UiO : Department of Mathematics
University of Oslo

Reserving under stochastic mortality

Erik Martinez Jensen

Master's Thesis, Spring 2022

This master's thesis is submitted under the master's programme *Stochastic Modelling, Statistics and Risk Analysis*, with programme option *Finance, Insurance and Risk*, at the Department of Mathematics, University of Oslo. The scope of the thesis is 60 credits.

The front page depicts a section of the root system of the exceptional Lie group E_8, projected into the plane. Lie groups were invented by the Norwegian mathematician Sophus Lie (1842–1899) to express symmetries in differential equations and today they play a central role in various parts of mathematics.

Abstract

In this thesis we review some fractional and non-fractional stochastic models for describing mortality rates in several countries. In particular, we provide empirical evidence that the geometric fractional Ornstein-Uhlenbeck process is not well suited for modeling the mortality rate in Norway.

As a natural continuation, we compute reserves of life insurance claims under Markovian mortality. Taking the classical approach from the theory of partial differential equations (PDEs), we derive the corresponding Thiele's PDE for prospective reserves under stochastic mortality. As Thiele's PDE is in general difficult to solve analytically, we give two algorithms to numerically approximate the solution surface. We finish by giving an application to premium calculation and reserving with a pension policy, given a geometric Ornstein-Uhlenbeck process.

Lastly, we have included extensive background material in order to introduce newcomers to actuarial life insurance and justify its mathematical foundation in a rigorous fashion.

Acknowledgements

I sincerely appreciate all the love and support I have received from friends, colleagues and family during my time as a math student at UiO. Special thanks to David Ruiz Baños for being an amazing supervisor and discussion partner.

Contents

CHAPTER 1

Introduction

1.1 A brief summary of the thesis

Uncertainty is the only certainty there is in the discipline of actuarial science and the task of an actuary is to provide security where there is insecurity. Predicative models which have a solid mathematical and statistical foundation are necessary in order for an actuary to assess a company's future liabilities. However, it is a common misconception among practitioners to believe that newer models with more accurate predictions are better than previously derived models. Indeed, that is not necessarily the case as there exist examples of models which are extremely computational demeaning and time consuming. That brings us over to the next point, which is that a good model should also be efficient and simple. Efficient in the sense that the model can provide quick information about a dynamic market and simple in the sense that the model is easy to implement, recalibrate and understand.

In life insurance, it is well known that there are two main risk factors for long term policies. Those are the returns on investments from the company and the population's mortality. In this thesis, we will first provide necessary preliminaries before focusing on the latter risk factor which we will tackle from two different perspectives.

The first approach, is to consider the difference in mortality rate from older generations to newer generations. The authors of [DO19] propose a stochastic mortality which follows a geometric fractional Ornstein-Uhlenbeck process in order to capture the generational effects such as medical advancements or global pandemics. We will in particular illustrate that Norwegian mortality generational rates do not seem to exhibit self-similarity, which is a characteristic property of fractional noises.

The second way, is to consider the insured's mortality rate. We will develop a general Markov model under the framework of Itô calculus. In particular, we have derived Thiele's PDE for prospective reserves under general insurance policies and Markovian mortality. We finish up with some important applications to pension schemes using a geometric Ornstein-Uhlenbeck process and highlight the importance of algorithmic efficiency.

1.2 Chapter overview

Chapter 2: The aim of this chapter, is to provide a simple and sufficient mathematical framework for our needs. We start with measure theory, where we follow the notation in [Lin17]. Fundamental, yet important concepts like measures and σ-algebras are introduced together with some important properties. We will then construct the Lebesgue integral from a measure theoretical standpoint and give some classical integration results which allow us to interchange the order of a limit and an integral. Secondly, we will discuss stochastic processes, conditional expectations as random variables and the different notions of martingales following [Øks03], in the sections on probability theory and stochastic analysis. An important process in Itô calculus, is the Brownian motion (Bm) which has many interesting properties. The reference we will use for stochastic calculus is [Bal17], which gives a thorough introduction to the stochastic integral and stochastic differential equations (SDEs). Lastly, we will go over some well established results regarding regular Markov chains.

Chapter 3: In this chapter, we will focus on fractional stochastic processes. We begin with a generalization of the Brownian motion known as the fractional Brownian motion (fBm) and explore some of its key properties as presented in [Nua95]. In particular, we will represent the fractional Brownian motion as a pathwise Riemann-Stieltjes integral for deterministic integrands, via an isometry between Hilbert spaces. Next we will look at the fractional generalization of the Ornstein-Uhlenbeck process and simulate some paths using the Euler-Maruyama method. Lastly, we will make sense of stochastic integrals with respect to a fractional Brownian motion as a pathwise Riemann-Stieltjes integral.

Chapter 4: We explore the assumptions made in the model for fractional mortality proposed by [DO19]. Furthermore, we present a simple estimation procedure for key parameters. Lastly, we conduct our own simulation study in order to get an intuition for the distribution of the estimators.

Chapter 5: [Gar18] emphasizes that the underlying and perceived Hurst parameter are not equal, which was assumed in [DO19]. As a consequence, we must take the Hurst parameter as an exogenous variable in the fractional mortality model. We divide the analysis in two cases and present the corresponding results. Finally, we conclude that fractional mortality might be a large population phenomenon and thus not a well suited model for Norwegian mortality rates.

Chapter 6: A brief introduction on life insurance mathematics in continuous time is given following [Kol12]. The main takeaway from this chapter is the explicit formula for prospective reserves and Thiele's ordinary differential equation (ODE), given that the states of the insured are deterministic quantities. An applications to premium calculation and reserving is also given under the Gompertz-Makeham's law of mortality in Norway.

Chapter 7: Thiele's PDE is derived under the assumption that the mortality rate is given as the solution of an Itô stochastic differential equation (SDE). As numerical methods are required in order to get an approximate solution of Thiele's PDE, we give a general algorithm for the explicit method and implicit method in our insurance setting. Lastly, we give an example related to pension where the mortality rate follows a geometric Ornstein-Uhlenbeck process and highlight different reserving methods.

Chapter 8: A short summary of the research conducted in this thesis together with some remarks on ideas for future work which expand on our model.

Appendix A: An attachment containing all the code we have implemented for simulations and computations throughout this thesis.

CHAPTER 2

Framework

We will start this chapter by giving an overview of the theory, notation and assumptions used throughout this thesis. We will assume that the reader is familiar with probability theory, integration theory and stochastic analysis. For more details about measure theory the reader should consult [Lin17] which most of the results here are gathered from.

2.1 Measure theory

Measure theory is the study of measures, which seeks to generalize the intuitive notion of size. This includes classical concepts from Euclidean geometry such as length, area and volume, but also more abstract ideas including probability.

Definition 2.1.1 (σ-algebra) Assume that X is a non-empty set. We say that a collection $\mathcal{A}$ of subsets of X is a σ-algebra if it satisfies:

(i) $\varnothing \in \mathcal{A}$,
(ii) If $A \in \mathcal{A} \implies A^c \in \mathcal{A}$,
(iii) If $A \in \mathcal{A} \implies \bigcup_{i=1}^{\infty} A_i \in \mathcal{A}$.

The sets in $\mathcal{A}$ are called the measurable sets of X and are the sets which are well enough behaved to be assigned a meaningful notion of size. The pair $(X, \mathcal{A})$ is called a measurable space.

A σ-algebra is used to model the possible events of an experiment. For instance, if we are interested in modelling the possible health states of an insured person, then the associated σ-algebra would contain all the possible information regarding the individual's health states.

Definition 2.1.2 (Generated σ-algebra) Let $X \neq \varnothing$ and let $\mathcal{B}$ be a collection of subsets of X. Then we define the σ-algebra generated by $\mathcal{B}$, denoted by $\sigma(\mathcal{B})$, to be the smallest σ-algebra containing $\mathcal{B}$. That is

$$\sigma(\mathcal{B}) = \bigcap_{\mathcal{B} \subseteq \mathcal{A}} \mathcal{A},$$

for all σ-algebras $\mathcal{A}$ on X.

This definition is particularly interesting for describing σ-algebras of events through observing random variables.

Example 2.1.3 (Borel σ-algebra on $\mathbb{R}$) Let $\mathcal{B}$ be the collection of all open subsets of $\mathbb{R}$. We say that $\sigma(\mathcal{B})$ is the Borel σ-algebra on $\mathbb{R}$ and we will denote it by $\mathcal{B}(\mathbb{R})$.

Definition 2.1.4 (Measure) Assume $(X, \mathcal{A})$ is a measurable space, we say that a function μ: $\mathcal{A} \to [0, \infty]$ is a measure on $(X, \mathcal{A})$ if it satisfies:

(i) $\mu(\varnothing) = 0$
(ii) (Countable additivity) If $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ and $A_i \cap A_j = \varnothing$ for all natural numbers $i \neq j$ then:

$$\mu(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} \mu(A_i).$$

We call the triple $(X, \mathcal{A}, \mu)$ a measure space.

Definition 2.1.5 (Finite measure) We say that a measure is finite on the measurable space $(X, \mathcal{A})$ if and only if

$$\mu(X) < \infty.$$

Definition 2.1.6 (Probability space) We say that a measure space $(X, \mathcal{A}, \mu)$ is a probability space if μ is finite and in particular $\mu(X) = 1$. We will denote probability spaces by $(\Omega, \mathcal{A}, P)$.

Probability spaces will be essential in our modeling situations as they contain all the information of an experiment. In most of the cases, our σ-algebras will be generated by observable random variables and the probabilities defined through the distributions of these random variables.

Definition 2.1.7 (σ-finite measure) We say that a measure is σ-finite if there exists a sequence $\{A_i\}_{i=1}^{\infty} \in \mathcal{A}$ with $\mu(A_i) < \infty$ for all $i \in \mathbb{N}$ such that

$$X = \bigcup_{i \in \mathbb{N}} A_i.$$

Definition 2.1.8 (Null set) Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that a set $N \subseteq X$ is a null set if there exists a set $A \in \mathcal{A}$ such that $N \subseteq A$ with $\mu(A) = 0$.

We will denote the collection of all null sets by $\mathcal{N}$. And we say that the measure space is complete if $\mathcal{N} \subseteq \mathcal{A}$.

Example 2.1.9 (Lebesgue measure) Let us consider a concrete example where our measure space is $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$. Here $\mathbb{R}$ is the set of real numbers. $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra of subsets of $\mathbb{R}$. The Lebesgue measure λ: $\mathcal{B}(\mathbb{R}) \to [0, \infty]$ defined by

$$\lambda((a, b)) = b - a$$

for $a \leq b$.

Note that the Lebesgue measure on $\mathbb{R}$ is an example of a non-finite measure since $\lambda(\mathbb{R}) = \infty$. However it is still σ-finite since

$$\mathbb{R} = \bigcup_{n \in \mathbb{N}} (-n, n)$$

is a countable covering of $\mathbb{R}$ with measurable sets, each with Lebesgue measure $2n$.

Proposition 2.1.10 (Continuity of measure) Let $\{A_n\}_{n\in\mathbb{N}} \in \mathcal{A}$ be a monotonic sequence on a measure space $(X, \mathcal{A}, \mu)$.

(i) If the sequence is increasing, then

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

(ii) If the sequence is decreasing and $\mu(A_1) < \infty$, then

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

Terminology 2.1.11 (Almost everywhere) Let $(X, \mathcal{A}, \mu)$ be a measure space. We say that a property P holds almost everywhere, abbreviated to a.e., if

$$\mu(\{x \in X : \neg P\}) = 0.$$

Definition 2.1.12 (Absolutely continuous measures) Let μ and ν be measures on $(X, \mathcal{A})$. We say that ν is absolutely continuous with respect to μ if

$$\mu(A) = 0 \implies \nu(A) = 0, \qquad \forall A \in \mathcal{A}.$$

Symbolically we will write $\nu \ll \mu$. Moreover we say that ν and μ are equivalent measures if they share all the same null sets. This means that $\nu \ll \mu$ and $\mu \ll \nu$ which we will denote as $\nu \sim \mu$.

Definition 2.1.13 (Singular measures) Let μ and ν be measures on $(X, \mathcal{A})$. We say that ν is singular with respect to μ if there exists $N \in \mathcal{A}$ such that $\mu(N) = 0$ and $\nu(N^C) = 0$. We will write this as $\nu \perp \mu$.

Theorem 2.1.14 (Lebesgue's decomposition theorem) Let μ and ν be σ-finite measures on $(X, \mathcal{A})$. Then there are two distinct measures $\mu_{ac}, \mu_s : \mathcal{A} \to [0, \infty]$ such that

$$\mu = \mu_{ac} + \mu_s, \qquad \mu_{ac} \ll \nu, \qquad \mu_s \perp \nu.$$

We say that μ has been Lebesgue decomposed with respect to ν.

2.2 Integration theory

It is usually the case when learning mathematics that one encounters Riemann integrals before measure theory. The intuition is that the integral measures the area under some function. However, a different approach would be to generalize our understanding of size using measures and then use measures to define the Lebesgue integral. This will be our goal in this section. We will also include some applications of this theory. For more information about integration with measures read [Lin17].

Definition 2.2.1 (Measurable function) We say that a function $f : X \to [-\infty, \infty]$ on a measure space $(X, \mathcal{A}, \mu)$ is measurable if

$$f^{-1}[-\infty, r) \in \mathcal{A}$$

for all $r \in \mathbb{R}$.

Definition 2.2.2 (Simple function) We say that $f(x) = \sum_{i=1}^{n} a_i 1_{A_i}(x)$ is a nonnegative simple function whenever $\{A_i\}_{i=1}^{n} \in \mathcal{A}$ make a partition of X and $a_i \geq 0$ for all $i \in \{1, ..., n\}$, $n \in \mathbb{N}$. We define the integral of a simple function as

$$\int_X f d\mu = \sum_{i=1}^{n} a_i \mu(A_i).$$

Definition 2.2.3 (Integral of a nonnegative function) For a nonnegative measurable function f we define the integral as

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

where f_n is a nondecreasing sequence of simple functions converging pointwise to f.

Remark 2.2.4 (Existence of simple functions) We can always find a nonnegative sequence of simple functions f_n converging pointwise to f by cutting the interval $[0, 2^n)$ into the followings subintervals of length $\frac{1}{2^n}$

$$I_k = \Big[\frac{k}{2^n}, \frac{k+1}{2^n}\Big),$$

for $k \in [0, 2^{2n})$. Then the sets

$$A_k = f^{-1}(I_k)$$

are measurable and we can construct

$$f_n(x) = \sum_{k=0}^{2^{2n}-1} \frac{k}{2^n} 1_{A_k}(x) + 2^n 1_{\{x: f(x) \geq 2^n\}}.$$

By construction f_n is a nonnegative simple function approaching f from bellow pointwise. This is because for each unit increase of n, we cover more of $[0, \infty)$ and the intervals I_k get smaller. This is nondecreasing since each time we split the interval I_k into two new parts we either follow the same function values or make an upwards jump.

Definition 2.2.5 (Integrable function) We say that a nonnegative function f is integrable if it is both measurable and $\int_X f d\mu < \infty$.

Observation 2.2.6 (Nonnegative decomposition) Let f_+, f_- be nonnegative functions given by

$$f_+(x) = \begin{cases} f(x), & f(x) > 0 \\ 0, & f(x) \leq 0 \end{cases} \qquad f_-(x) = \begin{cases} -f(x), & f(x) < 0 \\ 0, & f(x) \geq 0 \end{cases}.$$

Then any function $f : X \to [-\infty, \infty]$ can be written as $f = f_+ - f-$. Also note that $|f| = f_+ + f-$.

Definition 2.2.7 (Integral of a real valued function) A function $f : X \to [-\infty, \infty]$ is called integrable if it is measurable and f_+ and f_- are integrable. Then the integral of f is given by

$$\int_X f d\mu = \int_X f_+ d\mu - \int_X f_- d\mu.$$

Observe that if

$$\int_X |f| d\mu < \infty \implies \left| \int_X f d\mu \right| < \infty,$$

which yields that f is integrable. Although it is a stronger statement, this is often a standard requirement.

Radon-Nikodym

Theorem 2.2.8 (Radon-Nikodym) Let μ and ν be measures on $(X, \mathcal{A})$. If μ is σ-finite then the following two statements are equivalent:

(i) $\nu \ll \mu$.
(ii) There exists a measurable function $f : X \to [0, \infty]$ such that $f = \frac{d\nu}{d\mu}$.

The Radon-Nikodym theorem is extremely useful as it gives a necessary and sufficient condition to guarantee the existence of a density function. We will use this theorem later on to apply a change of measure when computing integrals with respect to distribution functions.

Definition 2.2.9 (Density function) Let $(X, \mathcal{A}, \mu)$ be a measure space. Let $f : X \to [0, \infty]$ be a measurable and μ-integrable function. We define a new measure ν: $\mathcal{A} \to [0, \infty]$ by

$$\nu(A) = \int_A f d\mu, \qquad A \in \mathcal{A}.$$

We say that f is the density function μ-a.e. We will denote the density by $f = \frac{d\nu}{d\mu}$ and also refer to it as the Radon-Nikodym derivative.

Product spaces

Definition 2.2.10 (Product σ-algebra) Let $(X_1, \mathcal{A}_1)$ and $(X_2, \mathcal{A}_2)$ be measurable spaces. We say that $\mathcal{A}_1 \otimes \mathcal{A}_2$ is the product σ-algebra generated by the Cartesian product of measurable sets in $\mathcal{A}_1$ and $\mathcal{A}_2$. That is

$$\mathcal{A}_1 \otimes \mathcal{A}_2 = \sigma(\{A_1 \times A_2 : A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2\}).$$

Definition 2.2.11 (Product measure) Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be measure spaces. We say that $\mu_1 \times \mu_2$ is a product measure on the measurable space $(X_1 \times X_2, \mathcal{A}_1 \otimes \mathcal{A}_2)$ if it satisfies the property

$$\mu_1 \times \mu_2(A_1 \times A_2) = \mu_1(A_1)\mu_2(A_2),$$

for all $A_1 \in \mathcal{A}_1, A_2 \in \mathcal{A}_2$.

Theorem 2.2.12 (Tonelli-Fubini's theorem) Let $(X_1, \mathcal{A}_1, \mu_1)$ and $(X_2, \mathcal{A}_2, \mu_2)$ be σ-finite measure spaces and let $f : X_1 \times X_2 \to \mathbb{R}$ be a $\mathcal{A}_1 \otimes \mathcal{A}_2$-measurable function. Then

$$\int_{X_1} \Big(\int_{X_2} |f| d\mu_2 \Big) d\mu_1 = \int_{X_2} \Big(\int_{X_1} |f| d\mu_1 \Big) d\mu_2 = \int_{X_1 \times X_2} |f| d(\mu_1 \times \mu_2).$$

Furthermore, if one of the integrals above are finite then f is integrable and we have that

$$\int_{X_1} \Big(\int_{X_2} f d\mu_2 \Big) d\mu_1 = \int_{X_2} \Big(\int_{X_1} f d\mu_1 \Big) d\mu_2 = \int_{X_1 \times X_2} f d(\mu_1 \times \mu_2).$$

Convergence theorems

We would like to interchange limits and integrals so that the following equation holds

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

Unfortunately the above equation does not hold in general as we will see in this next example.

Example 2.2.13 (Limits and integrals do not commute) Let λ be the usual Lebesgue measure on $\mathbb{R}$ and define the sequence of functions $f_n : \mathbb{R} \to \mathbb{R}$ by

$$f_n(x) = \begin{cases} n^2, & x \in (0, \frac{1}{n^2}) \\ 0, & else. \end{cases}$$

We see that for all $x \in \mathbb{R}$

$$\int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)d\lambda = \int_{\mathbb{R}} 0 d\lambda = 0.$$

However, the same is not true if we interchange the limit and integral above since

$$\lim_{n\to\infty}\int_{\mathbb{R}} f_n(x)d\lambda = \lim_{n\to\infty} 1 = 1.$$

Fortunately this problem of interchanging limits and integrals is already well studied in measure theory. We will therefore state some key results.

