y(0,s)ds + \int_0^t \int_t^S a(u,s)dsdu + \int_0^t \int_t^S b(u,s)dsdW_u \\ &= \int_0^S Y(0,s)ds + \int_0^t \int_u^S a(u,s)dsdu + \int_0^t \int_u^S b(u,s)dsdW_u \\ &- \int_0^t Y(0,s)ds - \int_0^t \int_u^t a(u,s)dsdu - \int_0^t \int_u^t b(u,s)dsdW_u \\ &= \int_0^S Y(0,s)ds + \int_0^t \int_u^S a(u,s)dsdu + \int_0^t \int_u^S b(u,s)dsdW_u \\ &- \int_0^t Y(0,s)ds - \int_0^t \int_0^s a(u,s)duds - \int_0^t \int_0^s b(u,s)dW_uds \\ &= \int_0^S Y(0,s)ds + \int_0^t \int_u^S a(u,s)dsdu + \int_0^t \int_u^S b(u,s)dsdW_u \\ &- \int_0^t \left(Y(0,s) + \int_0^s a(u,s)du + \int_0^s b(u,s)dW_u\right)ds \\ &= \mathcal{I}(0,S) + \int_0^t \left(\int_u^S a(u,s)ds - Y(u,u)\right)du + \int_0^t \left(\int_u^S b(u,s)ds\right)dW_u\end{aligned}$$

□

Corollary 6.2. *For all $j,k \in \mathcal{S}$ we define $Y_j(t,s) = \frac{1}{v_t} E_{Q^W}[v_s \dot{a}_j(s) | \mathcal{F}_t^W]$ and $Y_{j,k}(t,s) = \frac{1}{v_t} E_{Q^W}[v_s \dot{a}_{j,k}(s) | \mathcal{F}_t^W]$ and assume that these are Itô processes with t-dynamics*

$$dY_j(t,s) = a_j(t,s)dt + b_j(t,s)dW_t,$$
$$dY_{j,k}(t,s) = a_{j,k}(t,s)dt + b_{j,k}(t,s)dW_t,$$

satisfying the conditions of Lemma 6.1. We then have that

$$\begin{aligned} V_i(t,A_j) = V_i(0,A_j) &- \mathbb{1}(i=j)\int_0^t Y_j(u,u)du \\ &- \int_0^t \int_u^\infty \sum_{k\in\mathcal{S}} \mu_{i,k}(u) p_{k,j}(u,s) Y_j(u,s) ds du \\ &+ \int_0^t \int_u^\infty p_{i,j}(u,s) a_j(u,s) ds du \\ &+ \int_0^t \int_u^\infty p_{i,j}(u,s) b_j(u,s) ds dW_u, \end{aligned}$$

$$\begin{aligned} V_i(t,A_{j,k}) = V_i(0,A_{j,k}) &- \mathbb{1}(i=j)\int_0^t \mu_{j,k}(u) Y_{j,k}(u,u)du \\ &- \int_0^t \int_u^\infty \sum_{k\in\mathcal{S}} \mu_{i,k}(u) p_{k,j}(u,s) Y_{j,k}(u,s) ds du \\ &+ \int_0^t \int_u^\infty p_{i,j}(u,s) \mu_{j,k}(s) a_{j,k}(u,s) ds du \\ &+ \int_0^t \int_u^\infty p_{i,j}(u,s) \mu_{j,k}(s) b_{j,k}(u,s) ds dW_u. \end{aligned}$$

Proof. We will prove this statement for

$$V_i(t,A_j) = \int_t^S p_{i,j}(t,s) Y(t,s) ds,$$

$$V_i(t,A_j) = \int_t^S p_{i,j}(t,s) \mu_{j,k}(s) Y(t,s) ds,$$

note that this result then easily extends to $S = \infty$ by the dominated convergence theorem.

We first see that this assumption yields

$$V_i(t,A_j) = \int_t^S g_j(t,s,Y_j(t,s)) ds,$$
$$V_i(t,A_{j,k}) = \int_t^S g_{j,k}(t,s,Y_{j,k}(t,s)) ds,$$

where

$$g_j(t,s,x) = p_{i,j}(t,s)x,$$
$$g_{j,k}(t,s,x) = p_{i,j}(t,s)\mu_{j,k}(s)x.$$

Applying Theorem 2.25, Itô's lemma, we get that the $g_j(t,s,Y_j(t,s))$'s and $g_{j,k}(t,s,Y_{j,k}(t,s))$'s are Itô processes in t on the following form.

$$g_j(t,s,Y_j(t,s)) = g_j(0,s,Y_j(0,s)) + \int_0^t \frac{\partial p_{i,j}(u,s)}{\partial u} Y_j(u,s)du + \int_0^t p_{i,j}(u,s)a_j(u,s)du + \int_0^t p_{i,j}(u,s)b_j(u,s)dW_u, \quad (6.4)$$

$$g_{j,k}(t,s,Y^s_{j,k}(t)) = g_{j,k}(0,s,Y^s_{j,k}(0)) + \int_0^t \frac{\partial p_{i,j}(u,s)}{\partial u} \mu_{j,k}(s)Y_{j,k}(u,s)du + \int_0^t p_{i,j}(u,s)\mu_{j,k}(s)a_{j,k}(u,s)du + \int_0^t p_{i,j}(u,s)\mu_{j,k}(s)b_{j,k}(u,s)du. \quad (6.5)$$

Note that since both $p_{i,j}(u,s)$ and $\mu_{j,k}(s)$ are continuous they are bounded on $[0,S]\times[0,S]$, which means that the g's still satisfy the conditions in Lemma 6.1.

From Theorem 2.28, the backwards Kolmogorov equation we also have that

$$\frac{\partial p_{i,j}(u,s)}{\partial u} = -\sum_{k\in\mathcal{S}} \mu_{i,k}(u)p_{k,j}(u,s).$$

By plugging this into equation (6.4) and (6.5) along with applying Lemma 6.1 we get

$$\begin{aligned} V_i(t,A_j) = V_i(0,A_j) &- \int_0^t g_j(u,u,Y_j(u,u))du \\ &- \int_0^t \int_u^S \sum_{k\in\mathcal{S}} \mu_{i,k}(u)p_{k,j}(u,s)Y_j(u,s)dsdu \\ &+ \int_0^t \int_u^S p_{i,j}(u,s)a_j(u,s)dsdu \\ &+ \int_0^t \int_u^S p_{i,j}(u,s)b_j(u,s)dsdW_u, \end{aligned}$$

$$\begin{aligned} V_i(t,A_{j,k}) = V_i(0,A_{j,k}) &- \int_0^t g_{j,k}(u,u,Y_{j,k}(u,u))du \\ &- \int_0^t \int_u^S \sum_{k\in\mathcal{S}} \mu_{i,k}(u)p_{k,j}(u,s)Y_{j,k}(u,s)dsdu \\ &+ \int_0^t \int_u^S p_{i,j}(u,s)\mu_{j,k}(s)a_{j,k}(u,s)dsdu \\ &+ \int_0^t \int_u^S p_{i,j}(u,s)\mu_{j,k}(s)b_{j,k}(u,s)dsdW_u. \end{aligned}$$

Finally we note that $p_{i,j}(u,u) = \mathbb{1}(i=j)$, this implies

$$\begin{aligned} g_{j,k}(u,u,Y_{j,k}(u,u)) &= p_{i,j}(u,u)\mu_{j,k}(u)Y_{j,k}(u,u) \\ &= \mathbb{1}(i=j)\mu_{j,k}(u)Y_{j,k}(u,u), \end{aligned}$$

$$\begin{aligned} g_j(u,u,Y_j(u,u)) &= p_{i,j}(u,u)Y_j(u,u) \\ &= \mathbb{1}(i=j)Y_j(u,u), \end{aligned}$$

which yields the desired formula. □

With this result we will, in many cases, get that the mathematical reserves are Itô processes. This means that they are both continuous and progressive, furthermore, it provides another explicit formula that could be used for computations.

6.3 Applications to Continuous Time CD Ladders with Minimum Guarantee

Zero-coupon bonds are generally offered with better rates than what an investment in a bank account would yield while also locking in the interest rate and thus protecting your investment from market volatility. This, however, comes at the cost of liquidity as the investment cannot be readily accessed before the maturity of the bond.

We will in this section consider a "CD ladder", where CD stands for certificate of deposit, another name for zero-coupon bonds. A CD ladder is an investment strategy where the investor purchases zero-coupon bonds with equal time to maturity at regularly spaced intervals. For example, one could purchase zero-coupon bonds with a five year time to maturity quarterly. This means that every quarter the investor would receive a set amount of money to be spent or reinvested into the ladder. This allows the cash to be more readily available and thus maintaining liquidity while still providing the regular benefits of investing in zero-coupon bonds.

We then consider a pension scheme based on a continuous version of such a CD ladder, with payments starting at some time S_0 and ending upon death. We assume that at any time $s \geq S_0$ the insured purchases zero-coupon bonds with maturity at time $s+\Delta$ at a deterministic rate of $K_{s+\Delta}$ from the insurance company. We further assume that at the same time the insured will receive a deterministic payout at a rate of K_s, either from previously purchased bonds or otherwise financed as part of the pension contract, this means the insured will receive payments at a rate of $K_s - K_{s+\Delta}P(s,s+\Delta)$.

Lastly we consider a deterministic minimum payout guarantee at time s, denoted $\widehat{K}_s$, which means the insurance company would have to pay an additional amount at a rate equal to

$$\left(K_{s+\Delta}P(s,s+\Delta) - K_s + \widehat{K}_s\right)^+,$$

where $(\cdot)^+ = max(0,\cdot)$. This brings the total rate of the payout received at time s to

$$K_s - K_{s+\Delta}P(s,s+\Delta) + \left(K_{s+\Delta}P(s,s+\Delta) - K_s + \widehat{K}_s\right)^+ \geq \widehat{K}_s.$$

When the insured dies they will leave a collection of bonds, these can then either be paid out to the beneficiaries of the insured, either over a time period of Δ or settled with a lump sum. Another way to settle this is by letting the bonds pass to the insurance company, effectively annulling them. We will for now consider the latter for simplicity, but will tackle generalizations later.

We will here consider two possible states, living, denoted by $*$, and dead, denoted by $\dagger$. We assume that the insured cannot return from the dead state to the living state. Furthermore, in order to compute the value of these policies, we will use forward measures. As such we will assume that our ELMM Q^W is an EMM and that $\sigma(t,T)$, the volatility process of the forward rates $f(t,T)$, is deterministic. With this established we move on to define our policy functions for $s \geq S_0$ as

$$\begin{aligned}
\dot{a}_\dagger(s) &= a_{*,\dagger}(s) = a_{\dagger,*}(s) = 0, \\
\dot{a}_*(s) &= K_s - K_{s+\Delta}P(s,s+\Delta) + \left(K_{s+\Delta}P(s,s+\Delta) - K_s + \widehat{K}_s\right)^+.
\end{aligned}$$

We note that the only non-zero mathematical reserve is the one corresponding to being in state $*$, the living state. From Corollary 6.1 we get that the fair price of such an agreement at time t, assuming the insured is alive, must be

$$\begin{aligned}
V_*^+(t,A) =& V_*(t,A_*) \\
=& \int_t^\infty p_{*,*}(t,s)\frac{1}{v_t}E_{Q^W}\left[v_s\dot{a}_*(s)|\mathcal{F}_t^W\right]ds \\
=& \int_t^\infty p_{*,*}(t,s)\frac{1}{v_t}E_{Q^W}\left[v_sK_s|\mathcal{F}_t^W\right]ds \\
&- \int_t^\infty p_{*,*}(t,s)\frac{1}{v_t}E_{Q^W}\left[v_sK_{s+\Delta}P(s,s+\Delta)|\mathcal{F}_t^W\right]ds \\
&+ \int_t^\infty p_{*,*}(t,s)\frac{1}{v_t}E_{Q^W}\left[v_s\left(K_{s+\Delta}P(s,s+\Delta) - K_s + \widehat{K}_s\right)^+|\mathcal{F}_t^W\right]ds.
\end{aligned}$$

To compute an explicit formula we first note that by Theorem 2.29 or Example 2.7 we have an explicit representation of the transition probability by

$$p_{*,*}(t,s) = exp\left(-\int_s^t \mu_{*,\dagger}(u)du\right). \tag{6.6}$$

However, to keep notation compact we will still write the probability in the less explicit form as $p_{*,*}(t,s)$. Secondly, by our definition of zero-arbitrage pricing or by Proposition 4.2 we have

$$\frac{1}{v_t}E_{Q^W}\left[v_sK_s\middle|\mathcal{F}_t^W\right] = P(t,s)K_s,$$

and by our assumption that Q^W is an EMM, which implies that the process $u \mapsto v_uP(u,s+\Delta)$ is a martingale under Q^W, we get

$$\frac{1}{v_t}E_{Q^W}\left[v_sK_{s+\Delta}P(s,s+\Delta)\middle|\mathcal{F}_t^W\right] = P(t,s+\Delta)K_{s+\Delta}.$$