Theorem 2.2.14 (Monotone convergence theorem) If $\{f_n(x)\}_{n\in\mathbb{N}}$ is an increasing sequence of nonnegative, measurable functions such that $f(x) = \lim_{n\to\infty} f_n(x)$ for all $x \in X$ and f is μ-integrable, then

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

Theorem 2.2.15 (Fatou's lemma) If $\{f_n(x)\}_{n\in\mathbb{N}}$ is a sequence of nonnegative measurable functions, then

$$\liminf_{n\to\infty}\int_X f_n(x)d\mu \geq \int_X \liminf_{n\to\infty} f_n(x)d\mu.$$

Theorem 2.2.16 (Dominated convergence theorem) Assume $g : X \to [0,\infty]$ is a nonnegative, integrable function and that $\{f_n(x)\}_{n\in\mathbb{N}}$ is a sequence of measurable functions converging pointwise to f. If $|f_n(x)| \leq g(x)$ for all n, then

$$\lim_{n\to\infty}\int_X |f_n(x) - f(x)|d\mu = 0.$$

Since convergence in norm implies convergence of the norms, we have that

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

2.3 Probability theory

Probability theory is a part of measure theory with additional structure. We will in this section apply our knowledge of measure theory to motivate the Riemann-Stieltjes integral and other probabilistic results. This section is based on the books [Wal12], [Øks03] and [Bal17].

Definition 2.3.1 (Filtration) We say that a collection of σ-algebras $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0} \subseteq \mathcal{A}$ is a filtration on $(\Omega, \mathcal{A})$ if

$$0 \leq s \leq t \implies \mathcal{F}_s \subseteq \mathcal{F}_t.$$

In insurance one can think of σ-algebras as information. A filtration extends this idea by modeling the development of information by time.

Definition 2.3.2 (P-augmented filtration) A filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}$ on $(\Omega, \mathcal{A}, P)$ is called P-augmented if the collection of all nullsets under the probability measure P, denoted by $\mathcal{N}$, is contained in $\mathcal{F}_0$. That is

$$\mathcal{N} \subseteq \mathcal{F}_0.$$

Definition 2.3.3 (Right continuous filtration) A filtration $\mathcal{F} = \{\mathcal{F}_t\}_{t\geq 0}$ on $(\Omega, \mathcal{A})$ is called right continuous if

$$\mathcal{F}_t = \bigcap_{u>t} \mathcal{F}_u, \qquad t \geq 0.$$

Definition 2.3.4 (Random variable) A mapping $X : \Omega \to \mathbb{R}$ which is measurable with respect to $\mathcal{A}$ is called a random variable.

Remark 2.3.5 (Pushforward measure) Every random variable induces a probability measure μ_X on $\mathcal{B}(\mathbb{R})$ defined by

$$\mu_X(B) = P(X^{-1}(B)), \qquad B \in \mathcal{B}(\mathbb{R}).$$

The pushforward measure is also called the law of X.

Theorem 2.3.6 (Image measure) Let $(X, \mathcal{A}, \mu)$ be a measure space and let $(Y, \mathcal{C})$ be a measurable space. Assume also that we have a measurable transformation $T : X \to Y$ and a measurable function $f : Y \to \mathbb{R}$. Then the integral of f exists as a function on $(Y, \mathcal{C}, \mu_T)$ if and only if the integral of $f \circ T$ exists as a function on $(X, \mathcal{A}, \mu)$. If that is the case the following equation holds

$$\int_X f \circ T d\mu = \int_Y f d\mu_T.$$

Notice that f is integrable if and only if $f \circ T$ is integrable.

Definition 2.3.7 (Distribution function) We define $F : \mathbb{R} \to [0, 1]$ as the distribution function of a random variable X given by

$$F(x) = P(X(\omega) \leq x).$$

Remark 2.3.8 (Important relation) When working with abstract probability spaces it is often a good idea to use the induced probability measure μ_X to prove theoretical results. However, in practical applications it is often convenient to characterize the distribution of a random variable by its distribution function. An important and trivial relationship between the two concepts is

$$F(x) = P(X(\omega) \leq x) = P(X^{-1}(-\infty, x]) = \mu_X(-\infty, x]).$$

Definition 2.3.9 (Expectation) Let $X : \Omega \to \mathbb{R}$ be a random variable and $f : \mathbb{R} \to \mathbb{R}$ be a measurable function. If

$$\int_\Omega |f(X(\omega))| dP < \infty,$$

we define the expectation of $f(X)$ by

$$E[f(X)] = \int_\Omega f(X(\omega)) dP = \int_\mathbb{R} f(x) d\mu_X.$$

If $\mu_X \ll \lambda$ we can simplify the expectation to a Lebesgue-integral by using the Radon-Nikodym theorem to find a density g.

$$E[f(X)] = \int_\mathbb{R} f(x) d\mu_X = \int_\mathbb{R} f(x) \frac{d\mu_X}{d\lambda} d\lambda = \int_\mathbb{R} f(x) g(x) dx.$$

Definition 2.3.10 (Variance) Let X be a random variable with finite second moment, i.e $E[X^2] < \infty$, then we define the variance of X as

$$Var(X) = E\Big[\big(X - E[X]\big)^2\Big] = E[X^2] - E[X]^2.$$

Definition 2.3.11 (Covariance) Let X, Y be two random variables. We define their covariance by

$$Cov(X, Y) = E\Big[\big(X - E[X]\big)\big(Y - E[Y]\big)\Big] = E[XY] - E[X]E[Y].$$

Definition 2.3.12 (Moment generating function) Let X be a random variable. We define the moment generating function $M : \mathbb{R} \to \mathbb{R}$ of X by

$$M_X(t) = E[e^{tX}],$$

whenever the expectation exists.

Definition 2.3.13 (Characteristic function) Let X be a random variable. We define the function $\phi_X : \mathbb{R} \to \mathbb{C}$ as its characteristic function by

$$\phi_X(t) = E[e^{itX}], \qquad t \in \mathbb{R}.$$

Remark 2.3.14 (Characteristic functions and moment generating functions) An important observation about the characteristic function is that it always exists unlike the moment generating function. While the moment generating function may run into integrability problems, the characteristic functions does not as in the latter case we integrate around the unit circle. However, both functions are useful for studying distributional properties of a random variable.

Borell-Cantelli

Definition 2.3.15 (Limit superior) Let $(X, \mathcal{A})$ be a measurable space. We define the limit superior of a sequence of sets $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ by

$$\limsup_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} \bigcup_{k=n}^{\infty} A_n = \{A_n \text{ happens for infinitely many } n\}.$$

Definition 2.3.16 (Limit inferior) Let $(X, \mathcal{A})$ be a measurable space. We define the limit inferior of a sequence of sets $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ by

$$\liminf_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_n = \{A_n \text{ happens for all but finitely many } n\}.$$

Definition 2.3.17 (Independent sets) Let $(\Omega, \mathcal{A}, P)$ be a probability space. We say that a sequence of sets $\{A_n\}_{n\in\mathbb{N}} \subseteq \mathcal{A}$ are independent if

$$P\Big(\bigcap_{n=1}^{\infty} A_n\Big) = \prod_{n=1}^{\infty} P(A_n).$$

Theorem 2.3.18 (Borel-Cantelli lemma) Assume $\{A_n\}_{n\in\mathbb{N}}$ is a sequence of events.

(i) If $\sum_{i=1}^{\infty} P(A_n) < \infty$, then $P\Big(\limsup_{n\to\infty} A_n\Big) = 0$.

(ii) If $\sum_{i=1}^{\infty} P(A_n) = \infty$ and the events A_n's are independent, then $P\Big(\limsup_{n\to\infty} A_n\Big) = 1$.

Borel-Cantelli lemma is a well known result from probability theory and is often referred to as a 0-1 law.

Riemann-Stieltjes integral

Definition 2.3.19 (Mesh) Let $\Pi = \{x_i\}_{i=0}^{n-1} \subseteq [a, b]$ be a partition, $a = x_0 < ... < x_{n-1} = b$. We define its mesh by

$$|\Pi| = \max_{0\leq i\leq n-1} |x_{i+1} - x_i|.$$

Definition 2.3.20 (p-variation) Let $\Pi = \{x_i\}_{i=0}^{n-1} \subseteq [a, b]$ be a partition and $f : [a, b] \to \mathbb{R}$ a function. We define the p-variation $V^p_{[a,b]}(\Pi, f)$ by

$$V^p_{[a,b]}(\Pi, f) = \lim_{|\Pi|\to 0} \sum_{i=0}^{n-1} |f(x_{i+1}) - f(x_i)|^p, \qquad p \geq 1.$$

We will say that f has bounded total variation if the sum above converges for $p = 1$. Likewise we will say that f has bounded quadratic variation if it has bounded 2-variation.

Definition 2.3.21 (Riemann-Stieltjes integral) Let $f, g : [a, b] \to \mathbb{R}$ be real valued functions. We define the Riemann-Stieltjes integral of f with respect to g by

$$\int_a^b f dg = \lim_{|\Pi| \to 0} \sum_{i=0}^{n-1} f(c_i)\Big(g(x_{i+1}) - g(x_i)\Big),$$

whenever the limit exists for all partition Π of $[a, b]$ and each $c_i \in [x_i, x_{i+1}]$.

Remark 2.3.22 (Existence of the Riemann-Stieltjes integral) We would like to impose that $\int_a^b f dg$ is well-defined. A sufficient condition for the existence of the Riemann-Stieltjes integral is that g is of bounded total variation and that f is continuous on $[a, b]$. Indeed it can be checked that

$$\Big| \int_a^b f dg \Big| = \Big| \sum_{i=0}^{n-1} f(c_i)\Big(g(x_{i+1}) - g(x_i)\Big) \Big|$$

$$\leq \sup_{c \in [a,b]} |f(c)| \sum_{i=0}^{n-1} \Big| g(x_{i+1}) - g(x_i) \Big| \leq \sup_{c \in [a,b]} |f(c)| V^1_{[a,b]}(\Pi, g) < \infty.$$

Where we have used the triangle inequality. That f is continuous on a compact set such that f must be bounded by the extreme value theorem. Finally, that g is of bounded total variation.

Remark 2.3.23 (Computing Riemann-Stieltjes integral) Let $f, g : \mathbb{R} \to \mathbb{R}$. If f is continuous and g is a.e. differentiable with at most a finite number of discontinuities occurring at $x_1, x_2, x_3, ..., x_n \in \mathbb{R}$, then we define the Riemann-Stieltjes integral of f with respect to g by

$$\int_{\mathbb{R}} f(x) dg(x) = \int_{\mathbb{R}} f(x) g'(x) dx + \sum_{i=1}^{n} f(x_i)(g(x_i) - g(x_i-)).$$

Here $g(x_i-) = \lim_{x \uparrow x_i} g(x)$ denotes the left limit of g at x_i. If we are fortunate enough that g is continuous and a.e. differentiable on the whole domain, then the Riemann-Stieltjes integral is simply just a Riemann integral.

Inequalities

Theorem 2.3.24 (Chebyshev's inequality) Let X be a centered random variable and let $p, \lambda > 0$, then

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

In particular, for $p = 2$ we have that

$$P(|X| \geq \lambda) \leq \frac{1}{\lambda^2} Var(X).$$

Chebyshev's inequality can be useful in applications since it yields an upper bound on the probability that the absolute deviation from the mean is outside of a given threshold. On the other hand, the upper bound may often be too coarse.

Theorem 2.3.25 (Schwarz's inequality) Let X, Y be random variables with finite absolute second moments, then

$$E[|XY|] \leq E[X^2]^{\frac{1}{2}} E[Y^2]^{\frac{1}{2}}.$$

Schwarz's inequality is perhaps the most important inequality when operating with inner product spaces.

Theorem 2.3.26 (Jensen's inequality) Let $\phi : (a, b) \to \mathbb{R}$ be a convex function and let $X : \Omega \to (a, b)$ be a random variable . If both X and $\phi(X)$ are integrable, then

$$\phi(E[X]) \leq E[\phi(X)].$$

Jensen's inequality is vital for convex analysis but also in probability theory. Important applications in probability theory include the derivation of Lyapunov's inequalities and the existence of characteristic functions.

Theorem 2.3.27 (Lyapunov's inequalities) Let X be a random variable and let $1 \leq p \leq q$, then

$$(i) \quad E[|X|]^p \leq E[|X|^p] \qquad\qquad (ii) \quad E[|X|^p]^{\frac{1}{p}} \leq E[|X|^q]^{\frac{1}{q}}.$$

Lyapunov's inequalities are useful when doing analysis in L^p- spaces.

Convergence in what sense?

We are often interested in convergence of a random variable of the type $\lim_{n\to\infty} X_n = X$. However this notation is ambiguous without specifying what type of convergence we are exploring. To avoid confusion, we will briefly summarize the most important types of convergence as well as the relationships between them.

Definition 2.3.28 (Pointwise convergence) We say that a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges pointwise to X if for all $\omega \in \Omega$ we have that

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

Definition 2.3.29 (Convergence almost surely) We say that a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges almost surely, often abbreviated a.s., to X if there exists a null set N, under the probability measure P, such that for all $\omega \notin N$.

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

Definition 2.3.30 (Convergence in probability) We say that a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges in probability to X if for all $\omega \in \Omega$ we have that

$$\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0,$$

for every $\epsilon > 0$.

Definition 2.3.31 (Convergence in L^p) Let $p \geq 1$, we say that a sequence $\{X_n\}_{n\in\mathbb{N}}$ of random variables converges in L^p to X if for all $\omega \in \Omega$ we have that

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

Definition 2.3.32 (Convergence in distribution) Let $\{F_n\}_{n\in\mathbb{N}}$ and F be distribution functions of the random variables $\{X_n\}_{n\in\mathbb{N}}$ and X respectively. We say that $\{X_n\}_{n\in\mathbb{N}}$ converges to X in distribution whenever

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

at all points where F is continuous.

Convergence relations

(i) If $\{X_n\}_{n\in\mathbb{N}}$ converges to X in L^p or almost surely, then $\{X_n\}_{n\in\mathbb{N}}$ converges to X in probability.

(ii) If $\{X_n\}_{n\in\mathbb{N}}$ converges to X in probability, then $\{X_n\}_{n\in\mathbb{N}}$ converges to X in distribution.

(iii) If $\{X_n\}_{n\in\mathbb{N}}$ converges to X in probability, then there exists a subsequence $\{X_{n_k}\}_{k\in\mathbb{N}}$ converging to X almost surely.

2.4 Stochastic analysis

The goal of this section is to explore stochastic processes and their properties. We will focus on Brownian motion and martingale theory which will be vital for what is to come. The content is gathered from [Øks03] and [Bal17].

Definition 2.4.1 (Stochastic process) A stochastic process $\{X_t\}_{t\in\mathcal{T}}$ is a collection of random variables indexed by $\mathcal{T}$. Common choices of $\mathcal{T}$ include

$$\mathcal{T} = \mathbb{N}\cup\{0\},\ \mathcal{T} = [0,T],\ \mathcal{T} = [0,\infty), \qquad\qquad T > 0.$$

Note that a stochastic process can be viewed as a measurable mapping X: $\Omega\times\mathcal{T}\to\mathbb{R}$. By fixing $t\in\mathcal{T}$ and letting $\omega\mapsto X_t(\omega)$ we get a random variable. By fixing $\omega\in\Omega$ and letting $t\mapsto X_t$ we get a function, known as a sample path of the process, or also called a trajectory.

Definition 2.4.2 (Adapted process) A stochastic process $\{X_t\}_{t\in\mathcal{T}}$ is adapted to a filtration $\{\mathcal{F}_t\}_{t\in\mathcal{T}}$ provided that X_t is $\mathcal{F}_t$-measurable for all $t\in\mathcal{T}$.

Definition 2.4.3 (Modification) Let $X = \{X_t\}_{t\in\mathcal{T}}$ and $Y = \{Y_t\}_{t\in\mathcal{T}}$ be stochastic processes. We say that X is a modification of Y if

$$P(X_t = Y_t) = 1, \qquad \forall t\in\mathcal{T}.$$

Definition 2.4.4 (Hölder continuity) Let $f:\mathbb{R}\to\mathbb{R}$ be a function. We say that f is Hölder continuous if there exist constants $C,\gamma>0$ such that

$$|f(x)-f(y)| < C|x-y|^{\gamma}, \qquad \forall x,y\in\mathbb{R}.$$

Theorem 2.4.5 (Kolmogorov's continuity theorem) Let $X = \{X_t\}_{t\in\mathcal{T}}$ be a stochastic process with constants $\alpha,\beta,C>0$ satisfying the equation

$$E[|X_t - X_s|^{\beta}] \le C|t-s|^{1+\alpha} \qquad s\le t.$$

Then there exists almost surely a modification $Y = \{Y_t\}_{t\in\mathcal{T}}$ of X that is Hölder continuous for all $\gamma\in(0,\frac{\alpha}{\beta})$ on every bounded time interval.

Definition 2.4.6 (Brownian motion) A stochastic process $B = \{B_t\}_{t\in[0,\infty)}$ is called a Brownian motion starting in $x\in\mathbb{R}$ if

(i) $P(B_0 = x) = 1$.
(ii) $B_t - B_s$ is independent of $\mathcal{F}_s$ for all s $\le$ t.
(iii) $B_t - B_s$ is normally distributed with mean zero and variance $t-s$, i.e $B_t - B_s \sim N(0, t-s)$.

Example 2.4.7 (Brownian motion has a continuous modification) First note that the Brownian motion can be written in the following form

$$B_t - B_s = (t-s)^{\frac{1}{2}}Z, \qquad Z\sim N(0,1),$$

where the equality is in the sense of law.

Now we take the absolute value and exponentiate both sides by β. Lastly, we apply the expectation on both sides which yields

$$E[|B_t - B_s|^{\beta}] = (t-s)^{\frac{\beta}{2}}E[|Z|^{\beta}].$$

This is in the form of Kolmogorov's continuity theorem with $\alpha = \frac{\beta}{2} - 1$ and $C = E[|Z|^\beta] < \infty$. Thus there is almost surely a modification of a Brownian motion which is Hölder continuous with $\gamma \in (0, \frac{1}{2} - \frac{1}{\beta})$. Since $\beta > \frac{1}{2}$, we have that $\gamma \in (0, \frac{1}{2})$.