Lastly, we note that, assuming $K_{s+\Delta} > 0$, we have

$$\begin{aligned}
&\frac{1}{v_t} E_{Q^W}\left[v_s \left(K_{s+\Delta} P(s, s+\Delta) - K_s + \widehat{K}_s \right)^+ \middle| \mathcal{F}_t^W \right] \\
&= K_{s+\Delta} \frac{1}{v_t} E_{Q^W}\left[v_s \left(P(s, s+\Delta) - \frac{K_s - \widehat{K}_s}{K_{s+\Delta}} \right)^+ \middle| \mathcal{F}_t^W \right],
\end{aligned}$$

which is the value at time t of $K_{s+\Delta}$ units of a European call on $P(s, s+\Delta)$ with strike price $(K_s - \widehat{K}_s)/K_{s+\Delta}$. By Theorem 4.2 this equals

$$\begin{aligned}
&K_{s+\Delta} P(t, s+\Delta) \varphi^+\Big(t, s, P(t, s+\Delta), P(t, s)\Big) \\
&- (K_s - \widehat{K}_s) P(t, s) \varphi^-\Big(t, s, P(t, s+\Delta), P(t, s)\Big),
\end{aligned}$$

where we have defined

$$\varphi^{\pm}(t, s, x, y) = \Phi\left(\frac{log\left(\frac{K_{s+\Delta} x}{(K_s - \widehat{K}_s) y} \right) \pm \frac{1}{2} \int_t^s |v_{s,s+\Delta}(u)|^2 du}{\sqrt{\int_t^s |v_{s,s+\Delta}(u)|^2 du}} \right),$$

in order to keep our notation compact. We may also recall the definition $v_{s,s+\Delta}(u) = \int_s^{s+\Delta} \sigma(u, v) dv$, and that Φ is the standard normal distribution function.

Combining these equations we get an explicit formula for V_*^+.

$$\begin{aligned}
V_*^+(t, A) =& V_*(t, A_*) \\
=& \int_t^\infty p_{*,*}(t, s) P(t, s) K_s ds \\
&- \int_t^\infty p_{*,*}(t, s) P(t, s+\Delta) K_{s+\Delta} ds \\
&+ \int_t^\infty p_{*,*}(t, s) K_{s+\Delta} P(t, s+\Delta) \varphi^+\Big(t, s, P(t, s+\Delta), P(t, s)\Big) ds \\
&- \int_t^\infty p_{*,*}(t, s) (K_s - \widehat{K}_s) P(t, s) \varphi^-\Big(t, s, P(t, s+\Delta), P(t, s)\Big) ds.
\end{aligned} \tag{6.7}$$

In addition to the explicit formula we also note that our calculations have shown that $Y(t, s) = \frac{1}{v_t} E_{Q^W}[v_s \dot{a}_j(s) | \mathcal{F}_t^W]$ is made up from twice differentiable functions of Itô processes. This means that Y itself is an Itô process by Itô's lemma (Theorem 2.25). Applying Corollary 6.2 we then see that the prospective reserve must be an Itô process which guarantees that it is continuous and progressive in t.

As previously mentioned we can also consider a death benefit based on the remaining bonds. If the insured dies at time s there will be a collection of zero-coupon bonds that would grant payouts at a rate of K_u for $u \in [s, s+\Delta]$. The fair price of such a cash flow at time s is therefore

$$E_{Q^W}\left[\frac{1}{v_s} \int_s^{s+\Delta} v_u K_u du \middle| \mathcal{F}_s^W \right] = \int_s^{s+\Delta} P(s, u) K_u du.$$

With this we see that the insurance company could reasonably offer this sum as a death benefit in this pension scheme, essentially buying back all bonds at the time of death. With this adjustment we get $a_{*,\dagger}(s) = \int_s^{s+\Delta} P(s,u)K_u du$, which yields an explicit formula for the reserve at time t.

$$
\begin{aligned}
&V_*(t, A_{*,\dagger}) \\
&= \frac{1}{v_t} \int_t^\infty E_{Q^W} \left[v_s a_{*,\dagger}(s) | \mathcal{F}_t^W \right] p_{*,*}(t,s) \mu_{*,\dagger}(s) ds \\
&= \frac{1}{v_t} \int_t^\infty E_{Q^W} \left[v_s E_{Q^W} \left[\frac{1}{v_s} \int_s^{s+\Delta} v_u K_u du \middle| \mathcal{F}_s^W \right] \middle| \mathcal{F}_t^W \right] p_{*,*}(t,s) \mu_{*,\dagger}(s) ds \\
&= \frac{1}{v_t} \int_t^\infty E_{Q^W} \left[\int_s^{s+\Delta} v_u K_u du \middle| \mathcal{F}_t^W \right] p_{*,*}(t,s) \mu_{*,\dagger}(s) ds \\
&= \int_t^\infty \left(\int_s^{s+\Delta} P(t,u) K_u du \right) p_{*,*}(t,s) \mu_{*,\dagger}(s) ds.
\end{aligned}
\tag{6.8}
$$

By noting that $V_*(t, A_*)$ is unaffected by the change in $a_{*,\dagger}$ we get an explicit formula for the new prospective reserve by combining equation (6.7) and (6.8) to get

$$
\begin{aligned}
V_*^+(t, A) =& V_*(t, A_*) + V_*(t, A_{*,\dagger}) \\
=& \int_t^\infty p_{*,*}(t,s) P(t,s) K_s ds \\
&- \int_t^\infty p_{*,*}(t,s) P(t, s+\Delta) K_{s+\Delta} ds \\
&+ \int_t^\infty p_{*,*}(t,s) K_{s+\Delta} P(t, s+\Delta) \varphi^+ \Big(t, s, P(t, s+\Delta), P(t,s) \Big) ds \\
&- \int_t^\infty p_{*,*}(t,s) (K_s - \widehat{K}_s) P(t,s) \varphi^- \Big(t, s, P(t, s+\Delta), P(t,s) \Big) ds \\
&+ \int_t^\infty \left(\int_s^{s+\Delta} P(t,u) K_u du \right) p_{*,*}(t,s) \mu_{*,\dagger}(s) ds.
\end{aligned}
\tag{6.9}
$$

Chapter 7

Implementation of CD Ladder Insurance Pricing under HJM

In this chapter we will give an example of how to implement the pricing methods developed in earlier chapters for the pension scheme based on a continuous time CD ladder considered in Section 6.3. We will here consider the case without a death benefit, i.e. the prospective reserve given by equation (6.7).

7.1 Model Specification

For the forward rate we will make the following assumptions, here a, b and c will be real parameters.

- The initial forward curve is on the form

$$f(0,T) = b + (f(0,0) - b)\, e^{-aT}.$$

  We will refer to b as the reversion level and a as the speed of mean reversion.

- The drift term $\alpha(t,T)$ is assumed to equal 0 under P.

- The volatility term is taken to be one-dimensional and on the form

$$\sigma(t,T) = ce^{-a(T-t)}.$$

  We will refer to c as the instantaneous volatility.

With these assumptions we get

$$f(t,T) = b + (f(0,0) - b)\, e^{-aT} + \int_0^t ce^{-a(T-s)} dW_s. \tag{7.1}$$

For convenience we will define

$$\psi(t) = (f(0,0) - b) + \int_0^t ce^{as} dW_s,$$

which implies $f(t,T) = b + e^{-aT}\psi(t)$.

Remark 7. *We may note that this implies by inserting $T = t$ into equation (7.1) we get*

$$r_t = b + (r_0 - b)\, e^{-at} + \int_0^t ce^{-a(t-s)} dW_s.$$

From this we get

$$\begin{aligned}
\int_0^t a(b - r_s)ds &= -a\,(r_0 - b) \int_0^t e^{-as} ds - a \int_0^t \int_0^s ce^{-a(s-u)} dW_u ds \\
&= (r_0 - b)\, e^{-at} - (r_0 - b) - a \int_0^t \int_0^s ce^{-a(s-u)} dW_u ds \\
&= (r_0 - b)\, e^{-at} + b - r_0 - a \int_0^t \int_u^t ce^{-a(s-u)} ds dW_u \\
&= (r_0 - b)\, e^{-at} + b - r_0 + \int_0^t ce^{-a(t-u)} dW_u - \int_0^t c dW_u \\
&= r_t - r_0 - \int_0^t c dW_u.
\end{aligned}$$

This means that r must satisfy the stochastic differential equation

$$dr_t = a(b - r_t)dt + c dW_t,$$

which is a common way to define the famous Vašíček short-rate model.

As for the pension scheme specifications we will assume that the payout function K_s and the minimum guarantee $\widehat{K}_s$ are both at a constant value adjusted for inflation. The rate of inflation, ρ, will be assumed constant. To state these assumptions precisely we have

$$\begin{aligned}
K_s &= K_0 e^{\rho s}, \\
\widehat{K}_s &= \widehat{K}_0 e^{\rho s}.
\end{aligned}$$

We will also assume that K_0 is on the form

$$K_0 = \frac{\widehat{K}_0}{1 - e^{(\rho - b)\Delta}}.$$

Note that if the forward rates were constant and equal to the reversion level b this assumption would yield

$$K_s - K_{s+\Delta} P(s, s+\Delta) = \widehat{K}_s.$$

This value of K_0 therefore assures that the CD-ladder would be self-financing in this case.

For the transition rates we will be using the mortality basis known as K2013 [4]. K2013 is a letter published by Finanstilsynet, the Norwegian Supervisory Authority of Norway, detailing mortality rates which Norwegian insurance companies have to comply with. A more exact description is given in the next section, see equation (7.4).

7.2 Formulas for Exact Computation

These model assumption also allows us to more exactly calculate some aspects of our pricing equations, let us first consider the function

$$\varphi^{\pm}\Big(t,s,P(t,s{+}\Delta),P(t,s)\Big)=\Phi\left(\frac{log\left(\frac{K_{s+\Delta}P(t,s+\Delta)}{(K_s-\widehat{K}_s)P(t,s)}\right)\pm\frac{1}{2}\int_t^s|v_{s,s+\Delta}(u)|^2du}{\sqrt{\int_t^s|v_{s,s+\Delta}(u)|^2du}}\right).$$

We note from Definition 3.2 that $P(t,T)=exp\left(-\int_t^T f(t,s)ds\right)$, which means we get

$$\begin{aligned}log\left(\frac{P(t,s+\Delta)}{P(t,s)}\right)&=log\left(\frac{exp\left(-\int_t^{s+\Delta}f(t,u)du\right)}{exp\left(-\int_t^s f(t,u)du\right)}\right)\\&=-\int_s^{s+\Delta}f(t,u)du\\&=-\int_s^{s+\Delta}\left(b+e^{-au}\psi(t)\right)du\\&=-b\Delta-\left(e^{-as}-e^{-a(s+\Delta)}\right)\frac{\psi(t)}{a}.\end{aligned}\tag{7.2}$$

Using the same calculations we also get an explicit formula for $P(t,s)$ by

$$P(t,s)=exp\left(-b(s-t)-\left(e^{-at}-e^{-as}\right)\frac{\psi(t)}{a}\right).$$

Likewise, we can compute other parts of φ, from basic calculus we get

$$\begin{aligned}\int_t^s|v_{s,s+\Delta}(u)|^2du&=\int_t^s\left(\int_s^{s+\Delta}\sigma(u,v)dv\right)^2du\\&=\int_t^s\left(\int_s^{s+\Delta}ce^{-av}e^{au}dv\right)^2du\\&=c^2\int_t^s e^{2au}du\left(\int_s^{s+\Delta}e^{-av}dv\right)^2\\&=\frac{c^2}{2a^3}\left(e^{2as}-e^{2at}\right)\left(e^{-as}-e^{-a(s+\Delta)}\right)^2\\&=\frac{c^2}{2a^3}\left(1-e^{-2a(s-t)}\right)\left(1-e^{-a\Delta}\right)^2.\end{aligned}\tag{7.3}$$

For our survival probabilities the K2013 mortality for a person aged x in calendar year $u>2013$ is given by

$$\mu(x,u)=\mu(x,2013)\left(1+\frac{w(x)}{100}\right)^{u-2013}.\tag{7.4}$$

This is the mortality at age x in 2013 scaled down with a factor based on $w(x)$, the mortality reduction. These values are specified by

$$\text{Men: } w(x) = min(2.671548 - 0.17248x + 0.001485x^2, 0)$$
$$\text{Women: } w(x) = min(1.287968 - 0.10109x + 0.000814x^2, 0),$$

and

$$\text{Men: } 1000\,\mu(x, 2013) = 0.241752 + 0.004536 \cdot 10^{(0.051x)}$$
$$\text{Women: } 1000\,\mu(x, 2013) = 0.085411 + 0.003114 \cdot 10^{(0.051x)}.$$

We recall equation (6.6) which allows for computation of the survival probabilities using the K2013 parameters.

$$p_{*,*}(t, s) = exp\left(-\int_s^t \mu_{*,\dagger}(u)du\right).$$

Finally we will need a way of simulating $\psi(t)$, which is the only non-deterministic part of our equations. We note from Theorem 2.23 that $\psi(t)$ is normally distributed with expectation $f(0,0) - b$ and a variation which can be computed by Theorem 2.25, Itô's lemma, to be

$$\begin{aligned} Var[\psi(t)] &= \int_0^t (ce^{as})^2\, ds \\ &= c^2 \int_0^t e^{2as} ds \\ &= \frac{c^2}{2a}\left(e^{2at} - 1\right). \end{aligned} \tag{7.5}$$

For our computations we will give time in units of years with $t = 0$ in the current year of 2021. The parameters will also be set to the following values.