Example 2.4.8 (Brownian motion is not differentiable) Recall the definition of a differentiable function $f : \mathbb{R} \to \mathbb{R}$. We say that f is differentiable at a point $x \in \mathbb{R}$ if the following limit exists and is finite,

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

We will show that the Brownian motion is not differentiable in distribution. Define a random variable X_h by

$$X_h = \frac{B_{t+h} - B_t}{h}.$$

Then it is straight forward to compute its expectation and variance using the properties of Brownian motion.

$$E[X_h] = E\Big[\frac{B_{t+h} - B_t}{h}\Big] = \frac{1}{h}E[B_h] = 0.$$

$$Var(X_h) = Var\Big(\frac{B_{t+h} - B_t}{h}\Big) = \frac{1}{h^2}Var(B_h) = \frac{1}{h^2}h = \frac{1}{h}.$$

Since a linear combination of independent normal random variables is again a normal random variable, we can write the following equality in distribution

$$X_h = \frac{Z}{\sqrt{h}}, \qquad Z \sim N(0,1).$$

We claim that X_h is unbounded in probability and as a result nowhere differentiable. Indeed for an arbitrary large number $N \in \mathbb{N}$ we have that

$$\lim_{h\to 0} P(|X_h| > N) = \lim_{h\to 0} P(|\frac{Z}{\sqrt{h}}| > N) = \lim_{h\to 0} P(|Z| > \sqrt{h}N)$$

$$= \lim_{h\to 0} P(Z < -\sqrt{h}N) + P(Z > \sqrt{h}N) = \lim_{h\to 0} 2P(Z > \sqrt{h}N)$$

$$= \lim_{h\to 0} 2\int_{\sqrt{h}N}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}} dz = \frac{2}{\sqrt{2\pi}} \lim_{h\to 0} \int_{-\infty}^{\infty} 1_{\{z>\sqrt{h}N\}} e^{\frac{-z^2}{2}} dz$$

$$= \frac{2}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \lim_{h\to 0} 1_{\{z>\sqrt{h}N\}} e^{\frac{-z^2}{2}} dz = \frac{2}{\sqrt{2\pi}} \int_{0}^{\infty} e^{\frac{-z^2}{2}} = \frac{2}{\sqrt{2\pi}} \frac{\sqrt{\pi}}{\sqrt{2}} = 1.$$

Where we have used the dominated convergence theorem to justify the interchange of limit and integral in the seventh equality. Note that the integrand is bounded by $e^{\frac{-z^2}{2}}$ which is integrable. It is possible to prove that Brownian motion is almost surely nowhere differentiable. The result is significantly stronger, but it comes with the cost of a more technical proof which can be found in [Wal12].

Example 2.4.9 (Brownian motion has bounded quadratic variation) We will show that the quadratic variation of Brownian motion converges in $L^2(P)$ to the length of the time interval. With out loss of generality we will consider the case where $\Pi = \{t_i\}_{i=0}^{n-1} \subseteq [0, T]$.

$$E[(\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2 - T)^2]$$

$$= E[(\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2)^2 - 2T\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2 + T^2]$$

Using $(\sum_i a_i)^2 = \sum_i a_i^2 + 2\sum_i \sum_{j<i} a_i a_j$ we can simplify the expression by linearity and independence.

$$= \sum_{i=0}^{n-1} E[(B_{t_{i+1}} - B_{t_i})^4] + 2\sum_{i=0}^{n-1}\sum_{j=0}^{i-1} E[(B_{t_{i+1}} - B_{t_i})^2]E[(B_{t_{j+1}} - B_{t_j})^2]$$

$$-2T\sum_{i=0}^{n-1} E[(B_{t_{i+1}} - B_{t_i})^2] + T^2$$

We will now use that if $X \sim N(0, \sigma^2)$, then $E[X^2] = \sigma^2$ and $E[X^4] = 3\sigma^4$.

$$= \sum_{i=0}^{n-1} 3(t_{i+1} - t_i)^2 + 2\sum_{i=0}^{n-1}\sum_{j=0}^{i-1}(t_{i+1} - t_i)(t_{j+1} - t_j) - 2T\sum_{i=0}^{n-1}(t_{i+1} - t_i) + T^2$$

We apply $(\sum_j a_j)^2 = \sum_j a_j^2 + 2\sum_j \sum_{i<j} a_j a_i$ in reverse and notice the following telescoping sum $\sum_{i=0}^{n-1}(t_{i+1} - t_i) = T$.

$$= 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2 + (\sum_{i=0}^{n-1}(t_{i+1} - t_i))^2 - 2T^2 + T^2 = 2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2$$

Finally we bound the sum by its mesh.

$$2\sum_{i=0}^{n-1}(t_{i+1} - t_i)^2 \leq 2|\Pi|\sum_{i=0}^{n-1}(t_{i+1} - t_i) = 2|\Pi|T \to 0, \qquad |\Pi| \to 0.$$

The random variable $(\sum_{i=0}^{n-1}(B_{t_{i+1}} - B_{t_i})^2 - T)^2$ is nonnegative by construction and since it approaches 0, it must be equal to 0 in $L^2(P)$.

Example 2.4.10 (Brownian motion has unbounded total variation) Assume that Brownian motion has bounded variation. The following estimate shows that this can not be the case as

$$V^2_{[0,T]}(\Pi, B_t) = \sum_{i=0}^{n-1} |B_{t_{i+1}} - B_{t_i}|^2 \leq \max_{0 \leq i \leq n-1} |B_{t_{i+1}} - B_{t_i}| V^1_{[0,T]}(\Pi, B_t).$$

Since Brownian motion has continuous paths $\max_{0 \leq i \leq n-1} |B_{t_{i+1}} - B_{t_i}| \to 0$ as $|\Pi| \to 0$. This would mean that the quadratic variation must also converge to zero, which is a clear contradiction! Thus our assumption about bounded variation must be false.

Definition 2.4.11 (Stopping time) We say that a mapping $\tau : \Omega \to [0, \infty]$ is a stopping time if

$$\{\tau \leq t\} \in \mathcal{F}_t,$$

for all $t \geq 0$.

Intuitively speaking, a stopping time is a rule for when to make a decision without requiring knowledge of future events.

Definition 2.4.12 (Càdlàg process) A process $X = \{X_t\}_{t \geq 0}$ is called a càdlàg process if it has right-continuous paths and existing left limits everywhere.

Definition 2.4.13 (Martingale, submartingale and supermartingale) An integrable and adapted process X is called

(i) a martingale if for all $s \leq t$, then $E[X_t|\mathcal{F}_s] = X_s$ a.e.
(ii) a submartingale if for all $s \leq t$, then $E[X_t|\mathcal{F}_s] \geq X_s$ a.e.
(iii) a supermartingale if for all $s \leq t$, then $E[X_t|\mathcal{F}_s] \leq X_s$ a.e.

It is useful to think of martingales as a process modeling a fair game. For example, tossing an unbiased coin where we gain 1$ for each heads and lose 1 $ for each tails. On the other hand, submartingales and supermartingales can be used to model profitable or disadvantageous games respectively.

Example 2.4.14 (Brownian motion is a martingale) We will prove the martingale property of the Brownian motion.

$$E[B_t|\mathcal{F}_s] = E[B_t - B_s + B_s|\mathcal{F}_s] = E[B_t - B_s|\mathcal{F}_s] + E[B_s|\mathcal{F}_s]$$

$$= E[B_t - B_s] + B_s = B_s$$

Where we have added and subtracted B_s in the first equality. Linearity of conditional expectations in the second equality. Independent increments and measurability of B_s in the third equality. Normally distributed increments with mean zero in the last equality.

Definition 2.4.15 (Predictable process) A stochastic process $A : [0, \infty) \times \Omega \to \mathbb{R}$ is called predictable with respect to a filtration $\{\mathcal{F}_t\}_{t\geq 0}$ if A is measurable with respect to the predictable σ-algebra given by

$$\mathcal{P} = \sigma\Big((s,t] \times F : s \leq t, F \in \mathcal{F}_s\Big).$$

Theorem 2.4.16 (Doob-Meyer decomposition) Let $\Pi = \{t_i\}_{i=0}^{n-1} \subseteq [0,t]$ be a partition and let $M = \{M_t\}_{t\geq 0}$ be a square integrable martingale. Then there exists a unique stochastic process $A = \{A_t\}_{t\geq 0}$ which is increasing, predictable and continuous such that

(i) $A_0 = 0$,

(ii) $M_t^2 - A_t$ is a martingale for all $t \geq 0$,

(iii) $A_t = \lim_{|\Pi \to 0|} \sum_{i=0}^{n-1} |M_{t_{i+1}} - M_{t_i}|^2$ in probability.

We call A the predictable compensator of M. Note that A can be computed as the quadratic variation of M, which is often written with the following notation

$$A_t = [M, M]_t = [M]_t. \tag{2.4.1}$$

Definition 2.4.17 (Local martingale) An adapted càdlàg process M $= \{M_t\}_{t\geq 0}$ is a local martingale if there are increasing stopping times τ_n, $n \in \mathbb{N}$ with $P(\lim_{n\to\infty} \tau_n = \infty) = 1$ such that

(i) the process stopped on $\{\tau_n > 0\}$

$$M_t^{\tau_n} = M_{t\wedge\tau_n} 1_{\{\tau_n > 0\}}$$

is a martingale for all n, that is

$$E[M_t^{\tau_n} | \mathcal{F}_s] = M_s^{\tau_n}, \qquad s \leq t.$$

(ii) And that the stopped process satisfies uniform integrability. That is

$$\lim_{m\to\infty} \sup_{t\geq 0} E[|M_t^{\tau_n}| 1_{\{|M_t^{\tau_n}| \geq m\}}] = 0.$$

Definition 2.4.18 (Semimartingale) An $\mathcal{F}$-adapted càdlàg process X is a semimartingale if

$$X_t = X_0 + A_t + M_t, \qquad t \geq 0.$$

Where A and M are càdlàg adapted processes such that A is of bounded variation with probability one and M is a local martingale.

Semimartingales are more informally known as "good integrators", that is processes for which one can define stochastic integrals. In the next chapter, we will review the construction of the Itô integral, i.e. with respect to a Brownian motion, which is one of the simplest examples of semimartingales. However, the construction of the stochastic integral can be extended to any general martingale.

2.5 Itô calculus

This section aims to construct Itô's integral which is a type of stochastic integral. The construction will be done for a general martingale integrator for the sake of generality before tackling the special case with a Brownian motion. We will also illuminate some application of Itô's calculus for solving stochastic differential equations. The source material has been gathered from [Øks03] and [Bal17].

Construction of the stochastic integral

We would like to define a stochastic integral of a process X with respect to a martingale M often denoted by

$$\int_0^T X_s dM_s.$$

However this is not a simple task as it gives rise to the following questions which we will tackle one by one:

(i) What does integration with respect to a martingale mean?
(ii) Is the integral well-defined and in what sense?
(iii) What types of problems can we solve with stochastic integrals?

It is natural to wonder if we can interpret the stochastic integral as a Riemann-Stieltjes integral with respect to a trajectory of a martingale. Unfortunately, this is not well-defined as the paths of a martingale are not of bounded variation. Which is a common assumption in constructing the Riemann-Stieltjes integral. Therefore, it is necessary to make sense of the stochastic integral in a different way and see its limitations.

Definition 2.5.1 (M^p-spaces) Let $[a, b] \subset [0, \infty)$ and $p \geq 1$. Then we define $M^p[a, b]$ as the space of equivalent classes of predictable processes such that

$$E\Big[\int_a^b |X_s|^p d[M]_s\Big] < \infty,$$

where $[M]$ denotes the quadratic variation of M as defined in (2.4.1).
We will mostly work with $M^2[a, b] \subset L^2\big([a, b] \times \Omega, \mathcal{B}[a, b] \otimes \mathcal{A}, d[M]_t \times P\big)$.

Definition 2.5.2 (Elementary processes) Let $\Pi = \{t_i\}_{i=0}^{n-1}$ be a partition of $[a, b]$ and let $\varphi_i : \Omega \to \mathbb{R}$ be bounded $\mathcal{F}_{t_i}$-measurable random variables. Then we define an elementary process $X : \Omega \times [a, b] \to \mathbb{R}$ by

$$X_t = \sum_{i=0}^{n-1} \varphi_i 1_{[t_i, t_{i+1})}(t), \qquad n \in \mathbb{N}.$$

Moreover, we define the stochastic integral of an elementary process with respect to a martingale process $M \in L^2(\Omega, \mathcal{A}, P)$ by

$$\int_a^b X_s dM_s = \sum_{i=0}^{n-1} \varphi_i (M_{t_{i+1}} - M_{t_i}).$$

Lemma 2.5.3 (Existence of an approximating sequence) Let $\Pi = \{t_i\}_{i=0}^{n-1}$ be a partition of $[a,b]$ and let for all i, $\{\varphi_{i,n}\}_{n\in\mathbb{N}}$ be a sequence of bounded $\mathcal{F}_{t_i}$-measurable random variables. If $\phi \in M^2[a,b]$, then there exists an approximating sequence of elementary processes $\{\phi_n\}_{n\in\mathbb{N}} \subseteq M^2[a,b]$ of the form

$$\phi_n = \sum_{i=0}^{n-1} \varphi_{i,n} 1_{[t_i,t_{i+1})}(t), \qquad n \in \mathbb{N},$$

such that

$$||\phi - \phi_n||_{L^2(P\times d[M]_t)} \to 0, \qquad n \to \infty.$$

Proposition 2.5.4 (Itô isometry for elementary processes) Let $\{\phi_n\}_{n\in\mathbb{N}}$ be an elementary process and $M \in L^2(P)$ a martingale. Then

$$E\Big[\Big(\int_a^b \phi_n dM_s\Big)^2\Big] = E\Big[\int_a^b \phi_n^2 d[M]_s\Big].$$

Proof. By definition we have that

$$E\Big[\Big(\int_a^b \phi_n dM_s\Big)^2\Big] = E\Big[\Big(\sum_{i=0}^{n-1} \varphi_{i,n}(M_{s_{i+1}} - M_{s_i})\Big)^2\Big]$$

$$= E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 (M_{s_{i+1}} - M_{s_i})^2\Big] \tag{2.5.1}$$

$$+2E\Big[\sum_{i=0}^{n-1}\sum_{j<i} \varphi_{i,n}(M_{s_{i+1}} - M_{s_i})\varphi_{j,n}(M_{s_{j+1}} - M_{s_j})\Big]. \tag{2.5.2}$$

We expand the square in (2.5.1) and use the martingale property such that

$$E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 (M_{s_{i+1}} - M_{s_i})^2\Big] = E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 (M_{s_{i+1}}^2 + M_{s_i}^2 - 2M_{s_{i+1}}M_{s_i})\Big]$$

$$= E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 (M_{s_{i+1}}^2 + M_{s_i}^2 - 2M_{s_i}E[M_{s_{i+1}}|\mathcal{F}_{s_i}])\Big]$$

$$= E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 (M_{s_{i+1}}^2 - M_{s_i}^2)\Big].$$

By Doob-Meyer decomposition we can write a martingale as the difference between a submartingale and the predictable compensator. This yields

$$E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 (M_{s_{i+1}}^2 - M_{s_i}^2)\Big]$$

$$= E\Big[\sum_{i=0}^{n-1} \varphi_{i,n}^2 \Big(M_{s_{i+1}}^2 - [M]_{s_{i+1}} - (M_{s_i}^2 - [M]_{s_i}) + [M]_{s_{i+1}} - [M]_{s_i}\Big)\Big]$$

$$= E\Big[\sum_{i=0}^{n-1}\varphi_{i,n}^2\Big(M_{s_{i+1}} - M_{s_i} + [M]_{s_{i+1}} - [M]_{s_i}\Big)\Big]$$

$$= E\Big[\sum_{i=0}^{n-1}\varphi_{i,n}^2\Big(E[M_{s_{i+1}} - M_{s_i}|\mathcal{F}_{s_i}] + [M]_{s_{i+1}} - [M]_{s_i}\Big)\Big]$$

$$= E\Big[\sum_{i=0}^{n-1}\varphi_{i,n}^2\Big([M]_{s_{i+1}} - [M]_{s_i}\Big)\Big] = E\Big[\int_a^b \phi_n d[M]_s\Big].$$

The result now follows since (2.5.2) is zero. Indeed

$$2E\Big[\sum_{i=0}^{n-1}\sum_{j<i}\varphi_{i,n}(M_{s_{i+1}} - M_{s_i})\varphi_{j,n}(M_{s_{j+1}} - M_{s_j})\Big]$$

$$= 2E\Big[\sum_{i=0}^{n-1}\sum_{j<i}\varphi_{i,n}E[M_{s_{i+1}} - M_{s_i}|\mathcal{F}_{s_i}]\varphi_{j,n}(M_{s_{j+1}} - M_{s_j})\Big] = 0.$$

■

We can now define the stochastic integral using a limit argument of elementary processes.

Step 1. (Existence) By Lemma 2.5.3 we know that the elementary functions are dense in $M^2[a,b]$.

Step 2. (Convergence) Let $I(\phi_n) \in L^2(P)$ be a sequence of stochastic integrals of elementary processes given by

$$I(\phi_n) = \sum_{i=0}^{n-1}\varphi_{i,n}(M_{s_{i+1}} - M_{s_i}),$$

then $\{I(\phi_n)\}_{n\in\mathbb{N}}$ is a convergent sequence, i.e. the following $L^2(P)$ limit exists

$$I(\phi_n) \to I(\phi), \qquad n \to \infty.$$

Since $L^2(P)$ is a Hilbert space, we know that all Cauchy sequences converge by completeness. Thus it suffices to show that $\{I(\phi_n)\}_{n\in\mathbb{N}}$ is a Cauchy sequence in order to prove convergence and existence of the limit. For $n, m \in \mathbb{N}$ we have that

$$E\Big[\Big(I(\phi_n) - I(\phi_m)\Big)^2\Big]$$

$$= E\Big[\Big(\sum_{i=0}^{n-1}\phi_{i,n}(M_{s_{i+1}} - M_{s_i}) - \sum_{i=0}^{n-1}\phi_{i,m}(M_{s_{i+1}} - M_{s_i})\Big)^2\Big]$$

$$= E\Big[\Big(\sum_{i=0}^{n-1}(\phi_{i,n} - \phi_{i,m})(M_{s_{i+1}} - M_{s_i})\Big)^2\Big]$$

$$= E\Big[\int_a^b (\varphi_n - \varphi_m)^2 d[M]_s\Big] = ||\varphi_n - \varphi_m||^2_{L^2(P\times d[M]_s)} \to 0, \qquad m, n \to \infty.$$

In the computation above, we have used linearity, Itô isometry and that $\{\phi_n\}_{n\in\mathbb{N}}$ is converging to ϕ in $L^2(P \times [M]_s)$. Hence we have a Cauchy sequence.

Step 3. (Well-defined limit) We must show that the stochastic integral is well-defined as the limit of a stochastic integral of elementary processes. This means that

$$I(\phi_n) \to I(\phi), \qquad n \to \infty,$$

is independent of the choice we make for the approximating sequence of elementary processes.

Let $\{\phi_n\}_{n\in\mathbb{N}}, \{\phi'_n\}_{n\in\mathbb{N}} \in L^2(P \times d[M]_s)$ be two distinct sequences of elementary processes that approximate $\phi \in L^2(P \times d[M]_s)$. Then

$$E\Big[\Big(I(\phi_n) - I(\phi'_n)\Big)^2\Big] = E\Big[\Big(I(\phi_n - \phi'_n)\Big)^2\Big] = E\Big[\int_a^b (\phi_n - \phi'_n)^2 d[M]_s\Big]$$

$$= E\Big[\int_a^b (\phi_n - \phi + \phi - \phi'_n)^2 d[M]_s\Big]$$

$$\leq 2E\Big[\int_a^b (\phi_n - \phi)^2 d[M]_s\Big] + 2E\Big[\int_a^b (\phi - \phi'_n)^2 d[M]_s\Big] \to 0, \qquad n \to \infty,$$

where we have used the following inequality

$$|x + y|^p \leq 2^{p-1}(|x|^p + |y|^p), \qquad x, y \in \mathbb{R}, \qquad p \geq 1.$$

This concludes the construction of the stochastic integral for a martingale integrator. But we will also mention some important properties of the stochastic integral.

Proposition 2.5.5 (Linearity) Let $X, Y \in M^2[a, b]$ and let $M \in L^2(P)$ be a martingale, then

$$\int_a^b (xX_s + yY_s)dM_s = x\int_a^b X_s dM_s + y\int_a^b Y_s dM_s, \qquad x, y \in \mathbb{R}.$$

Proof. We start by showing the result for elementary processes.

$$\int_a^b (xX_s + yY_s)dM_s = \sum_{i=0}^{n-1} \Big(x\varphi_i(M_{s_{i+1}} - M_{s_i}) + y\varphi'_i(M_{s_{i+1}} - M_{s_i})\Big)$$

$$= x\sum_{i=0}^{n-1}\Big(\varphi_i(M_{s_{i+1}}-M_{s_i})\Big)+y\sum_{i=0}^{n-1}\Big(\varphi_i'(M_{s_{i+1}}-M_{s_i})\Big)$$

$$= x\int_a^b X_s dM_s + y\int_a^b Y_s dM_s.$$

The result for a general $X, Y \in M^2[a,b]$ follows by applying the limit argument for elementary processes.

■

Proposition 2.5.6 (Itô isometry) Let $X, Y \in M^2[a,b]$ and let $M \in L^2(P)$ be a martingale, then

$$E\Big[\Big(\int_a^b X_s dM_s\Big)^2\Big] = E\Big[\int_a^b X_s^2 d[M]_s\Big], \tag{2.5.3}$$

$$E\Big[\int_a^b X_s dM_s \int_a^b Y_s dM_s\Big] = E\Big[\int_a^b X_s Y_s d[M]_s\Big]. \tag{2.5.4}$$

Proof. (2.5.3) follows from our previous proof of Itô's isometry using elementary processes and then applying the limit argument which we have constructed.

In order to prove (2.5.4) we use the general Itô isometry together with the polarization identity

$$xy = \frac{1}{4}\Big((x+y)^2-(x-y)^2\Big), \qquad x, y \in \mathbb{R}.$$

This yield

$$E\Big[\int_a^b X_s dM_s \int_a^b Y_s dM_s\Big]$$

$$= E\left[\frac{1}{4}\left(\Big(\int_a^b X_s dM_s + \int_a^b Y_s dM_s\Big)^2 - \Big(\int_a^b X_s dM_s - \int_a^b Y_s dM_s\Big)^2\right)\right]$$

$$= E\left[\frac{1}{4}\left(\Big(\int_a^b (X_s+Y_s) dM_s\Big)^2 - \Big(\int_a^b (X_s-Y_s) dM_s\Big)^2\right)\right]$$

$$= E\left[\frac{1}{4}\left(\int_a^b (X_s+Y_s)^2 d[M]_s - \int_a^b (X_s-Y_s)^2 d[M]_s\right)\right]$$

$$= E\left[\frac{1}{4}\left(\int_a^b \Big((X_s+Y_s)^2-(X_s-Y_s)^2 d[M]_s\Big)\right)\right]$$

$$= E\left[\frac{1}{4}\int_a^b 4X_sY_s d[M]_s\right] = E\left[\int_a^b X_sY_s d[M]_s\right]$$

■

Proposition 2.5.7 (Mean zero) Let $X \in M^2[a,b]$ and let $M \in L^2(P)$ be a martingale, then

$$E\Big[\int_a^b X_s dM_s\Big] = 0.$$

Proof. We once again start by showing the result for elementary processes.

$$\begin{aligned} E\Big[\int_a^b X_s dM_s\Big] &= E\Big[\sum_{i=0}^{n-1} \varphi_i (M_{s_{i+1}} - M_{s_i})\Big] \\ &= E\Big[\sum_{i=0}^{n-1} \varphi_i E[M_{s_{i+1}} - M_{s_i} | \mathcal{F}_{s_i}]\Big] = 0. \end{aligned}$$

The result for a general $X \in M^2[a,b]$ follows by applying the limit argument for elementary processes.

■

Stochastic differential equations

Now that we have defined the stochastic integral for a general martingale integrator, we will restrict the analysis to the Brownian motion. We have already seen that the stochastic integral is well-defined. If the integrator of a stochastic integral is a Brownian motion, we get the Itô integral, i.e.

$$\int_a^b X_s dB_s.$$

The Brownian motion has independent increments and constant mean, which implies that it is a martingale. In particular, having the property of independent increments simplifies many proofs from the construction of the stochastic integral as for example the proof of Proposition 2.5.4, but what is really crucial for the construction of a stochastic integral is the martingale property.

Now we will turn to defining the stochastic differential of Brownian motion. The naive candidate for a stochastic differential $\frac{dB_t}{dt}$ does not work as Brownian motion is almost surely nowhere differentiable. However, we will define the stochastic differential when working with an Itô process.

Definition 2.5.8 (Itô process) Let u be a stochastic process with a.s. integrable trajectories and v a stochastic process such that

$$E\Big[\int_a^b |v(s)|^2 ds\Big] < \infty.$$

Then the well-defined process

$$X_t = X_0 + \int_a^b u(s)ds + \int_a^b v(s)dB_s, \qquad s \in [a,b], \qquad X_0 \in \mathbb{R},$$

is called an Itô process. If we choose deterministic functions b, σ such that

$$u(s) = b(s, X_s), \qquad v(s) = \sigma(s, X_s),$$

we obtain the relation

$$X_t = X_0 + \int_a^b b(s, X_s)ds + \int_a^b \sigma(s, X_s)dB_s, \qquad s \in [a,b], \qquad X_0 \in \mathbb{R}. \quad (2.5.5)$$

The above definition makes sense since it is a semimartingale or a sum of a Lebesgue integral and an Itô integral. However, it is often convenient to write the equation in differential form

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t, \qquad t \in [a,b]. \quad (2.5.6)$$

This is just a matter of notation and terminology, the differential form in (2.5.6) has no other meaning than the integral form in (2.5.5). Equation (2.5.6) is called a stochastic differential equation, abbreviated to SDE, and is used to model the dynamics of a random system which evolves in time. Moreover, it is common to refer to $b(t, X_t)$ as the drift and $\sigma(t, X_t)$ as the diffusion of the process.

Theorem 2.5.9 (Itô's formula) Let $f \in C^{1,2}([0,T] \times \mathbb{R})$ and X an Itô process. Then

$$f(t,X_t) = f(0,X_0) + \int_0^t \Big(\frac{\partial}{\partial s} f(s,X_s)\Big) ds + \int_0^t \Big(\frac{\partial}{\partial x} f(s,X_s)\Big) dX_s$$
$$+\frac{1}{2}\int_0^t \Big(\frac{\partial^2}{\partial x^2} f(s,X_s)\Big) d[X,X]_s.$$

Or in differential form

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

Itô's formula can be thought as the chain rule for stochastic calculus and is often the key to solving stochastic differential equations.

Example 2.5.10 (Ornstein-Uhlenbeck) The solution of the following SDE is known as the Ornstein-Uhlenbeck process

$$dX_t = -\lambda X_t dt + \sigma dB_t,$$
$$X_0 = x_0,$$

where $\lambda, \sigma > 0$ and $x_0 \in \mathbb{R}$. The solution of this SDE can be obtained by Itô's formula using $f(t,X_t) = e^{\lambda t} X_t$. It is then easy to check that

$$\frac{\partial}{\partial t} f = \lambda f, \qquad \frac{\partial}{\partial x} f = e^{\lambda t}, \qquad \frac{\partial^2}{\partial x^2} f = 0, \qquad d[X,X]_t = \sigma^2 dt.$$

Inserting into Itô's formula we see that all the drift terms vanish and we are left with a diffusion term of the form

$$df = \sigma e^{\lambda t} dB_t.$$

Integrating on both sides we get

$$e^{\lambda t} X_t - X_0 = \int_0^t \sigma e^{\lambda s} dB_s.$$

Finally we isolate X_t and make use of the initial condition to get the final answer

$$X_t = x_0 e^{-\lambda t} + e^{-\lambda t} \int_0^t \sigma e^{\lambda s} dB_s.$$

From the last expression, it is easy to calculate the mean and variance of the process. Indeed

$$E[X_t] = E\Big[x_0 e^{-\lambda t} + e^{-\lambda t} \int_0^t \sigma e^{\lambda s} dB_s\Big]$$
$$= x_0 e^{-\lambda t} + e^{-\lambda t} E\Big[\int_0^t \sigma e^{\lambda s} dB_s\Big] = x_0 e^{-\lambda t},$$

where we have used that the Itô integral of a deterministic function has expectation zero. Moreover,

$$Var(X_t) = E\Big[\Big(e^{-\lambda t}\int_0^t \sigma e^{\lambda s} dB_s\Big)^2\Big]$$
$$= e^{-2\lambda t}\sigma^2 \int_0^t e^{2\lambda s} ds = \frac{\sigma^2}{2\lambda}(1-e^{-2\lambda t}),$$

where we have used Itô isometry in the second equality to arrive at a deterministic integral.

The solution process we found for the Ornstein-Uhlenbeck SDE is called a strong solution. We will end this section by defining the concept of a strong solution here below.

Definition 2.5.11 (Strong solution) A stochastic process $\{X_t\}_{t\geq 0}$ is called a strong solution of the following SDE

$$dX_t = b(t, X_t)dt + \sigma(t, X_t)dB_t$$
$$X_0 = x_0,$$

whenever the following conditions are satisfied.

(i) $\{X_t\}_{t\geq 0}$ is adapted to $\mathcal{F}_t$ on $(\Omega, \mathcal{A}, P)$ for all standard Brownian motions.

(ii) $P\Big(X_0 = x_0\Big) = 1$.

(iii) $P\Big(\int_0^t |b(s, X_s)|ds < \infty\Big) = 1$.

(iv) $P\Big(\int_0^t |\sigma(s, X_s)|^2 ds < \infty\Big) = 1$.

(v) $P\Big(X_t = x_0 + \int_0^t b(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s\Big) = 1$.

We end this section by stating an important result regarding existence and uniqueness of strong solutions.

Theorem 2.5.12 (Existence and uniqueness of strong solutions) Assume that the drift and diffusion of a SDE satisfies the global Lipschitz property and have at most linear growth. That is

$$|b(t,x) - b(t,y)| + |\sigma(t,x) - \sigma(t,y)| \leq L|x-y|,$$

for all $t \in [0,T], x, y \in \mathbb{R}$ and $L > 0$.

$$|b(t,x)| + |\sigma(t,x)| \leq C(1+|x|),$$

where $C > 0$.

If

$$E[X_0^2] < \infty,$$

then there exists a unique strong solution of SDE such that $E[X_t^2] < \infty$, for all $t \in [0,T]$.

2.6 Markov chains

This section will give a brief introduction to Markov chains and some of its applications to actuarial modelling in life insurance. The material is mostly gathered from [Kol12] with a few alternation of notation to better fit our framework.

Definition 2.6.1 (Markov chain with finite state space) Let X_t be a stochastic process taking values on a finite metric space S. We say that X_t is a Markov chain if

$$P(X_{t_{n+1}} = i_{n+1}|X_{t_1} = i_1, ..., X_{t_n} = i_n) = P(X_{t_{n+1}} = i_{n+1}|X_{t_n} = i_n)$$

for all $t_1 < t_2 < ... < t_{n+1} \in \mathbb{R}$ and $i_1, i_2, ..., i_{n+1} \in$ S with

$$P(X_{t_1} = i_1, ..., X_{t_n} = i_n) \neq 0.$$

One says that a Markov chain is "memoryless" since the future transition is independent of the past transitions. All that matters is the present state of the Markov chain.

Definition 2.6.2 (Transition probability). We say that a function is a transition probability if

$$p_{ij}(s, t) = P(X_t = j|X_s = i), \quad s \leq t, \quad i, j \in S.$$

Here, $p_{ij}(s, t)$ denotes the probability that X will be in state j at time t given that X was in state i at a previous time s.

Example 2.6.3 (Markov chain in insurance) We will apply the Markov chain to model the health of the insured in continuous time. Using actuarial notation we can define $S = \{*, \diamond, \dagger\}$ where $*$ means the insured is in good health, $\diamond$ signifies that the insured is sick or disabled and $\dagger$ means that the insured is deceased.

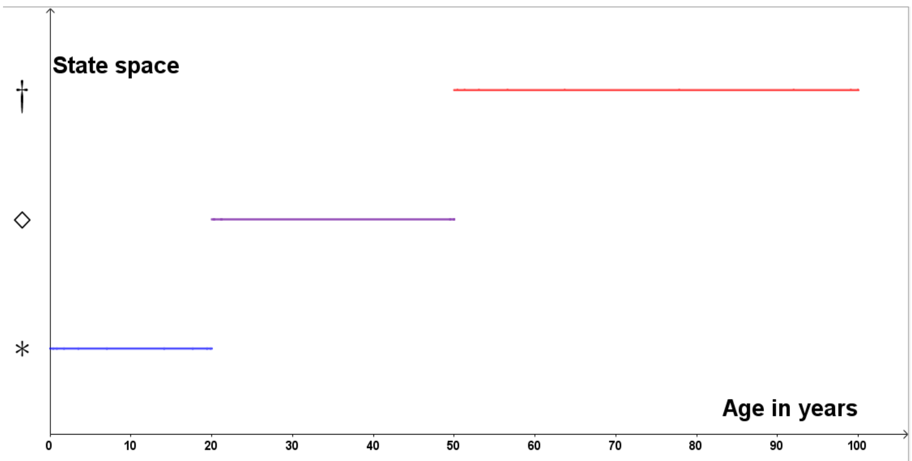

Figure 2.1: A sample path of the Markov chain.

We can see from the figure that the insured has been healthy up to the age of 20. Then some event happened which made him permanently disabled, this could be due to an accident or an underlying disease. Lastly at the age of 50 the individual has perished. With a life insurance policy each transition would initiate an insurance payout determined by the contract.

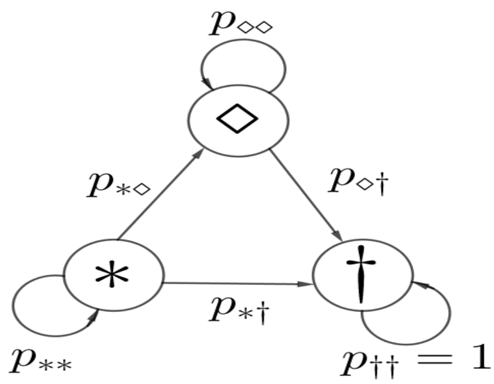

Figure 2.2: A transition diagram of the Markov chain.

By convention we only draw the transitions which can happen with a nonzero probability between time s and time t. We will also assume that once the insured dies, there is no possibility of resurrection. We say that state † is an absorbing state of the Markov chain since once the insured has entered this state, then the insured cannot leave it.

Definition 2.6.4 (Transition probability matrix) A matrix $P(s,t) = \{p_{ij}(s,t)\}_{i,j\in S}$ is called a transition probability matrix if

(i) $p_{ij}(s,t) \geq 0$,

(ii) $\sum_{j\in S} p_{ij}(s,t) = 1$, for all $i \in S$,

(iii) $p_{ij}(s,s) = 1_{i=j}$, whenever $P(X_s = i) \neq 0$.

Theorem 2.6.5 (Chapman-Kolmogorov equation) Let $X = \{X_t\}_{t\in\mathbb{R}}$ be a Markov chain with transition probability matrix $P(s,t)$. Then the following equation holds

$$P(s,t) = P(s,u)P(u,t), \qquad s \leq u \leq t.$$

Chapman-Kolmogorov gives us a way of decomposing transition probabilities into transition probabilities going through a middle time step.

Definition 2.6.6 (Transition rates) Let $X = \{X_t\}_{t\in\mathbb{R}}$ be a Markov chain with finite state space S. The transition rates are functions μ_i, μ_{ij} defined by

$$\mu_i(t) = \lim_{h\to 0^+} \frac{1 - p_{ii}(t, t+h)}{h}, \qquad t \in \mathbb{R},\ i \in S,$$

$$\mu_{ij}(t) = \lim_{h\to 0^+} \frac{p_{ij}(t, t+h)}{h}, \qquad t \in \mathbb{R},\ i, j \in S,\ i \neq j,$$

whenever the limits exist and are finite. Moreover, we define μ_{ii} by

$$\mu_{ii}(t) = -\mu_i(t), \qquad \text{for all } i \in S.$$

Definition 2.6.7 (Regular Markov chain) We say that a Markov chain $X = \{X_t\}_{t\in\mathbb{R}}$ is regular if the the transition rates $\mu_i(t)$, $\mu_{ij}(t)$ exist as continuous functions of t.

Remark 2.6.8 (Connecting rates and probabilities) In insurance, transition probabilities play a vital role in pricing policies. However they are not directly observable as in the case of transition rates. The key relation is to notice that transition rates are derivatives of the transition probabilities. Indeed for $i \neq j$

$$\mu_{ij}(t) = \lim_{h\to 0^+} \frac{p_{ij}(t, t+h)}{h} = \lim_{h\to 0^+} \frac{p_{ij}(t, t+h) - p_{ij}(t, t)}{h} = \frac{d}{ds} p_{ij}(s, t)\Big|_{s=t}.$$

Thus we may interpret $\mu_{ij}(t)h \approx p_{ij}(t, t+h)$ for infinitesimal small $h > 0$. This means that $\mu_{ij}(t)h$ is the probability of entering state j at time $t+h$ given that we are in state i at time t. We may informally say that $\mu_{ij}(t)$ is the "speed" of the transition. Similarly $\mu_i(t)h$ can be understood as the probability of leaving state i in the infinitesimal time interval $[t, t+h]$.

Definition 2.6.9 (Homogeneous Markov chain) We say that a Markov chain $X = \{X_t\}_{t\in\mathbb{R}}$ is homogeneous if it is time homogeneous. This means that the following equation hold for all $s, t \in \mathbb{R}$, $h > 0$ and $i, j \in S$ whenever $P(X_s = i) > 0$ and $P(X_t = i) > 0$

$$P(X_{s+h} = j | X_s = i) = P(X_{t+h} = j | X_t = i).$$

Example 2.6.10 (Homogeneous Markov chains) Intuitively, the transition probabilities in a homogeneous Markov chain only depend on the length of the time increment and not on the starting time. For instance, the probability of tossing heads on a fair coin is a homogeneous Markov chain. On the other hand, the probability of an individual surviving one more year certainly is not a homogeneous Markov chain since we know that mortality increases exponentially with age.

Definition 2.6.11 (Generator matrix) Let $X = \{X_t\}_{t\in\mathbb{R}}$ be a homogeneous Markov chain with finite state space. We say that the matrix $\Lambda(t) = \{\mu_{ij}(t)\}_{i,j\in S}$ is the generator matrix of X.

Observation 2.6.12 Indeed the generator matrix generates the behavior of the Markov chain. In particular $\Lambda(0)$ is given by the equation

$$\Lambda(0) = \lim_{h\to 0^+} \frac{P(h) - Id_n}{h}.$$

Where Id_n denotes the identity matrix of dimension n. Using the equation above together with a Taylor expansion we can completely reconstruct $P(t)$ by

$$P(t) = exp(t\Lambda(0)) = \sum_{n=0}^{\infty} \frac{t^n}{n!}\Lambda(0)^n.$$

Theorem 2.6.13 (Kolmogorov's differential equations) For a continuous time regular Markov chain $X = \{X_t\}_{t\in\mathbb{R}}$ with finite state space S, the following statements holds:

(Kolmogorov's backward equation)

$$\frac{d}{ds}P(s,t) = -\Lambda(s)P(s,t), \qquad s,t \in \mathbb{R}, \qquad s < t.$$

(Kolmogorov's forward equation)

$$\frac{d}{dt}P(s,t) = P(s,t)\Lambda(t), \qquad s,t \in \mathbb{R}, \qquad s < t.$$

Proof. We start with Kolmogorov's backward equation. Let $h > 0$, such that $s < s + h < t$. Then for small h:

$$\frac{P(s+h,t) - P(s,t)}{h} = \frac{1}{h}(P(s+h,t) - P(s,s+h)P(s+h,t))$$

$$= \frac{Id_n - P(s,s+h)}{h}P(s+h,t) \longrightarrow -\Lambda(s)P(s,t).$$

Where we have used Chapman-Kolmogorov's equation in the first equality and continuity of matrix multiplication to evaluate the limit as $h \longrightarrow 0^+$.
Now we do the same for Kolmogorov's forward equation.

$$\frac{P(s,t+h) - P(s,t)}{h} = \frac{1}{h}(P(s,t)P(t,t+h) - P(s,t))$$

$$= P(s,t)\frac{P(t,t+h) - Id_n}{h} \longrightarrow P(s,t)\Lambda(t).$$

Where we again have used Chapman-Kolmogorov's equation in the first equality and continuity of matrix multiplication to evaluate the limit as $h \longrightarrow 0^+$.