- Reversion level: $b = 5\%$.
- Current short-rate: $f(0,0) = r_0 = 3\%$.
- Instantaneous volatility: $c = 2\%$.
- Speed of mean reversion: $a = 50\%$.
- Initial minimum payout guarantee: $\widehat{K}_0 = 1$, note that if we considered the reserves as a function of $\widehat{K}_0$ we would get $V^+(t, \widehat{K}_0) = \widehat{K}_0 V^+(t, 1)$. This means that by setting the initial payout to 1 we effectively consider the payout in units of $\widehat{K}_0$.
- As a consequence of $\widehat{K}_0 = 1$ we get that $K_0 = 7.179162$.
- Rate of inflation: 2%.
- Time to maturity of bonds: $\Delta = 5$.

We also assume that the insured is a 39 year old woman that retires at age 67, $t = 28$. Payments will therefore start at $t = 28$, and for convention we will assume the chance of surviving past an age of 114, $t = 75$, to be negligible. This

convention is justified by the K2013 probability of the insured to reach this age, which equals

$$p_{*,*}(0,75) \approx 5.6 \cdot 10^{-7}.$$

These choices imply that the prospective reserve must be on the form

$$V_*^+(t,A) = \int_{\max(t,28)}^{75} I(t,s)ds,$$

where the integrand I is the regular integrand from equation (6.7),

$$\begin{aligned} I(t,s) = \; & p_{*,*}(t,s)P(t,s)K_s \\ & - p_{*,*}(t,s)P(t,s+\Delta)K_{s+\Delta} \\ & + p_{*,*}(t,s)K_{s+\Delta}P(t,s+\Delta)\varphi^+\Big(t,s,P(t,s+\Delta),P(t,s)\Big) \\ & - p_{*,*}(t,s)(K_s - \widehat{K}_s)P(t,s)\varphi^-\Big(t,s,P(t,s+\Delta),P(t,s)\Big). \end{aligned}$$

We finally note that the only stochastic term present in these equations is the term $\psi(t)$, so $V^+(t,A)$ could be considered a function of t and $\psi(t)$. Since ψ and r are directly related by $r_t = b + e^{-at}\psi(t)$ we could, and will, equivalently consider $V^+(t,A)$ as a function of t and r_t.

7.3 Results and Comments

Computing the reserve and dropping the notation of A in favour of r_t we get

$$V^+(0,A) = V^+(0,r_0) \approx 8.0.$$

However we may note that since we chose $\widehat{K}_0 = 1$ we are effectively getting the value of the reserve in units of the minimum guarantee, if we chose another value we would get

$$V^+(0,r_0) \approx 8.0 \cdot \widehat{K}_0.$$

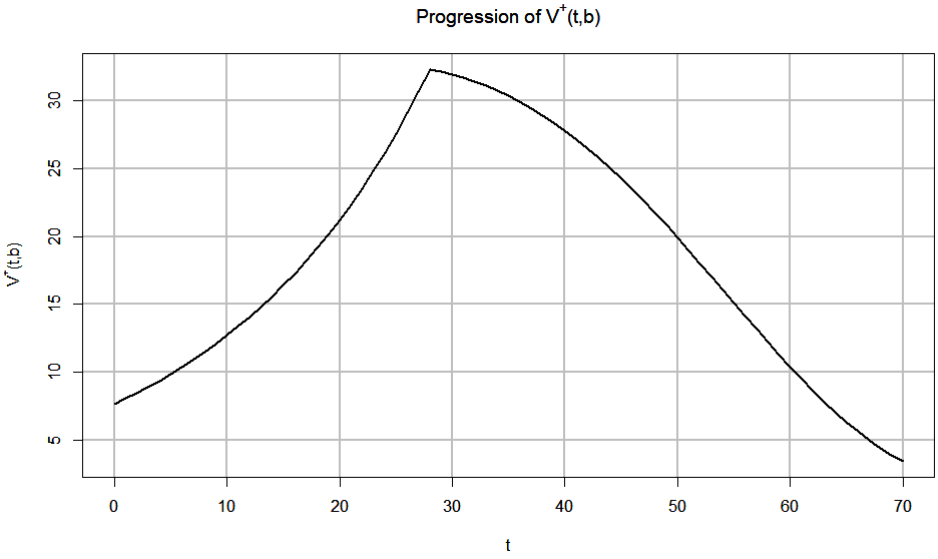
Our framework also allows to easily calculate the value of the reserve at $t > 0$. Let us for example consider the value of the reserve at retirement age, $t = 28$, under the assumption that $r_{28} = b$. direct computation then yields

$$V^+(28,b) \approx 32.3.$$

We could then, for example, consider the inflation-adjusted value of this to get a present value of

$$V^+(28,b)e^{-28\rho} \approx 32.3e^{-28\rho} \approx 18.5.$$

If we are interested in seeing how $V^+(t,r_t)$ develops in time we could get an impression by calculating $V^+(t,b)$ for an array of time points as shown in the following graph.



We see the reserves increase as we approach $t = 28$ as the payments get closer in time and thus less affected by the devaluation of the interest rate. As we pass $t = 28$, however, we lose out on more and more of the payments which causes the drop in value.

Of course, r_t is not known, but we can calculate the reserves for a reasonable range of values.