■

Theorem 2.6.14 (Calculating $p_{jj}(s,t)$) Let $X = \{X_t\}_{t\in\mathbb{R}}$ be a regular Markov chain. Then

$$p_{jj}(s,t) = exp\bigg(- \sum_{k\neq j} \int_s^t \mu_{jk}(u)du \bigg), \qquad \forall s \leq t,$$

whenever $P(X_s = j) > 0$.

Proof. The result follows from Kolmogorov's equations. The complete proof can be found in [Kol12].

■

CHAPTER 3

Fractional noise

This chapter will deal with fractional noise processes. First, we will introduce the fractional Brownian motion which we will then use to construct the fractional Ornstein-Uhlenbeck process.

3.1 Fractional Brownian motion

This section will offer a short introduction to a stochastic process known as the fractional Brownian motion. It is a generalisation of the Brownian motion in the sense that we are dealing with a centered Gaussian process. The main difference between these two processes is that fractional Brownian motion does not have independent increments. For a thorough guide to fractional Brownian motion and advanced stochastic calculus the reader can consult [Nua95] which we will follow in this section.

Definition 3.1.1 (Fractional Brownian motion) A centered Gaussian process starting from zero, $B^H = \{B_t^H, t \geq 0\}$ is called a fractional Brownian motion, abbreviated to fBm, of Hurst parameter $H \in (0,1)$ if the covariance function is given by

$$R_H(s,t) = Cov(B_s^H, B_t^H) = E[B_s^H B_t^H] = \frac{1}{2}(t^{2H} + s^{2H} - (t-s)^{2H}) \qquad s \leq t.$$

Proposition 3.1.2 (fBm's covariance function is homogeneous) For any $\alpha > 0$ the covariance function of fBm is homogeneous of degree $2H$. Meaning that the following equation holds

$$Cov(B_{\alpha s}^H, B_{\alpha t}^H) = \alpha^{2H} Cov(B_s^H, B_t^H).$$

Proof. Indeed this is easy to check by direct computation

$$Cov(B_{\alpha s}^H, B_{\alpha t}^H) = \frac{1}{2}\Big((\alpha t)^{2H} + (\alpha s)^{2H} - (\alpha t - \alpha s)^{2H}\Big)$$

$$= \alpha^{2H} \frac{1}{2}\Big(t^{2H} + s^{2H} - (t-s)^{2H}\Big) = \alpha^{2H} Cov(B_s^H, B_t^H).$$

■

Corollary 3.1.3 (fBm is self-similar of order H) For any $\alpha > 0$ the law of B_t^H and $\alpha^{-H} B_{\alpha t}^H$ are equal.
Proof. Recall that Gaussian distributions are uniquely determined by their mean and covariance function, hence it suffices to check that they are equal. The expectation is easy,

$$E[\alpha^{-H} B_{\alpha t}^H] = \alpha^{-H} E[B_{\alpha t}^H] = 0.$$

The covariance follows from the previous proposition

$$Cov(\alpha^{-H} B_{\alpha s}^H, \alpha^{-H} B_{\alpha t}^H) = \alpha^{-2H} Cov(B_{\alpha s}^H, B_{\alpha t}^H)$$

$$= \alpha^{-2H} \alpha^{2H} Cov(B_s^H, B_t^H) = Cov(B_s^H, B_t^H).$$

■

The self-similarity property is also reffered to as the fractal property of fBm.
Proposition 3.1.4 (Variance of a fBm increment) The variance of a fBm increment is given by

$$Var(B_t^H - B_s^H) = (t-s)^{2H}, \qquad s \leq t.$$

Proof. By simple computations we have that

$$Var(B_t^H - B_s^H) = E[(B_t^H - B_s^H)^2] = E[(B_t^H)^2 - 2B_t^H B_s^H - (B_s^H)^2]$$

$$= E[(B_t^H)^2] - 2E[B_t^H B_s^H] + E[(B_s^H)^2]$$

$$= t^{2H} - (t^{2H} + s^{2H} - (t-s)^{2H}) + s^{2H} = (t-s)^{2H}.$$

■

Corollary 3.1.5 (fBm has stationary increments) The two processes $B_t^H - B_s^H$ and B_{t-s}^H have the same law.
Proof. It is clear that both processes have mean 0 since they are centered Gaussian processes. What remains to show is that $Var(B_{t-s}^H) = Var(B_t^H - B_s^H)$. This is an immediate consequence of the previous proposition

$$Var(B_t^H - B_s^H) = Var(B_t^H) + Var(B_s^H) - 2Cov(B_s^H, B_t^H)$$

$$= t^{2H} + s^{2H} - 2(\frac{1}{2}(t^{2H} + s^{2H} - (t-s)^{2H})) = (t-s)^{2H}.$$

■

Proposition 3.1.6 (fBm has a continuous modification) There exists a continuous modification of the fBm.
Proof. This is a generalised proof of the case for the Brownian motion. Note that the fBm can be written in law as

$$B_t^H - B_s^H = (t-s)^H Z, \qquad Z \sim N(0,1).$$

Now we take the absolute value and exponentiate both sides by β. Lastly, we apply the expectation on both sides which yields

$$E[|B_t^H - B_s^H|^\beta] = (t-s)^{H\beta} E[|Z|^\beta].$$

This is in the form of Kolmogorov's continuity theorem with $\alpha = H\beta - 1$ and $C = E[|Z|^\beta] < \infty$. Thus there is almost surely a modification of a Brownian motion which is Hölder continuous with $\gamma \in (0, H - \frac{1}{\beta})$. Since $\beta > \frac{1}{H}$, we have that $\gamma \in (0, H)$.

■

Remark 3.1.7 (Choosing $H = \frac{1}{2}$) We observe that for the choice $H = \frac{1}{2}$ the fBm coincides with the standard Brownian motion. Moreover, the covariance function in the case of a Brownian motion gives 0 for all $s \leq t$. Thus the jointly normal increments of a Brownian motion are independent in disjoint intervals, which we already knew. However, this is not true for $H \neq \frac{1}{2}$.
Definition 3.1.8 (Long range dependence) We say that a stationary process X_t has long range dependence if the autocovariance function $\rho_H(n) = Cov(X_k, X_{k+n})$ satisfies the following equation for some constants $\alpha, \beta \in (0,1)$

$$\lim_{n\to\infty} \frac{\rho(n)}{\alpha n^{-\beta}} = 1.$$

Remark 3.1.9 (fBm has long range dependence) It is not hard to compute the autocovariance function of the fBm. In fact it is given by

$$\rho_H(n) = Cov(B_s^H - B_{s-1}^H, B_{s+n}^H - B_{s+n-1}^H) = \frac{1}{2}[(n+1)^{2H} + (n-1)^{2H} - 2n^{2H}].$$

We see that for $H < \frac{1}{2}$ the increments of the fBm are negatively correlated as $\rho_H(n) < 0$. One can use this property to model chaotic oscillating behavior like turbulence in aerodynamics. For $H > \frac{1}{2}$ we have positively correlated increments as $\rho_H(n) > 0$. This property can be useful when modeling stock prices with speculating investors.

Using Taylor series of second order we can approximate the asymptotic behaviour of the autocovariance function. This yields

$$\rho_H(n) \sim H(2H-1)n^{2H-2}, \qquad |n| \to \infty.$$

Setting $\alpha = H(2H-1)$ and $\beta = 2 - 2H$ we that the fBm exhibits long range dependency for $H \in (\frac{1}{2}, 1)$, because

$$\lim_{n\to\infty} \frac{\rho_H(n)}{H(2H-1)n^{2H-2}} = 1.$$

Some immediate consequences we can draw by comparing $\rho_H(n)$ to the asymptotic behavior is that the p-series n^{2H-2} converges for $2H-2 < -1 \implies H < \frac{1}{2}$. Hence

$$\sum_{n=1}^{\infty} |\rho_H(n)| = \infty, \qquad H > \frac{1}{2},$$

and

$$\sum_{n=1}^{\infty} |\rho_H(n)| < \infty, \qquad H < \frac{1}{2}.$$

Lemma 3.1.10 (Useful properties of p-variation) Let $1 \leq p < q$ and $V^p_{[a,b]}(\Pi, f)$ be the p-variation. Then

(i) $\quad V^q_{[a,b]}(\Pi, f) \leq V^p_{[a,b]}(\Pi, f)$,

(ii) $\quad V^p_{[a,b]}(\Pi, f) < \infty \implies V^q_{[a,b]}(\Pi, f) = 0$.

Proposition 3.1.11 (fBm is not a semimartingale) The fBm is not a semimartingale for $H \neq \frac{1}{2}$.

Proof. Let $p > 0$ and define the following sequence of random variables

$$X_{n,p} = n^{pH-1} \sum_{i=1}^{n} |B^H_{\frac{i}{n}} - B^H_{\frac{i-1}{n}}|^p, \qquad n \geq 1.$$

By the self-similarity property of fBm we have that $X_{n,p}$ has the same distribution as $Y_{n,p}$, where

$$Y_{n,p} = n^{-1} \sum_{i=1}^{n} |B^H_i - B^H_{i-1}|^p, \qquad n \geq 1.$$

Using that fBm has stationary and eregodic increments, it follows from the eregodic theorem that $Y_{n,p}$ converges in $L^1(\Omega)$ to $E[|B^H_1|^p]$ as n goes to infinity. Since $X_{n,p}$ has the same distribution it must converge to the same limit as $Y_{n,p}$. This yields that the p-variation

$$V_{n,p} = \lim_{n\to\infty} \sum_{i=1}^{n} |B^H_{\frac{i}{n}} - B^H_{\frac{i-1}{n}}|^p = \begin{cases} 0, & pH > 1 \\ \infty, & pH < 1 \end{cases}.$$

In the case where $H < \frac{1}{2}$, we can choose $p > 2$ such that $pH < 1$. Thus by part (i) of the Lemma 3.1.10 we have that

$$V^p_{[0,T]}(\Pi, B^H) = \infty \implies \infty \leq V^2_{[0,T]}(\Pi, B^H) \implies V^2_{[0,T]}(\Pi, B^H) = \infty.$$

If $H > \frac{1}{2}$ we can choose $\frac{1}{H} < p < 2$ such that $pH > 1$. By part (i) of the Lemma 3.1.10 we have that

$$V^p_{[0,T]}(\Pi, B^H) = 0 \implies V^2_{[0,T]}(\Pi, f) \leq 0 \implies V^2_{[0,T]}(\Pi, f) = 0.$$

However, if we choose $1 < p < \frac{1}{H}$ such that $pH < 1$. Then by the contrapositive statement of part (ii) in Lemma 3.1.10 we have that

$$V^p_{[0,T]}(\Pi, B^H) \neq 0 \implies V^1_{[0,T]}(\Pi, B^H) = \infty.$$

Remembering that all semimartingales have bounded quadratic variation, we conclude that fBm can not be a semimartingale for $H \neq \frac{1}{2}$.

■

The largest class of processes for which the Itô integral is well-defined is the class of semimartingales. As the fBm is not a semimartingale, we will need to develop a new meaning for the stochastic integral of the fBm. One approach is to make use of an isometry at the cost of considering deterministic integrands.

Fractional Brownian motion and isometry

Consider the interval $[0,T]$ with a fraction Brownian motion $B^H = \{B^H_t, t \in [0,T]\}$. We will be interested in studying the set of step functions, denoted by $\mathcal{E}$, on $[0,T]$. We will also consider the Hilbert space $\mathcal{H} = \overline{\mathcal{E}}$, that is the closure of set of step functions on $[0,T]$ with respect to the inner product

$$R_H(s,t) = \langle 1_{[0,s]}, 1_{[0,t]} \rangle_{\mathcal{H}}.$$

We see that the mapping $1_{[0,t]} \to B^H_t$ can be extended to an isometry between $\mathcal{H}$ and the Gaussian space $\mathcal{H}_1$ associated to B^H_t. Where $\mathcal{H}_1$ is the closed subspace of $L^2(\Omega, \mathcal{A}, P)$ whose elements are Gaussian random variables with mean zero. We will adopt the notation $\phi \to B^H(\phi)$ for this isometry.

Definition 3.1.12 (Isonormal Gaussian process) We say that a stochastic process $W = \{W(h), h \in \mathcal{H}\}$ defined in a complete probability space $(\Omega, \mathcal{A}, P)$ is an isonormal Gaussian process if W is a centered Gaussian family of random variables such that

$$E[W(h)W(g)] = \langle h, g \rangle_{\mathcal{H}}, \qquad \forall h, g \in \mathcal{H}.$$

It is the case that $\{B^H(\phi), \phi \in \mathcal{H}\}$ is an isonormal Gaussian process.

Moreover the standard Brownian motion on $[0,T]$ can be recovered by considering the Hilbert space of square integrable functions $L^2([0,T])$. We can choose

$$h = 1_{[0,t]}, \qquad g = 1_{[0,s]},$$

such that

$$W(h) = \int_0^T 1_{[0,t]}(s) dW_s,$$

and

$$E[W(h)W(g)] = t \wedge s.$$

We will now consider the special case where $H > \frac{1}{2}$. Then we can represent the covariance function of fBm as

$$R_H(s,t) = H(2H-1)\int_0^t\int_0^s |r-u|^{2H-2}dudr.$$

This implies that for any pair of step functions $\phi, \psi \in \mathcal{E}$ their inner product on $\mathcal{H}$ is given by

$$\langle \phi, \psi \rangle_{\mathcal{H}} = H(2H-1)\int_0^T\int_0^T |r-u|^{2H-2}\phi(r)\psi(u)dudr.$$

Furthermore, one can show that

$$|r-u|^{2H-2} = \frac{(ru)^{H-\frac{1}{2}}}{\beta(2-2H, H-\frac{1}{2})}\int_0^{r\wedge u} v^{1-2H}(r-v)^{H-\frac{3}{2}}(u-v)^{H-\frac{3}{2}}dv,$$

where β denotes the beta function defined as

$$\beta(x,y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}, \qquad x, y > 0,$$

with Γ being the gamma function.

We will be interested in the properties of the following square integrable kernel

$$K_H(s,t) = c_H s^{\frac{1}{2}-H}\int_s^t (u-s)^{H-\frac{3}{2}}u^{H-\frac{1}{2}}du, \qquad s < t.$$

Where c_H is the following H dependent constant

$$c_H = \left[\frac{H(2H-1)}{\beta(2-2H, H-\frac{1}{2})}\right]^{\frac{1}{2}}.$$

By combining the integral representation of $|r-u|^{2H-2}$ and $R_H(s,t)$ one can verify that

$$R_H(s,t) = \int_0^{s\wedge t} K_H(s,u)K_H(t,u)du.$$

An important fact is that the kernel K_H is differentiable with the following expression

$$\frac{\partial K_H}{\partial t}(s,t) = c_H\left(\frac{t}{s}\right)^{H-\frac{1}{2}}(t-s)^{H-\frac{3}{2}}.$$

This is important as it can be used to define the linear operator $K_H^* : \mathcal{E} \to L^2([0,T])$ defined by

$$K_H^*(\phi)(s) = \int_s^T \phi(t) \frac{\partial K_H}{\partial t}(t,s)dt, \qquad \phi \in \mathcal{E}.$$

Notice that K_H^* applied to the indicator function yields

$$K_H^*(1_{[0,t]})(s) = K_H(t,s)1_{[0,t]}(s).$$

We see that the operator K_H^* is an isometry that can be extended to the Hilbert space $\mathcal{H}$. Indeed

$$\langle K_H^* 1_{[0,s]}, K_H^* 1_{[0,t]} \rangle_{L^2([0,T])} = \langle K_H(s,\cdot)1_{[0,s]}, K_H(t,\cdot)1_{[0,t]} \rangle_{L^2([0,T])}$$

$$= \int_0^{s\wedge t} K_H(s,u)K_H(t,u)du = R_H(s,t) = \langle 1_{[0,s]}, 1_{[0,t]} \rangle_{\mathcal{H}}.$$

Inspect the process $B = \{B_t, t \in [0,T]\}$ defined by

$$B_t = B^H((K_H^*)^{-1}(1_{[0,t]})).$$

Then B is a Brownian motion. Indeed for $s,t \in [0,T]$ we have that

$$E[B_s B_t] = E[B^H((K_H^*)^{-1}(1_{[0,s]}))B^H((K_H^*)^{-1}(1_{[0,t]}))]$$

$$= \langle (K_H^*)^{-1}(1_{[0,s]}), (K_H^*)^{-1}(1_{[0,t]}) \rangle_{\mathcal{H}} = \langle 1_{[0,s]}, 1_{[0,t]} \rangle_{L^2([0,T])} = s \wedge t.$$

We now have the tools to give the following integral representation of fBm

$$B_t^H = \int_0^t K_H(s,t)dW_s.$$

Generalizing for any $\phi \in \mathcal{H}$ we get that

$$B^H(\phi)(t) = \int_0^t K_H^*(\phi)(s)dW_s.$$

3.2 Fractional Ornstein-Uhlenbeck process

This section will deal with the fractional Ornstein-Uhlenbeck process which can be obtained by solving a SDE with fBm as noise. The material gathered is based on [CKM03] which is an article covering advanced results on the fractional Ornstein-Uhlenbeck process. For our applications, we will need the result regarding existence and uniqueness of the Riemann-Stieltjes integral with respect to fBm.

Proposition 3.2.1 (Existence, uniqueness and continuity) Let $B^H = \{B_t^H, t \in \mathbb{R}\}$ be a fBm with Hurst parameter $H \in (0,1)$. Let $\xi \in \mathbb{R}$, $a \in [-\infty, \infty)$ and $\lambda, \sigma > 0$. Then for almost all $\omega \in \Omega$, the following statements hold.

(i) For all $a < t$

$$\int_a^t e^{\lambda u} dB_u^H(\omega),$$

exists as a Riemann-Stieltjes integral. Furthermore, the integral is explicitly given by

$$e^{\lambda t} B_t^H(\omega) - e^{\lambda a} B_a^H(\omega) - \lambda \int_a^t e^{\lambda u} B_u^H(\omega) du.$$

(ii) The function

$$\int_a^t e^{\lambda u} dB_u^H(\omega), \qquad a < t,$$

is almost surely continuous in t.

(iii) The distinct continuous function f that solves

$$f(t) = \xi - \lambda \int_0^t f(s) ds + \sigma B_t^H(\omega), \qquad t \geq 0,$$

has the solution

$$f(t) = e^{-\lambda t}\Big(\xi + \sigma \int_0^t e^{\lambda u} dB_u^H(\omega)\Big), \qquad t \geq 0.$$

(iv) The distinct continuous function that solves

$$f(t) = \sigma \int_{-\infty}^0 e^{\lambda u} dB_u^H(\omega) - \lambda \int_0^t f(s) ds + \sigma B_t^H(\omega), \qquad t \geq 0,$$

is given by

$$f(t) = \sigma \int_{-\infty}^t e^{-\lambda(t-u)} dB_u^H(\omega), \qquad t \geq 0.$$

Remark 3.2.2 (Langevin equation) Let $\xi \in \mathbb{R}$ and $\lambda, \sigma > 0$. The following SDE is called the Langevin equation

$$X_t = \xi - \lambda \int_0^t X_s ds + N_t, \qquad t \geq 0.$$

What is worth noticing is that the integral is with respect to the Lebesgue measure and not a stochastic process. This allows us to find path-wise solutions for the Langevin equation for more complicated noise processes $N = \{N_t, t \geq 0\}$ than the Brownian motion.

Example 3.2.3 (Fractional Ornstein-Uhlenbeck process) Let $\lambda, \sigma > 0$, $\xi \in \mathbb{R}$, $H \in (0,1)$ and $a \in [-\infty, \infty)$. By Proposition 3.2.1 we know that

$$\int_a^t e^{\lambda u} dB_u^H, \qquad a < t,$$

makes sense as a path-wise Riemann-Stieltjes integral, which we recall is almost surely continuous in t. Moreover, the distinct almost surely continuous process that solves the Langevin equation with fBm as noise, that is

$$X_t = \xi - \lambda \int_0^t X_s ds + \sigma B_t^H, \qquad t \geq 0,$$

is given by

$$Y_t^{H,\xi} = e^{-\lambda t}\Big(\xi + \sigma \int_0^t e^{\lambda u} dB_u^H\Big), \qquad t \geq 0.$$

Another solution to the Langevin equation with fBm as noise is

$$Y_t^H = \sigma \int_{-\infty}^t e^{-\lambda(t-u)} dB_u^H, \qquad t \geq 0,$$

with initial condition $\xi = Y_0^H$. It follows directly from the Gaussianity and stationary increments of the fBm that Y_t^H is a stationary Gaussian process. Since

$$Y_t^H - Y_t^{H,\xi} = e^{-\lambda t}(Y_0^H - \xi) \to 0, \text{ as } t \to \infty, \text{ almost surely,}$$

it is implied that all stationary solutions to the Langevin equation with fBm as noise must have the same distribution as Y_t^H. Finally, we say that $Y_t^{H,\xi}$ is a fractional Ornstein-Uhlenbeck process with initial condition ξ and Y_t^H a stationary fractional Ornstein-Uhlenbeck process.

Example 3.2.4 (Simulating paths) Let $\lambda, \sigma > 0$. In order to get some intuition of how the paths of the fractional Ornstein-Uhlenbeck process behave for different values of $H \in (0,1)$, we will simulate some paths using Euler-Maruyama's method. Without loss of generality, we will work with the following SDE

$$\begin{aligned} dX_t &= -\lambda X_t dt + \sigma dB_t^H, \qquad t \in [0,1], \\ X_0 &= 0. \end{aligned}$$

Using Euler-Maruyama's method the numerical scheme becomes

$$X_{t_{i+1}} - X_{t_i} \approx -\lambda X_{t_i}(t_{i+1} - t_i) + \sigma(B_{t_{i+1}}^H - B_{t_i}^H),$$

for any partition $t_0 = 0 < ... < t_n = 1$ of $[0,1]$. In particular, for a uniform partition of $[0,1]$ we have that $t_i = \frac{i}{n}$. Hence $t_{i+1} - t_i = \frac{1}{n}, \quad i \in \{0, ..., n\}$. Thus the method

$$X_{t_{i+1}} \approx X_{t_i} - \frac{\lambda}{n} X_{t_i} + \sigma(B_{t_{i+1}}^H - B_{t_i}^H).$$

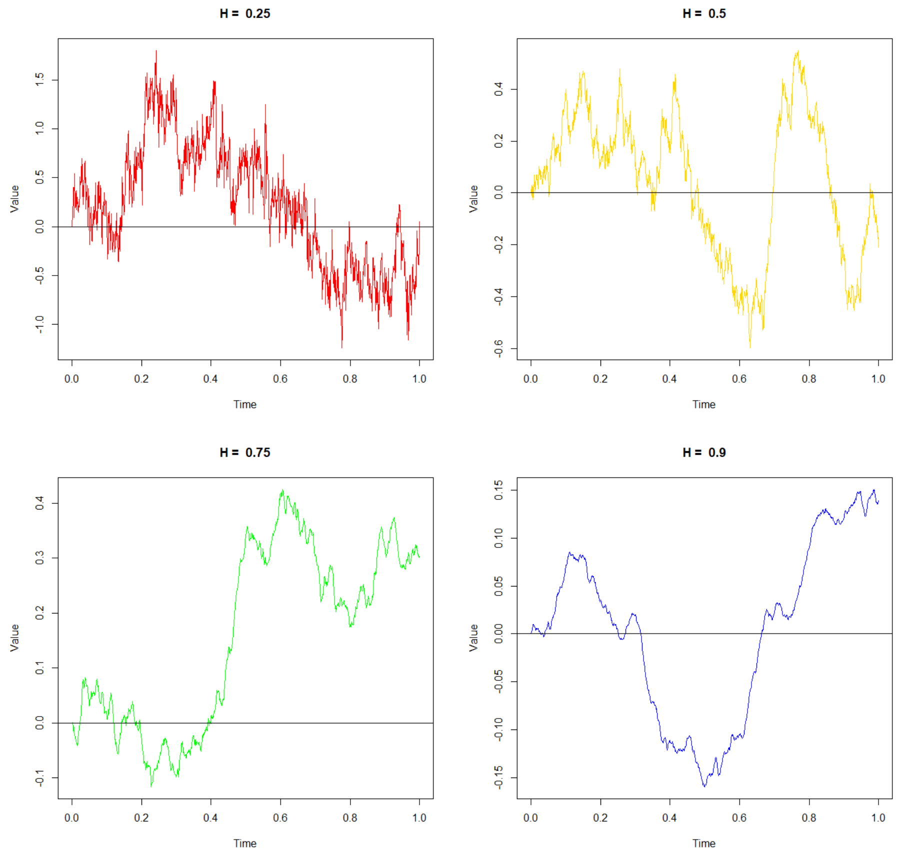

Figure 3.1: Simulated paths of the fractional Ornstein-Uhlenbeck process for mean reversion rate $\lambda = 2$, volatility $\sigma = 1$ and $n = 1000$ time steps using Euler-Maruyama's method. We observe that the paths get smoother as we increase H.

Remark 3.2.5 (Integration is well-defined) We refer to [PT00] where it is shown that for $H \in (\frac{1}{2}, 1)$ and two real-valued and measurable functions f, g satisfying

$$f, g \in \Big\{ f : \int_{\mathbb{R}} \int_{\mathbb{R}} |f(u)||f(v)||u - v|^{2H-2} du dv < \infty \Big\},$$

then the two integrals

$$\int_{\mathbb{R}} f(u) dB_u^H, \int_{\mathbb{R}} g(u) dB_u^H$$

are well-defined as limits of integrals of elementary functions. Finally, it is shown that

$$E\Big[\int_{\mathbb{R}} f(u) dB_u^H \int_{\mathbb{R}} g(u) dB_u^H \Big] = H(2H-1) \int_{\mathbb{R}} \int_{\mathbb{R}} f(u) g(v) |u - v|^{2H-2} du dv.$$

CHAPTER 4

Modeling mortality

In this chapter we propose an exponential fractional Ornstein-Uhlenbeck process to model Norwegian mortality rates.

4.1 Fractional mortality

This section is inspired by the work of [DO19] where the authors have used the exponential fractional Ornstein-Uhlenbeck process to model Italian mortality rates. The main reason for choosing such a model is that morality rates across generations are highly correlated albeit not perfectly correlated. Some explanations of this phenomenon include medical advancement and technological innovation which tend to increase life expectancy of future generations.

Remark 4.1.1 (Model assumptions and previous results) We will denote by $T > 0$ a fixed time horizon and by $\mu : [0, T] \times \Omega \to \mathbb{R}$ a stochastic process which models the mortality on the time span $[0, T]$. Given the mortality rate and an $\omega \in \Omega$, then by Theorem 2.6.14 the survival probability of an individual of age x in the time period $[s, t] \subseteq [0, T]$ is given by

$$p^{\omega}_{**}(x + s, x + t) = e^{-\int_{x+s}^{x+t} \mu(u,\omega)du} = e^{-\int_s^t \mu(x+u,\omega)du}.$$

Unfortunately, μ is not necessarily adapted to the available information denoted by $\mathcal{F}_t$. To circumvent this measurability issue, we will consider the adapted projection on to $\mathcal{F}_t$. This gives rise to the following conditional expectation

$$p_{**}(x + s, x + t) = E\Big[e^{-\int_s^t \mu(x+u)du}\Big|\mathcal{F}_t\Big].$$

In this thesis, we will model the mortality rate μ by a generalisation of the Milevsky-Promislow model. This means that μ will be of the following form

$$\mu(t) = \mu_0 e^{\alpha_0 t+\alpha_1 Y_t^H}, \qquad t \in [0, T],$$

where $\mu_0, \alpha_0, \alpha_1 \in \mathbb{R}$ are parameters. $Y_t^H = \{Y_t^H\}_{t\in[0,T]}$ is a fractional Ornstein-Uhlenbeck process with dynamics given by

$$dY_t^H = -\lambda Y_t^H + \sigma dB_t^H,$$

$$Y_0^H = 0,$$

where $B_t^H = \{B_t^H\}_{t\in[0,T]}$ is a fBm with with Hurst parameter $H \in (\frac{1}{2}, 1)$ and $\lambda, \sigma > 0$.
Interested readers may seek out [MP01] for the authentic paper written by Milevsky and Promislow where they instead consider the Ornstein-Uhlenbeck process driving the noise of the proposed mortality rate.

Recall that the meaning of the dynamics are given by the Langevin equation in integral form, that is

$$Y_t^H = -\lambda \int_0^t Y_s^H ds + \sigma B_t^H.$$

Furthermore, the authors of [CKM03] have shown that the solution of the Langevin equation with fBm as noise is given by

$$Y_t^H = \sigma \int_0^t e^{-\lambda(t-u)} dB_u^H.$$

We have seen in the previous chapter that the stochastic integral is well-defined as a pathwise Riemann-Stieltjes integral. By Proposition 3.2.1 we also know that Y_t^H is unique and has almost surely continuous paths. Although Y_t^H does not have the nice properties of being Markovian or a semimatingale for $H \in (\frac{1}{2}, 1)$, we do at least have that Y_t^H is Gaussian and ergodic. It can be shown using Fourier transforms, see [ZCY12] for details, that the variance of Y_t^H has an explicit representation of the form

$$Var(Y_t^H) = 2H\sigma^2 e^{-2\lambda t} \int_0^t s^{2H-1} e^{2\lambda s} ds.$$

In the special case where $H = \frac{1}{2}$ we get that

$$Var(Y_t^H) = \frac{\sigma^2}{2\lambda}(1 - e^{-2\lambda t}),$$

which coincides with the expression we calculated in Example 2.5.10 using Itô isometry, as expected.

Proposition 4.1.2 (Bounding the variance) Let $\lambda, \sigma > 0$ and $\alpha_1 = T^{-H}$. Furthermore, let $Y_t^H = \{Y_t^H\}_{t\in[0,T]}$ be a fractional Ornstein-Uhlenbeck process with $H \in (0, 1)$. Then

$$Var(\alpha_1 Y_t^H) \leq \sigma^2.$$

Proof. A simple calculation gives that

$$\begin{aligned} Var(\alpha_1 Y_t^H) &= \alpha_1^2 Var(Y_t^H) = \alpha_1^2 2H\sigma^2 \int_0^t s^{2H-1} e^{-2\lambda(t-s)} ds \\ &\leq \alpha_1^2 2H\sigma^2 \int_0^t s^{2H-1} ds = \alpha_1^2 2H\sigma^2 \frac{s^{2H}}{2H}\Big|_{s=0}^{s=t} = \alpha_1^2 \sigma^2 t^{2H} \\ &= T^{-2H}\sigma^2 t^{2H} = \sigma^2 \Big(\frac{t}{T}\Big)^{2H} \leq \sigma^2, \qquad 0 \leq s \leq t \leq T. \end{aligned}$$

Hence $Var(\alpha_1 Y_t^H)$ is bounded by a time independent constant. We can use α_1 as a tool to control the variance of Y_t^H.

■

4.2 Estimation of parameters

We will now focus on developing a method for estimating $\alpha_0, \alpha_1, \lambda, \sigma$ and H which all appear in the model for fractional mortality.

Assumption 4.2.1 (H is preserved) Since the estimation of H is quite complicated and sensitive, we will make use of a simplifying assumption. The authors of [Yer+14] have experimented with the assumption that the Hurst parameter of the fBm B_t^H driving the Gaussian noise in the fractional Ornstein-Uhlenbeck SDE and the fractional Ornstein-Uhlenbeck process Y_t^H are equal. Thus we will from now on assume that H can be chosen equal for both processes. Although this is a critique worthy assumption, it is also a necessary assumption for estimation purposes.

Remark 4.2.2 (Estimating α_0) We want to find an estimator for α_0 by minimizing the square sum errors, abbreviated to SSE. Without loss of generality, we will momentarily assume that $\alpha_1 = 1$. Taking the natural logarithm of the mortality rate we obtain

$$ln\ \mu(t) = ln\ \mu_0 + \alpha_0 t + Y_t^H.$$

We will distinguish between the observed mortality rate μ_i and the expected log-mortality rate $\widehat{z}_i = E[ln\ \mu_i]$, both at time t_i, where the latter is given by

$$\widehat{z}_i = ln\ \mu_0 - \alpha_0 t_i, \qquad i = \{1, ..., n\}.$$

Then the SSE is given by

$$SSE = \sum_{i=1}^{n} \Big(ln\ \mu_i - \widehat{z}_i\Big)^2 = \sum_{i=1}^{n} \Big(ln\ \mu_i - ln\ \mu_0 - \alpha_0 t_i\Big)^2.$$

Differentiating SSE with respect to α_0 and setting the result equal to zero we obtain

$$\frac{\partial SSE}{\partial \alpha_0} = -2 \sum_{i=1}^{n} \Big(ln\ \mu_i - ln\ \mu_0 - \alpha_0 t_i\Big) t_i = 0.$$

Finally, by solving the equation with respect to α_0 we get the estimator

$$\widehat{\alpha}_0 = \frac{\sum_{i=1}^{n} \Big(ln\ \mu_i - ln\ \mu_0\Big) t_i}{\sum_{i=1}^{n} t_i^2}.$$

It is trivial that $\widehat{\alpha}_0$ minimizes SSE by looking at the second partial derivative with respect to α_0. Indeed, we have that

$$\frac{\partial^2 SSE}{\partial \alpha_0^2} = -2 \sum_{i=1}^{n} -t_i^2 = 2 \sum_{i=1}^{n} t_i^2 > 0,$$

which shows that SSE is convex in α_0.

We proceed by introducing some terminology which will come in handy when estimating the remaining parameters.

Definition 4.2.3 ($C^{1+\alpha}$) We denote by $C^{1+\alpha}(\mathbb{R},\mathbb{R})$ the set of all functions $g:\mathbb{R}\to\mathbb{R}$ which are continuously differentiable and satisfy

$$\sup_x |g'(x)| + \sup_{x\neq y} \frac{|g'(x)-g'(y)|}{|x-y|^\alpha} < \infty, \qquad \alpha \in (0,1), \qquad x,y\in\mathbb{R}.$$

Definition 4.2.4 (First order quadratic variation) Let $X=\{X_t\}_{t\in[0,T]}$ be a real valued process. We define the first order quadratic variation of X by

$$V_n^1(X) = \sum_{i=0}^{n-1}\Big(X_{t_{i+1}} - X_{t_i}\Big)^2, \qquad n\in\mathbb{N},$$

$$t_i = i\frac{T}{n}, \qquad i=\{0,...,n\}.$$

Theorem 4.2.5 If $f:[0,T]\to\mathbb{R}$ is a Lipschitz continuous function and $g\in C^{1+\alpha}([0,T],\mathbb{R})$, then

$$\lim_{n\to\infty} n^{2H-1}V_n^1(X) = \int_0^T g^2(X_t)dt,$$

where X is the solution of

$$X_t = X_0 + \int_0^t f(X_s)ds + \int_0^t g(X_s)dB_s^H, \qquad X_0\in\mathbb{R}, \qquad H\in\Big(\frac{1}{2},1\Big).$$

Proof. See [Mel11] for proof.

■

Remark 4.2.6 (Estimating H) Let $H\in\Big(\frac{1}{2},1\Big)$, then a reasonable estimator of H suggested in [Mel11] is given by

$$\widehat{H} = \frac{1}{2} - \frac{1}{2ln(2)}ln\Big(\frac{V_{2n}^1(X)}{V_n^1(X)}\Big), \qquad n\to\infty,$$

where V_{2n}^1 is the first order quadratic variation of the entire sample path and V_n^1 is the first order quadratic variation of the subset

$$\Big\{X_i : i=2j, 0\le j\le \Big[\frac{n}{2}\Big]\Big\},$$

$[x]$ denotes the integer part of a real number x.

The Hurst estimator relies heavily on the self-similarity property of the fBm and the asymptotic behaviour of the first order quadratic variation. Indeed we have that

$$V_n^1 = \sum_{i=0}^{n-1}\Big(X_{t_{i+1}} - X_{t_i}\Big)^2 \sim \sum_{i=0}^{n-1} X_{\frac{T}{n}}^2 \sim nX_{\frac{T}{n}}^2 \sim n\Big(\frac{T}{n}\Big)^{2H}X_1^2.$$

Similarly

$$V_{2n}^1 = \sum_{i=0}^{2n-1} \Big(X_{t_{i+1}} - X_{t_i}\Big)^2 \sim \sum_{i=0}^{2n-1} X_{\frac{T}{2n}}^2 \sim 2nX_{\frac{T}{2n}}^2 \sim 2n\Big(\frac{T}{2n}\Big)^{2H} X_1^2.$$

Taking the ratio we get that

$$\frac{V_{2n}^1}{V_n^1} \sim 2\Big(\frac{T}{2n} : \frac{T}{n}\Big)^{2H} = 2\Big(\frac{1}{2}\Big)^{2H}, \qquad n \to \infty.$$

Applying the logarithm on both sides yield that

$$ln\Big(\frac{V_{2n}^1}{V_n^1}\Big) = (1-2H)ln(2), \qquad n \to \infty.$$

Finally, we solve the equation with respect to H. This gives us the estimator

$$\widehat{H} = \frac{1}{2} - \frac{1}{2ln(2)} ln\Big(\frac{V_{2n}^1(X)}{V_n^1(X)}\Big), \qquad n \to \infty.$$

Remark 4.2.7 (Estimating σ) We have as an immediate consequence of Theorem 4.2.5 that we can estimate a constant volatility σ by rearranging the integral equation

$$n^{2H-1} V_n^1(X) = \int_0^T \sigma^2 dt = \sigma^2 T.$$

Solving for σ we get the estimator

$$\widehat{\sigma} \approx \sqrt{\frac{n^{2H-1}V_n^1(X)}{T}}.$$

Remark 4.2.8 (Estimating λ) Hu and Nualart have shown in [HN10] that the fractional Ornstein-Uhlenbeck process has the following long term variance

$$\lim_{T\to\infty} Var(Y_T) = \lim_{T\to\infty} E[Y_T^2] = \lim_{T\to\infty} \frac{1}{T}\int_0^T Y_s^2 ds \overset{a.s.}{=} \sigma^2 \lambda^{-2H} H\Gamma(2H).$$

Rearranging the equation above with respect to λ, we arrive at the estimator

$$\widehat{\lambda} = \left(\frac{1}{\sigma^2 H\Gamma(2H)T} \int_0^T Y_s^2 ds\right)^{-\frac{1}{2H}}.$$

We will mostly work with discrete observations, therefore we will approximate the second moment by the empirical second moment given by

$$\widehat{\mu}_2 = \frac{1}{N}\sum_{i=1}^N X_{t_i}^2.$$

This gives rise to the discrete estimator

$$\widehat{\lambda} = \left(\frac{\widehat{\mu}_2}{\sigma^2 H\Gamma(2H)}\right)^{-\frac{1}{2H}}.$$

Theorem 4.2.9 (Strongly consistent estimators) Assume that the conditions of Theorem 4.2.5 are satisfied and that $H \in (\frac{1}{2}, \frac{3}{4})$. Then

$$\lim_{n\to\infty} (\widehat{H}, \widehat{\sigma}, \widehat{\lambda}) \overset{a.s.}{\to} (H, \sigma, \lambda).$$

Example 4.2.10 (Numerical simulations) We have simulated 1 000 paths of the fractional Ornstein-Uhlenbeck process each with 1 000 discrete observations using Euler-Maruyama's method, using a time horizon of 10 years, to see how well the estimators perform in a controlled environment. We summarize our findings in a table.