To select a range we note that

$$r_t = b + e^{-at} \int_0^t c e^{au} dW_u$$

is normally distributed with mean

$$b + e^{-at}(r_0 - b)$$

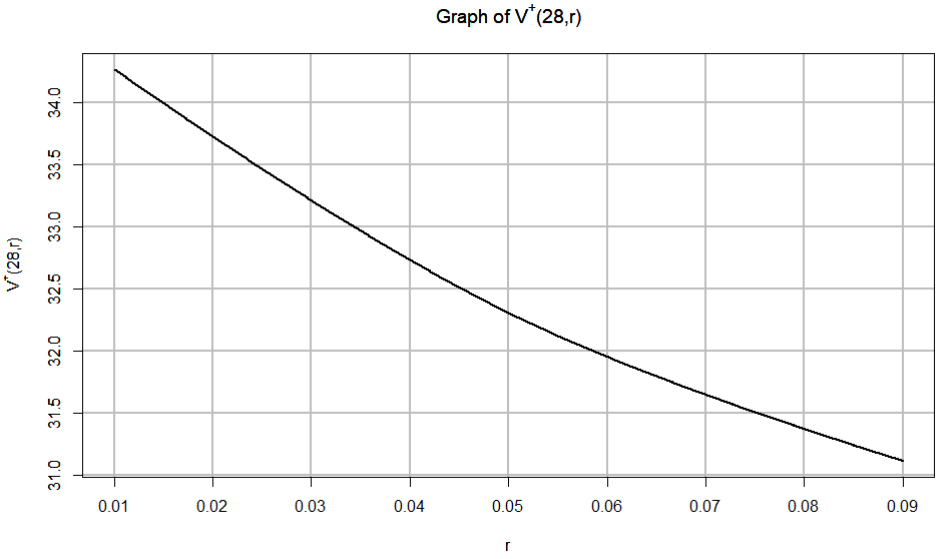
and a variance of

$$c^2 e^{-2at} \int_0^t e^{2au} du = \frac{c^2}{2a} \left(1 - e^{-2at}\right).$$

This means that when $t \to \infty$ we get an expectation of b and a variance of $\frac{c^2}{2a}$. We could then choose our range to be the long-run 95% confidence interval of r_t, which equals

$$\left[b - 1.96\sqrt{\frac{c^2}{2a}}, b + 1.96\sqrt{\frac{c^2}{2a}} \right] \approx [b - 4\%, B + 4\%] = [1\%, 9\%].$$

Using this range we get the following graph of values.



We see that V^+ decreases in r, this is expected as a high interest rate devalues future payouts.

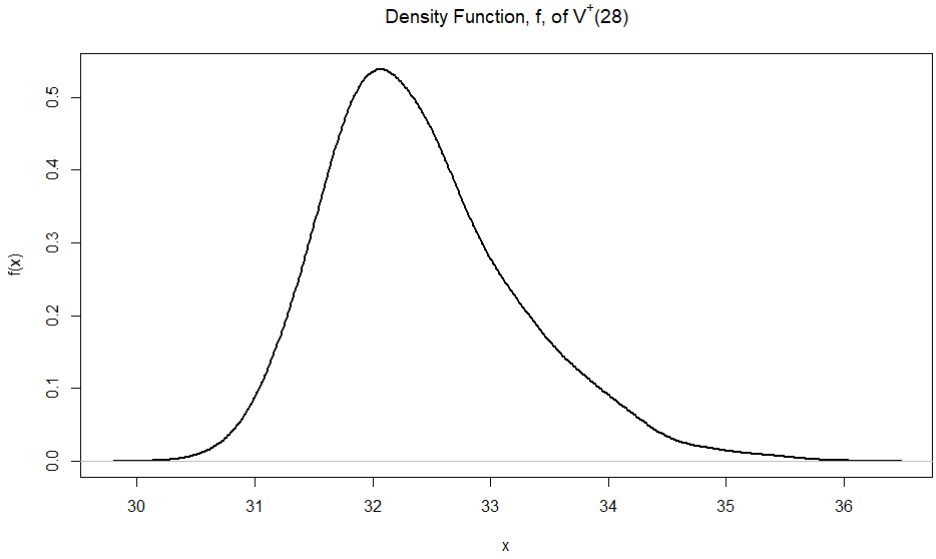
Furthermore, since r_0 is known we can also calculate the distribution $V^+(28)$, and therefore several statistical properties, of $V^+(28)$ numerically. To produce the following graph we used 2000 simulations of r_{28} and a Gaussian kernel density estimation with the standard deviation of the kernels equal to 0.226. More explicitly, the distribution is approximated by the graph of

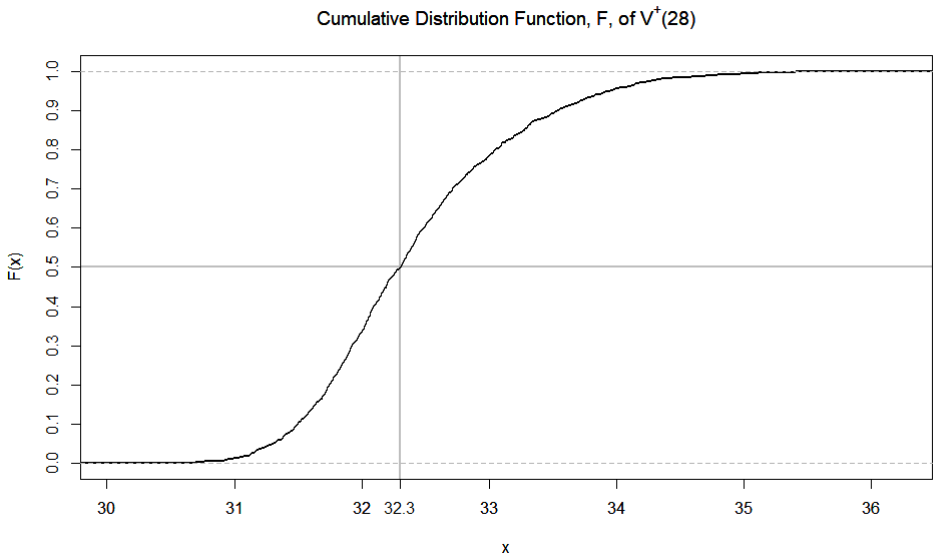
$$D(x) = \sum_{i=1}^{2000} \phi(x, \bar{V}_i, 0.226),$$

where $\{\bar{V}_i\}$ is the collection of simulated values of $V^+(28)$ and

$$\phi(x, \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

is the density function of a normal variable with mean μ and standard deviation σ. From these simulations we also get the empirical cumulative distribution function as shown.