Estimate	$\widehat{H}$	$\widehat{\sigma}$	$\widehat{\lambda}$
Mean	0.599523	0.01258286	0.3280835
Median	0.6013932	0.01258386	0.2280266
Standard deviation	0.02752146	0.0003005425	0.3240334

Table 4.1: Summary of estimates. The true values were $H = 0.6$, $\sigma = 0.01$ and $\lambda = 0.4$.

As we see the estimators produce fairly accurate estimates in the sense that they are close to the true value and have a small standard deviation. The most robust estimators seems to be $\widehat{H}$ and $\widehat{\sigma}$ which become better the smaller the mesh size becomes. In our simulations we used the mesh size $|\Pi| = 0.01$. Finally, we see that $\widehat{\lambda}$ is difficult to estimate since it by construction relies on the asymptotic behaviour of the fractional Ornstein-Uhlenbeck process. Indeed, improving the performance of $\widehat{\lambda}$ would require a long time horizon and many observations in order to maintain a small mesh size $|\Pi|$. Unfortunately, this comes with a significant computational cost.

In the rest of this example we will try to determine the distribution of the estimators using the R-package "fitdistrplus". In order to get a decent initial guess of the distribution of $\widehat{H}$ we will make a Cullen and Frey plot.

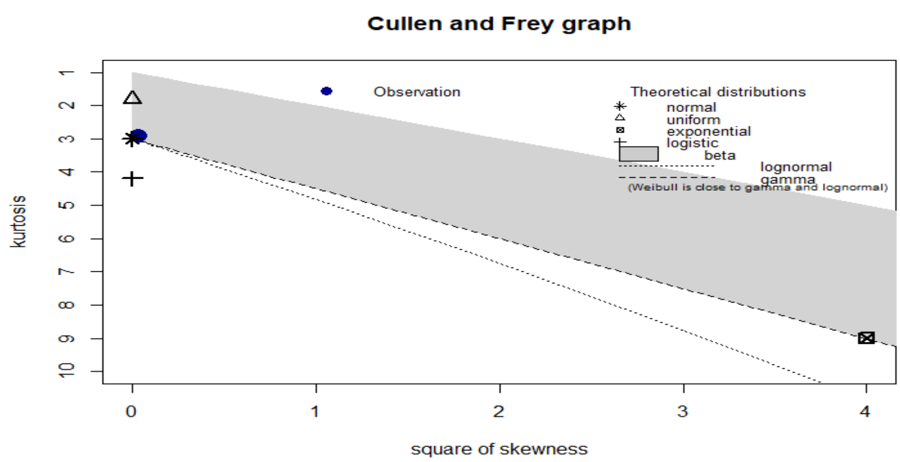

Figure 4.1: Cullen and Frey plot for $\widehat{H}$.

The plot seems to indicate that $\widehat{H}$ follows a normal distribution which we will analyze in more detail.

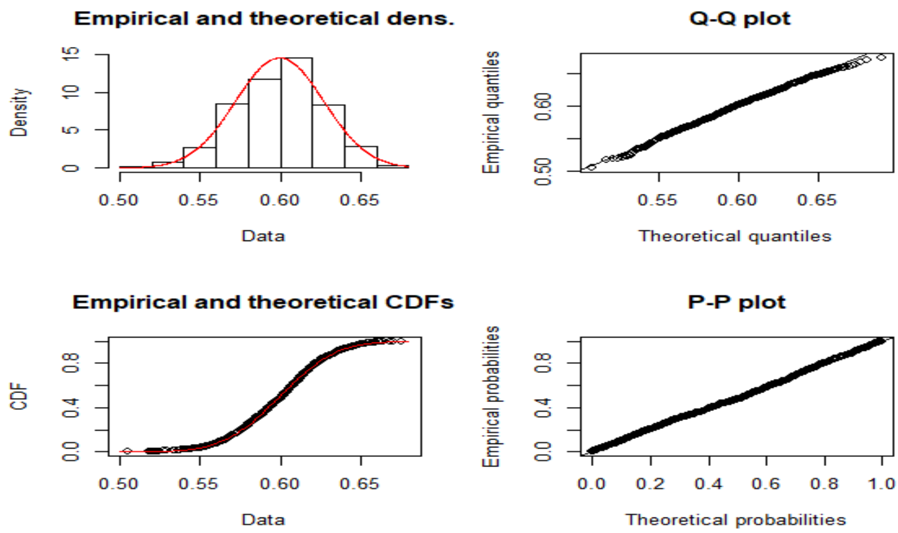

Figure 4.2: Normality check for $\widehat{H}$.

Once again the distribution of $\widehat{H}$ does not differ significantly from a normal distribution which makes normality a safe assumption.

Following the same procedure for $\widehat{\sigma}$, we make a Cullen and Frey plot.

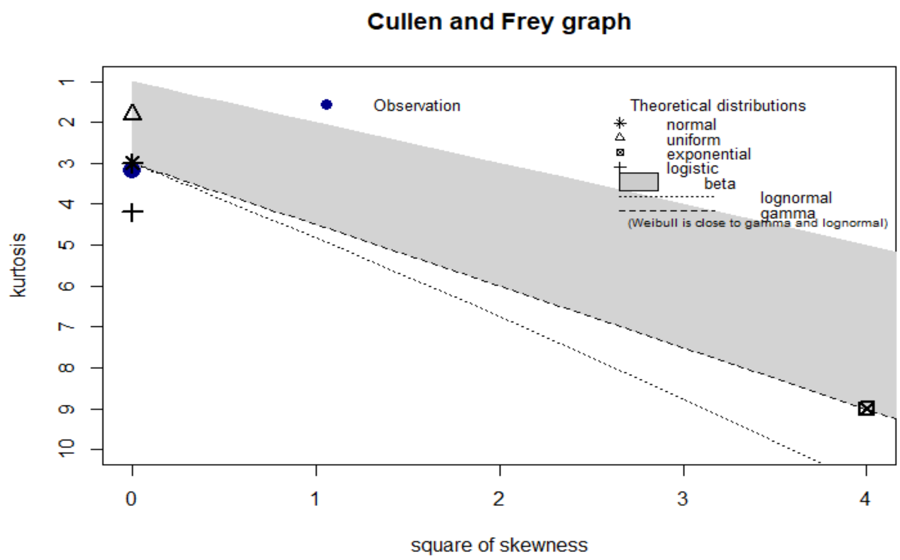

Figure 4.3: Cullen and Frey plot for $\widehat{\sigma}$.

It seem that $\widehat{\sigma}$ is also closely related to the normal distribution, but by the shape of the estimator it is possibly closer to a log-normal distribution. We investigate this by doing a normality check of $ln(\widehat{\sigma})$.

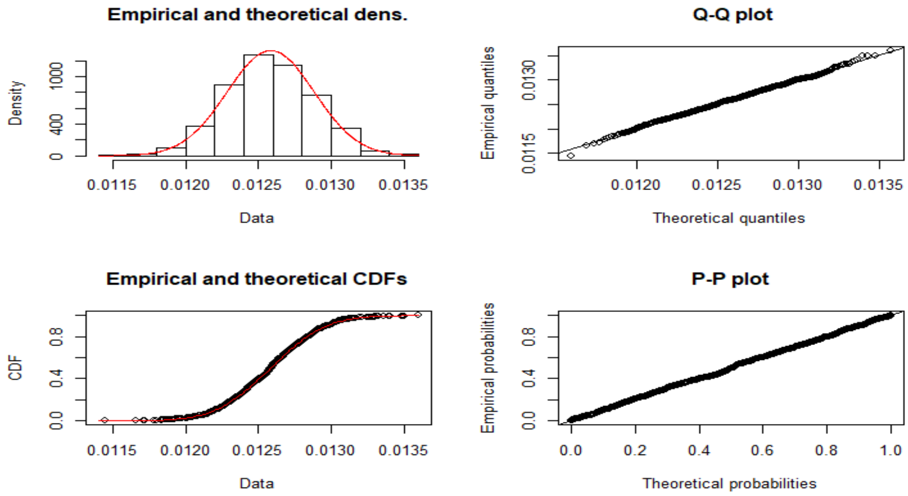

Figure 4.4: Normality check for $ln(\widehat{\sigma})$.

The plots insinuates that our assumption regarding log-normality of $\widehat{\sigma}$ is realistic. Finally, we determine the distribution of $\widehat{\lambda}$. We initiate by creating the Cullen and Frey plot of $\widehat{\lambda}$.

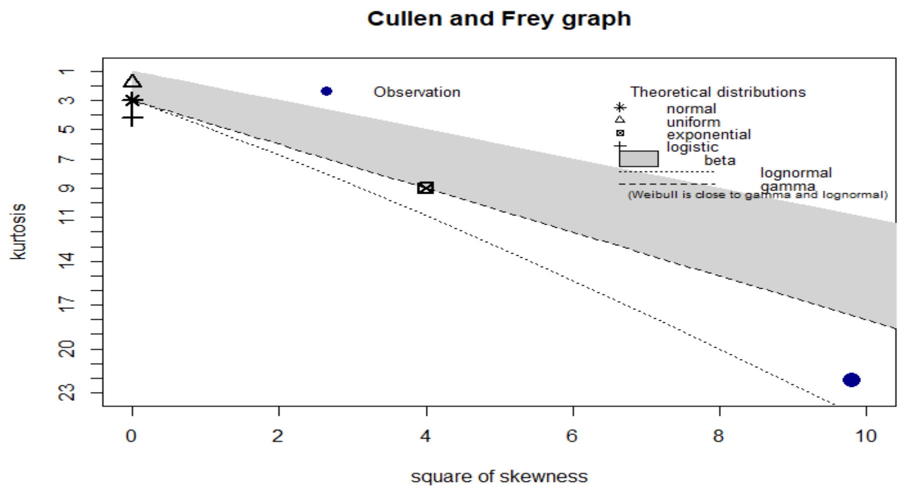

Figure 4.5: Cullen and Frey plot for $\widehat{\lambda}$.

After inspecting the graph we see that the distribution has an extremely high skewness. Fortunately, the blue data point is still close to the dashed line for a log-normal distribution. We repeat the procedure by taking the logarithm and check for normality.

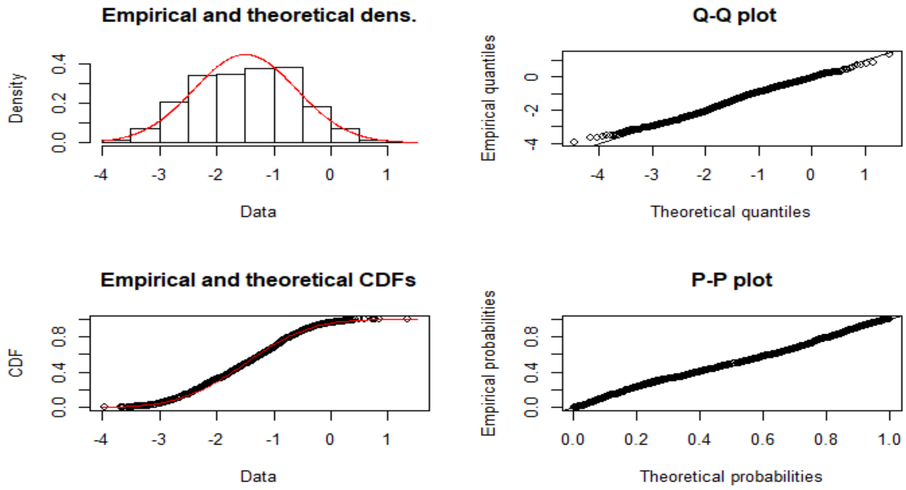

Figure 4.6: Normality check for $\ln(\widehat{\lambda})$.

The normality test checks out even in this extreme case. Thus we will work under the assumption that $\widehat{\lambda}$ follows a strongly skewed log-normal distribution.

CHAPTER 5

Modelling fractional mortality in Norway

In this chapter we will apply our model of fractional mortality, developed in Section 4.1, on Norwegian mortality rates. The data we have collected is gathered from [SSBb] and includes mortality rates from 1996 to 2020. The data will be used to calibrate the estimators we presented and tested using simulations in Section 4.2. Furthermore, we will use the estimates we obtain to run a new simulation study in which we will compare the historical mortality rates with the estimated mortality rates given by our model.

5.1 Fractional Ornstein-Uhlenbeck log-mortality

We present the estimation results of the parameters that are independent of the Hurst parameter H.

Remark 5.1.1 (Overview of the model) As a quick reminder we assumed that mortality rate is given by a geometric type fractional Ornstein-Uhlenbeck process of the form

$$\mu(t) = \mu_0 e^{\alpha_0 t + \alpha_1 Y_t^H}, \qquad t \in [0, T],$$

where Y_t^H is the fractional Ornstein-Uhlenbeck process which can be represented as

$$Y_t^H = \sigma \int_0^t e^{-\lambda(t-u)} dB_u^H.$$

We also recall the following estimators which were derived by maximum likelihood estimation and first order quadratic variation methods.

$$\widehat{H} = \frac{1}{2} - \frac{1}{2ln(2)} ln\Big(\frac{V_{2n}^1(X)}{V_n^1(X)}\Big),$$

$$\widehat{\sigma} = \sqrt{\frac{n^{2H-1}V_n^1(X)}{T}},$$

$$\widehat{\lambda} = \left(\frac{\widehat{\mu}_2}{\sigma^2 H\Gamma(2H)}\right)^{-\frac{1}{2H}},$$

$$\widehat{\alpha}_0 = \frac{\sum_{i=1}^{n}\Big(ln\ \mu_i - ln\ \mu_0\Big)t_i}{\sum_{i=1}^{n} t_i^2},$$

Calculating α_0 and α_1 can be done directly from the observed mortality rates. On the other hand, it is first necessary to transform our data into observations of the fractional Ornstein-Uhlenbeck process in order to estimate H, σ and λ. The reason being that the mortality rates follows a geometric type fractional Ornstein-Uhlenbeck process by assumption. Transforming the data is easy since the mortality function is injective, which we can use to invert the mortality with respect to Y_t^H. Discretizing the equation we get that

$$Y_i^H = \frac{1}{\alpha_1}\Bigg(ln\Big(\frac{\mu_i}{\mu_0}\Big) - \alpha_0 t_i\Bigg).$$

Remark 5.1.2 (Initial results) We now present our findings of how the endogenous parameters depend on age and gender. As the parameter α_0 is easy to estimate we present it first.

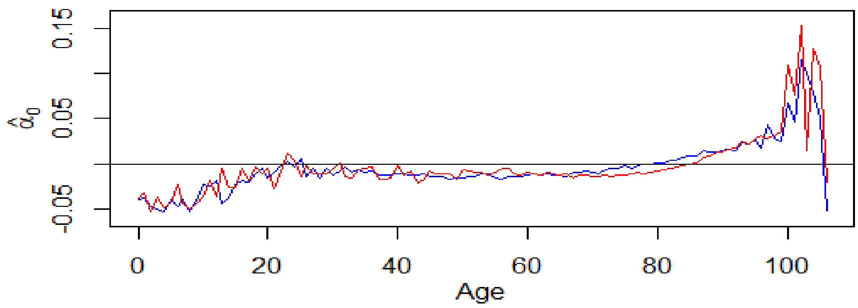

Figure 5.1: Plot showing the evolution of $\widehat{\alpha}_0$ for men (blue) and women (red). It is reasonable that $\widehat{\alpha}_0$ is negative for the most part at early ages as this indicates a deterministic decrease in mortality rate. After the age of 80 we see that $\widehat{\alpha}_0$ is positive, reflecting a deterministic increase in mortality rate.

As we have previously discussed α_1 is an exogenous variable as it can be chosen arbitrarily. Note that we still have a degree of freedom in the parameters appearing in the fractional Ornstein-Uhlenbeck process. For simplicity, and in order to avoid changing the interpretation of the fractional Ornstein-Uhlenbeck parameters, we will set $\alpha_1 = 1$.

The estimation of the Hurst parameter proved to be troublesome for our data, we will explain why later. As a consequence we are forced to consider H as an exogenous variable. A naive approach would be to consider the three cases $H = \{\frac{1}{4}, \frac{1}{2}, \frac{3}{4}\}$.

The first case $H = \frac{1}{4}$ does not make much sense as it would entail a mortality rate with anti persistent behaviour and negative correlation between generations at the time of same age. Empirically, it is observed in most countries that the mortality rates tends to decrease for newer generations as advancements in medicine and technology improve living conditions. Hence it seems reasonable to restrict ourselves to the last two cases.

As for the estimated volatility $\widehat{\sigma}$ and the estimated drift $\widehat{\lambda}$ we have a dependency on the Hurst parameter H. We will therefore consider them separately for each $H = \{\frac{1}{2}, \frac{3}{4}\}$.

5.2 Results of fractional mortality in Norway for $H = 0.5$

The remaining parameters are estimated and presented with the main result of this model given that $H = 0.5$.

Remark 5.2.1 (Results for $H = \frac{1}{2}$) We start this remark by presenting the estimated volatility and estimated drift.

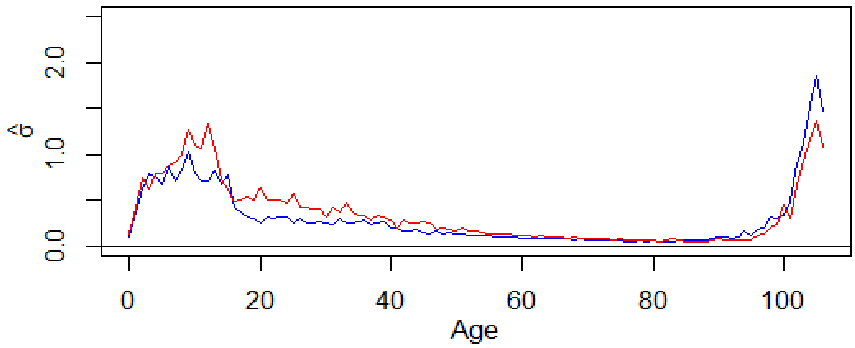

Figure 5.2: Plot showing the evolution of $\widehat{\sigma}$ for men (blue) and women (red). It is reasonable that $\widehat{\sigma}$ is largest for the earliest and oldest ages as the spread in mortality rate becomes huge. In addition, we see that woman tend to have a higher volatility compared to men until age 80.

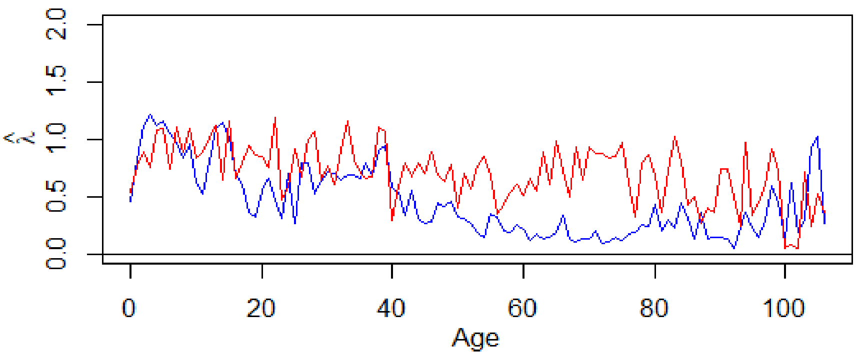

Figure 5.3: Plot showing the evolution of $\widehat{\lambda}$ for men (blue) and women (red). We observe that $\widehat{\lambda}$ tends to follow the shape of $\widehat{\sigma}$.