We may note the marked median in the cumulative distribution graph at $x = 32.3 \approx V^+(28, b)$. Since r_{28} is normally distributed its mean and median coincide, this along with $V^+(28, r)$ being monotone in r forces the median of $V^+(28, r)$ to $V^+(28, E[r]) \approx V^+(28, b)$. The distribution of $V^+(28, r)$, however, is not symmetric as it has a slightly heavier tail for large values. To better understand the source of this we consider the payout rate of the policy

$$K_s - K_{s+\Delta}P(s, s+\Delta) + \left(K_{s+\Delta}P(s, s+\Delta) - K_s + \widehat{K}_s\right)^+.$$

We see that while high interest rates reduce the value of future payment it also increases the payout function, therefore somewhat lessening the impact that high interest rates have on V^+. In the case of low interest rates however, the payout function is bounded by the minimum guarantee which allows V^+ to rise unrestrained as r decreases. This can perhaps be more clearly seen by analysing a simpler example, we consider constant mortality $\mu_{*,\dagger}$ and interest rate r, as would approximately be valid in the short-term. If we then also use a constant payout function $\widehat{K}$, which would occur when the interest rate is sufficiently low, we would get

$$V^+(t,r) = \int_t^\infty \widehat{K} e^{-(s-t)r} e^{-(s-t)\mu_{*,\dagger}} ds = \frac{\widehat{K}}{r + \mu_{*,\dagger}}.$$

Since this value increases with the inverse of r we see that the short-term impact of low interest rates would contribute to the heavy tails in the distribution of V^+.

As a final note we could now compute any number of statistics from the distribution of V^+, some of which are listed below.

Variance	Expectation	Median	10'th percentile	90'th percentile
0.652	32.4	32.3	31.5	33.5

Chapter 8

Final Conclusions and Further Work

We have now demonstrated how our framework can be used to explicitly compute reserves for a wide range of payout functions under HJM at both current time and at future time points. This theory also extends past the example where we considered payout functions based purely on the interest rate, we could for example consider unit-linked insurance based on the performance of a fund or other financial instrument. In order to apply this framework to other insurance policies we would need to develop pricing formulas for the payout functions $\dot{a}_j$ and $a_{j,k}$. Whenever that is possible the framework would then immediately provide us with explicit formulas to compute relevant information about the reserves.

More general work still remains when it comes to numerical methods. While we have developed two explicit ways of computing and simulating the reserve we could still benefit from an analogue to Thiele's differential equation. Such an equation could possibly be developed for sufficiently nice classes of interest rate and payout function models, which could yield more efficient computation algorithms for the reserves.

The presented example can also be more thoroughly analysed, in particular through comparisons to other pension policies. Furthermore, we could also compute the distribution of the stochastic prospective reserve under P without much issue. Such simulations are highly relevant for insurance companies as they are required to maintain sufficient liquidity under e.g. the Solvency II directive.

Appendix A

Appendix: Code

A.1 K2013 survival probabilities

The following code was used to compute survival probabilities in accordance with the K2013 mortality basis.

Variables used in this code are

- G: Denotes gender
- x: Denotes age in year 2013, not current age.
- t, s: Denotes calendar year (assumed $t, s > 2013$)

Mortality Reduction:

```
w <- function(G, x){
  if(G==0){
    return( min(2.671548-0.17248*x+0.001485*x^2,0) )
  }
  if(G==1){
    return( min(1.287968-0.10109*x+0.000814*x^2,0) )
  }
}
```

Mortality in 2013:

```
mu.kol.2013 <- function(G, x){
  #men
  if(G==0){
    return( (0.241752+0.004536*10^(0.051*x))/1000 )
  }
  #women
  if(G==1){
    return( (0.085411+0.003114*10^(0.051*x))/1000 )
  }
}
```

Adjusted Mortality in Year t:

```
mu.kol <- function(G, x,t){
  return(mu.kol.2013(G, x+t-2013)
      *(1+w(G, x+t-2013)/100)^(t-2013))
}
```

Survival Probability:

```
p_surv <- function(G,x,t,s){
  temp_int <- integrate(f=Vectorize(mu.kol),
      lower=t,upper=s,G=G,x=x)[[1]]
  return(exp(-temp_int))
}
```

A.2 CD Ladder Functions

The code in this section contains functions relevant for computing the value of the reserve in our "CD ladder"-based pension scheme. Note that these functions assume that the following global variables are defined.

- a: The speed of mean reversion
- b: Reversion level
- c: Short-rate volatility
- $delta$: Time to maturity for bonds
- $r0$: Short-rate value at time 0
- rho: Rate of inflation
- K: Payout function at time 0
- $Kmin$: Minimum payout at time 0
- $S0$: Start age of retirement payout
- $Smax$: Maximum age

Computing K_0:

```
Kmaker <-function(){
  return( Kmin / (1-exp(delta*(rho-b) ) ) )
}
```

Simulating $\psi(t)$:

```
generatepsi <- function(t,n=1){
  mean_temp <- r0 - b
  sd_temp <- sqrt( (c^2/(2*a))*(exp(2*a*t)-1) )
  stdnorm_temp <- rnorm(n,mean=mean_temp,sd=sd_temp)
  return(stdnorm_temp)
}
```

Converting r_t to $\psi(t)$:

```
rtopsi <- function(r_t,t){
  return( (r_t-b)*exp(a*t) )
}
```

Converting $\psi(t)$ to r_t:

```
psitor <- function(psi,t){
  return( b+exp(-a*t)*psi )
}
```

Zero-Coupon Bond Value:

```
Pts <- function(t,s,psi_t){
  return(exp(-b*(s-t)-(exp(-a*t)-exp(-a*s))*(psi_t/a) ))
}
```

Variation Factor in Pricing Formula for Call:

```
var_call <- function(t,s){
  return( (c^2/(2*a^3))*(1-exp(2*a*(t-s)))
      *(1-exp(-a*delta))^2 )
}
```

Computing $\varphi^{\pm}$:

```
#Returns list with phi+ in [1] and phi- in [2]
phi <- function(t,s,x,y){
  log_term <- log(  (K*exp(rho*(s+delta))*x)
  /((K-Kmin)*exp(rho*s)*y)  )
  var_term <- var_call(t,s)

  if (var_term==0){
    var_term <- 0.00000001
  }

  total_term_plus  <- (log_term+var_term/2)/sqrt(var_term)
  total_term_minus <- (log_term-var_term/2)/sqrt(var_term)
  out_plus  <- pnorm(total_term_plus)
  out_minus <- pnorm(total_term_minus)
  return(c(out_plus,out_minus))
}
```

Computing the Integrand of the Reserve:

Note that the default values imply the insured is a woman aged 31 in the year 2013.

```
integrand_lifereserve <- function(t,s,psi_t,age13=31,gender=1){
  surv      <- p_surv(gender,age13,2021+t,2021+s)
  P_sd      <- Pts(t,s+delta,psi_t)
  P_s       <- Pts(t,s,psi_t)
  phi_temp  <- phi(t,s,P_sd,P_s)
  phi_plus  <- phi_temp[1]
  phi_minus <- phi_temp[2]

  payout_term  <- surv*P_s *K*exp(rho*s)
  payment_term <- surv*P_sd*K*exp(rho*(s+delta))
  call_term1   <- surv*P_sd*K*exp(rho*(s+delta))*phi_plus
  call_term2   <- surv*P_s *(K-Kmin)*exp(rho*s) *phi_minus
  return(payout_term-payment_term+call_term1-call_term2)
}
```

Computing the Reserve:

```
lifereserve <- function(t,psi_t,n=1000,age13=31,gender=1){
  lower        <- max(t,S0-2021+2013-age13)
  upper        <- Smax-2021+2013-age13
  dt           <- (upper-lower)/n
  number_range <- (1:n-1)*dt + lower
  int_temp     <- sum( Vectorize(integrand_lifereserve)(t,
    number_range,psi_t,age13,gender) )
  int_temp     <- int_temp*dt
  return(int_temp)
}
```

Graphing $V^+(t,b)$:

```
time_list    <- 0:70
reserve_list <- Vectorize(lifereserve)(time_list,0)

plot(time_list,reserve_list,type="l",xlab = "t", lwd=2,
    ylab = expression("V"^"+"*"(t,b)"),main =
        expression("Progression of V"^"+"*"(t,b)"))
abline(v=seq(0,70,10), col="gray",lwd=2)
abline(h=seq(5,30,5), col="gray",lwd=2)

#Plot again to overlay graph over grid
par(new=TRUE)
plot(time_list,reserve_list,type="l",xlab = "t", lwd=2,
    ylab = expression("V"^"+"*"(t,b)"),main =
        expression("Progression of V"^"+"*"(t,b)"))
```

Graphing $V^+(28, r)$:

```
r_list           <- b+(0:800-400)/(10000)
temptime         <- 28
psi_list         <- Vectorize(rtopsi)(r_list,28)
new_reserve_list <- Vectorize(lifereserve)(28,psi_list)

plot(r_list,new_reserve_list,type="n",
    ylab=expression(paste("V"^"+"*"(28,r)")), xlab="r",
        main = expression("Graph of V"^"+"*"(28,r)"),
            xaxt="n",lwd=2)
axis(1,at=seq(0.01,0.09,0.01))
abline(v=seq(0.01,0.09,0.01), col="gray",lwd=2)
abline(h=seq(31,34.5,0.5),    col="gray",lwd=2)

#Plot again to overlay graph over grid
par(new=TRUE)
plot(r_list,new_reserve_list,type="n",
    ylab=expression(paste("V"^"+"*"(28,r)")), xlab="r",
        main = expression("Graph of V"^"+"*"(28,r)"),
            xaxt="n",lwd=2)
```

Simulating $V^+(28)$:

```
set.seed(51)
new_psi_list         <- generatepsi(28,2000)
new_new_reserve_list <- Vectorize(lifereserve)(28,new_psilist)
```

Plot Density of $V^+(28)$:

```
plot(density(new_new_reserve_list,bw=0.226),lwd=2,main =
    expression("Density Function, f, of V"^"+"*"(28)"),
        xlab="x",ylab = "f(x)")
```

Plot Cumulative Distribution of $V^+(28)$:

```
plot(ecdf(new_new_reserve_list),lwd=2,main=
    expression("Cumulative Distribution Function, F, of V"^"+"*"(28)"),
        ylab = "F(x)",yaxt="n")

axis(2,at=seq(0,1,0.1))
axis(1,at=c(32.3))
abline(v=c(32.3), col="gray",lwd=2)
abline(h=c(0.5),  col="gray",lwd=2)

#Plot again to overlay graph over grid
par(new=TRUE)
plot(ecdf(new_new_reserve_list),lwd=2,main=
    expression("Cumulative Distribution Function, F, of V"^"+"*"(28)"),
        ylab = "F(x)",yaxt="n")
```

Bibliography

[1] Bacinello, Anna Rita, and Svein-Arne Persson. "Design and Pricing of Equity-Linked Life Insurance Under Stochastic Interest Rates." SSRN Electronic Journal, 2001. https://doi.org/10.2139/ssrn.276701.

[2] Cohen, Samuel N., and Robert J. Elliot. *"Stochastic Calculus and Applications, 2nd edition*. New York: Birkhäuser, 2015

[3] Filipović, Damir. *Term-Structure Models, A Graduate Course*. Berlin: Springer Berlin, 2009.

[4] Finanstilsynet. https://www.finanstilsynet.no/nyhetsarkiv/brev/2013/nytt-dodelighetsgrunnlag-i-kollektiv-pensjonsforsikring-k2013/

[5] Gaillardetz, Patrice. "Valuation of Life Insurance Products under Stochastic Interest Rates." Insurance: Mathematics and Economics 42, no. 1 (2008): 212–26. https://doi.org/10.1016/j.insmatheco.2007.02.009.

[6] Koller, Michael. *Stochastic Models in Life Insurance*. 2nd ed. Berlin: Springer, 2012.

[7] Lowther, George. *Properties of Feller Processes*. https://almostsuremath.com/2010/07/19/properties-of-feller-processes

[8] Lin, X. Sheldon. Introductory Stochastic Analysis for Finance and Insurance. Hoboken, NJ: Wiley-Interscience, 2006.

[9] McDonald, John N., and N. A. Weiss. *A Course in Real Analysis*. 2nd ed. Amsterdam: Elsevier/Academic Press, 2013.

[10] Persson, Svein-Arne. *"Stochastic Interest Rate in Life Insurance: The Principle of Equivalence Revisited."* Scandinavian Actuarial Journal 1998, no. 2 (1998): 97–112. https://doi.org/10.1080/03461238.1998.10413996.

[11] Rolski, Tomasz, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugels. *Stochastic Processes for Insurance and Finance*: August 5, 1998. Chichester: Wiley, 1999.

[12] Zaglauer, Katharina, and Daniel Bauer. "Risk-Neutral Valuation of Participating Life Insurance Contracts in a Stochastic Interest Rate Environment." Insurance: Mathematics and Economics 43, no. 1 (2008): 29–40. https://doi.org/10.1016/j.insmatheco.2007.09.003.

[13] Øksendal, Bernt. *Stochastic Differential Equations: an Introduction with Applications*. 6th ed. Heidelberg: Springer, 2013.