Subsequently we present the main result of this model for men and women given the selected representative ages $\{0, 15, 30, 50, 70, 90\}$. We have simulated 10 000 paths of the mortality rates in order to compare the simulated mean mortality rate with the historical mortality rate. We will also include an empirical 95% confidence interval.

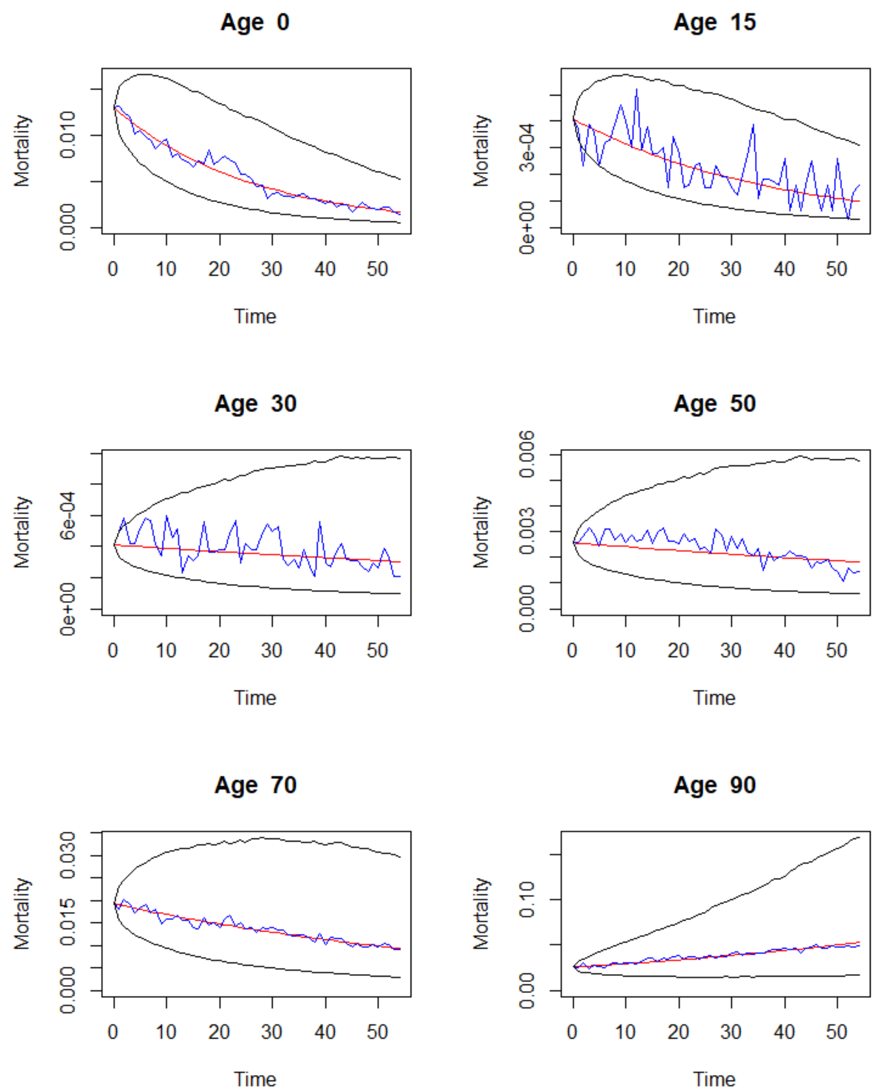

Figure 5.4: Results for women. The simulated mean mortality rate (red), historical mortality rate (blue) and an empirical 95% confidence interval (black).

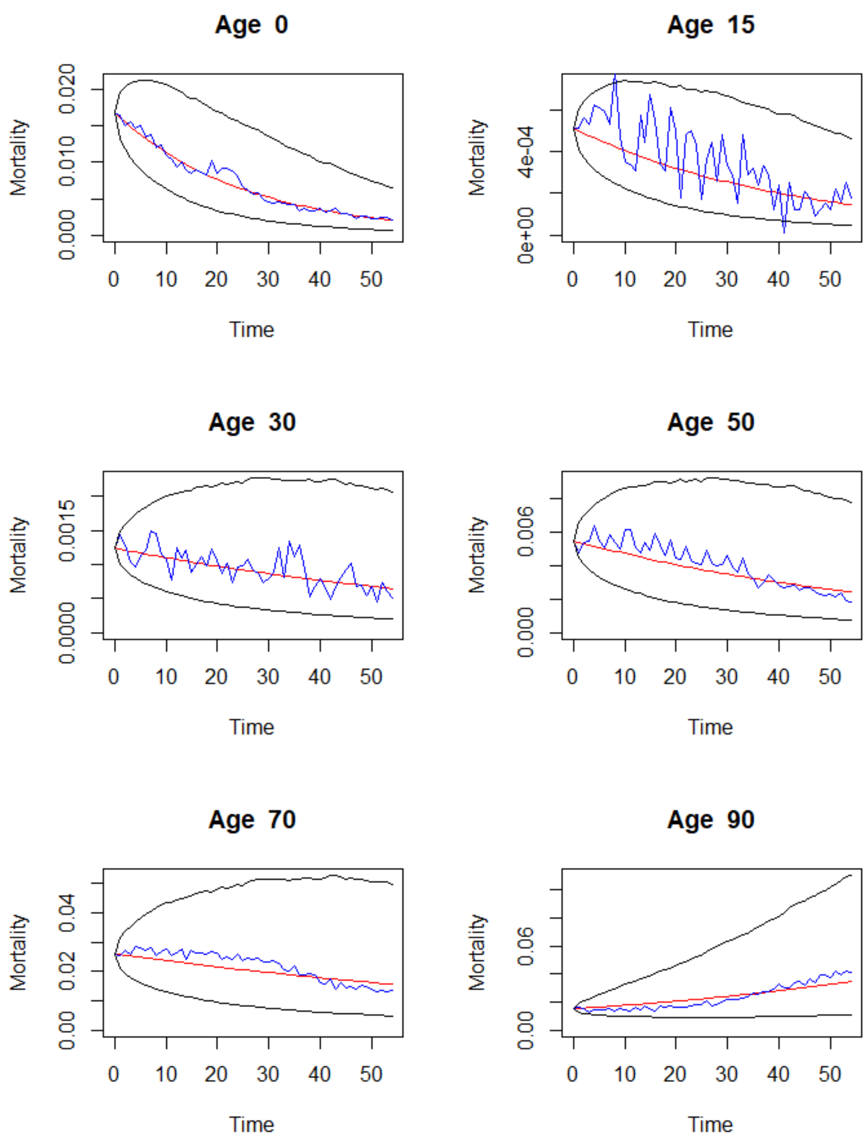

Figure 5.5: Results for men. The simulated mean mortality rate (red), historical mortality rate (blue) and an empirical 95% confidence interval (black).

First, we observe that the 95% confidence interval for men and women is wide, but yet there are instances of rare events where the historical mortality rate intersects the 95% confidence interval. The simulated mean mortality lies close to the historical mortality rate for our selected ages.

5.3 Results of fractional mortality in Norway for $H = 0.75$

The remaining parameters are estimated and presented with the main result of this model given that $H = 0.75$.

Remark 5.3.1 (Results for $H = \frac{3}{4}$) We repeat the analysis with $H = \frac{3}{4}$. First, we present the estimated volatility and estimated drift.

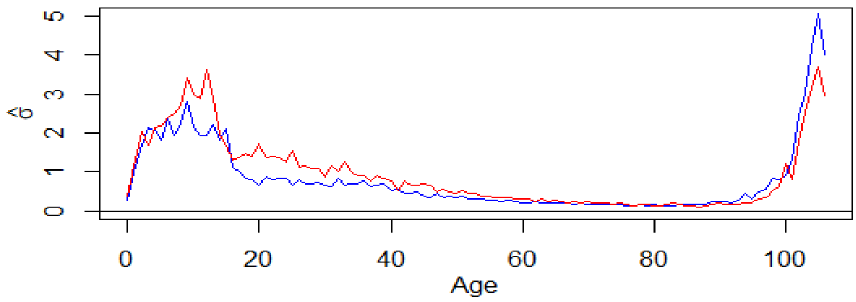

Figure 5.6: Plot showing the evolution of $\widehat{\sigma}$ for men (blue) and women (red). It is reasonable that $\widehat{\sigma}$ is largest for the earliest and oldest ages as the spread in mortality rate becomes huge. In addition, we see that woman tend to have a higher volatility compared to men until age 80. Finally, we see that $\widehat{\sigma}$ increases with H.

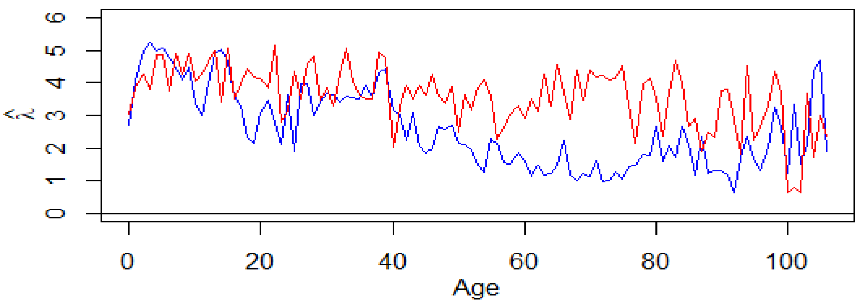

Figure 5.7: Plot showing the evolution of $\widehat{\lambda}$ for men (blue) and women (red). We observe that $\widehat{\lambda}$ tends to follow the shape of $\widehat{\sigma}$. And once again, we see that $\widehat{\lambda}$ increases with H.

For the last time, we present the main result of this model for men and women given the selected representative ages $\{0, 15, 30, 50, 70, 90\}$. We have simulated 10 000 paths of the mortality rates in order to compare the simulated mean mortality rate with the historical mortality rate. We will also include an empirical 95% confidence interval.

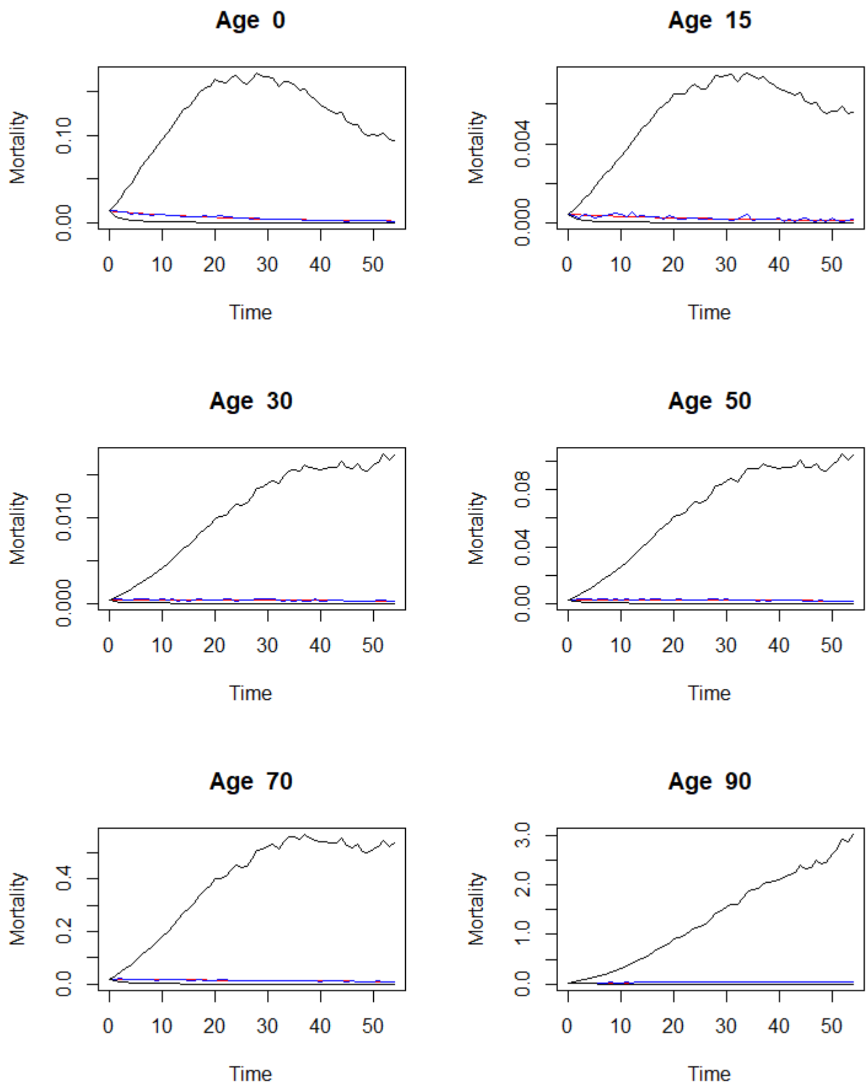

Figure 5.8: Results for women. The simulated mean mortality rate (red), historical mortality rate (blue) and an empirical 95% confidence interval (black).

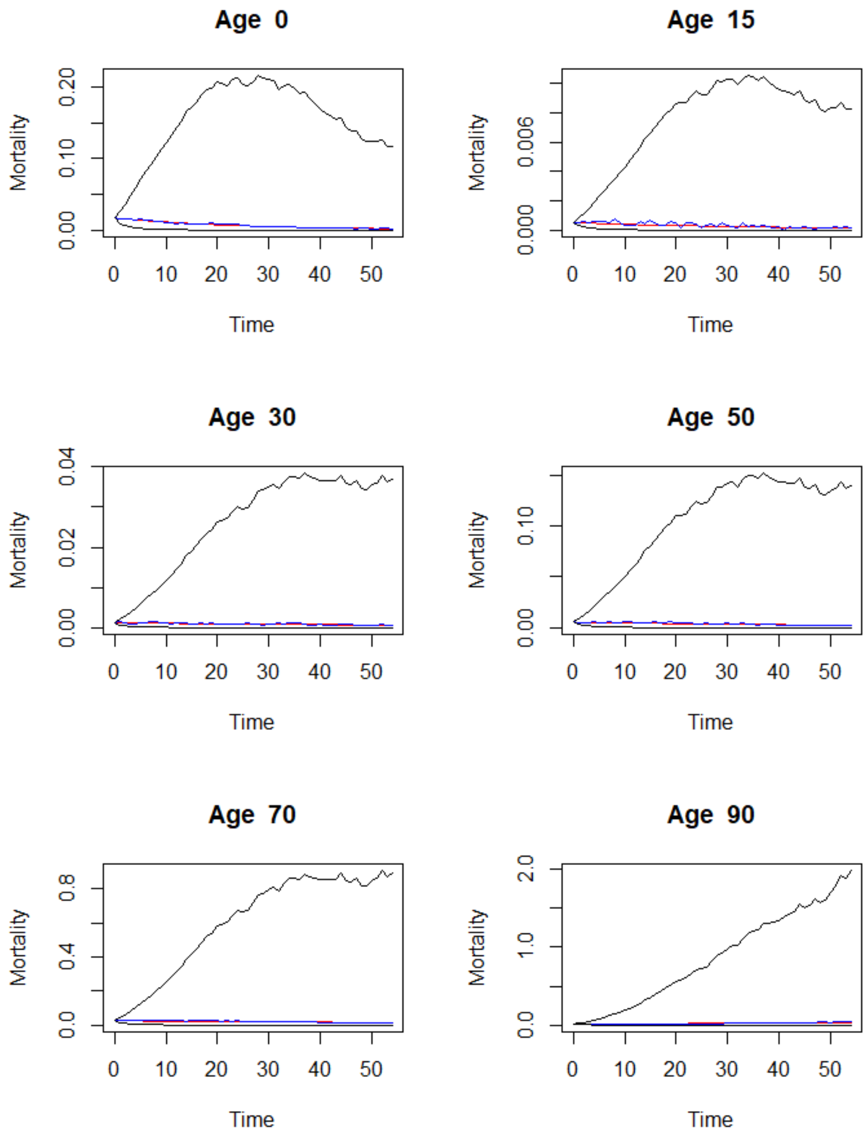

Figure 5.9: Results for men. The simulated mean mortality rate (red), historical mortality rate (blue) and an empirical 95% confidence interval (black).

Since both $\widehat{\sigma}$ and $\widehat{\lambda}$ increase with H, we see a tremendous amount of uncertainty. So much that we are forced to question the underlying assumptions made in this model together with the validity of our results.

5.4 Discussion of the model

It is pointed out in [Gar18] that there exist different notions of a Hurst parameter, which we will use to criticize the model for fractional mortality in Norway. Moreover, as Norwegian mortality rates do not seem to be self-similar, we will also explore the possibility that larger countries exhibit more self-similar mortality rates.

Definition 5.4.1 (Underlying Hurst exponent) Let Y be a time series with underlying noise stemming from a fractional Brownian motion B^H. Then we say that H is the underlying Hurst exponent of Y, that is the Hurst exponent of B^H.

Definition 5.4.2 (Perceived Hurst exponent) Let $\widehat{H}(B^H)$ be an estimator of the fractional Brownian motion's Hurst exponent and let Y be a time series with underlying noise stemming from a fractional Brownian motion B^H. Then we define $\widehat{H}(Y)$ as the perceived Hurst exponent of Y.

The interpretation of the Hurst exponent H is different for B^H and Y. As usual, we say that a fBm with $H > \frac{1}{2}$ is persistent. This simply means that increments of the fBm have a positive autocorrelation. However, this does not mean that the increments of Y have positive autocorrelated increments. Even more alarming is the fact that any attempt to estimate the underlying Hurst parameter from observations of Y are utterly meaningless, as the fBm is not well specified for Y. The main problem is that Y only behaves similarly to a fBm on small scales. Whereas a mean reversion effect dominates the dynamics of Y at high scales. Hence if Y has a perceived $H > \frac{1}{2}$ we can only expect autocorrelation to be positive between close increments of short time duration.

We will now go into details regarding the estimation problems of the Hurst parameter H. Recall that the Hurst parameter is a measurement of self-similarity in a stochastic process. Mathematically, that is

$$B_t = \alpha^{-H} B_{\alpha t}, \qquad \forall \alpha > 0, \qquad t \geq 0.$$

If we think of the fractional Brownian motion B_t^H, we have already shown that it is self-similar of order H in Corollary 3.1.3. A simple way to verify this fact would be to generate one path of the fractional Brownian motion and plot

$$\frac{B_{\alpha t}}{B_t^H} = \alpha^H,$$

which should be similar to a constant straight line. Keeping this in mind, we can see from the Langevin representation of the fractional Ornstein-Uhlenbeck process that Y_t^H is not self-similar. Indeed, we have that

$$\frac{Y_{\alpha t}^H}{Y_t^H} = \frac{Y_0^H - \lambda \int_0^{\alpha t} Y_s^H ds + \sigma B_{\alpha t}^H}{Y_0^H - \lambda \int_0^t Y_s^H ds + \sigma B_t^H},$$

which is clearly not a constant. The difficulty here resides with the complexity of the integral.

The non self-similarity of the mortality observations from Y_t^H, is in itself not problematic as it is only the underlying noise from the fractional Brownian motion which is required to be self-similar. However, there is no reason to believe that the perceived Hurst parameter from Y_t^H is equal to the underlying Hurst parameter from B_t^H. Unless we are dealing with a negligible drift of course, resulting in Y_t^H and B_t^H being similar processes. Thus the assumption that we adopted from [Yer+14], in Assumption 4.2.1 in order to estimate the underlying Hurst parameter, was dubious.

As a consequence of the dubious assumption, the only choice we had to proceed forward was to fix the Hurst parameter of the underlying self-similar noise in Y_t^H outside of the model. The dubious assumption is also the reason our attempt to estimate the underlying Hurst parameter, with the estimator we gave in Remark 4.2.6, led to nonsensical values.

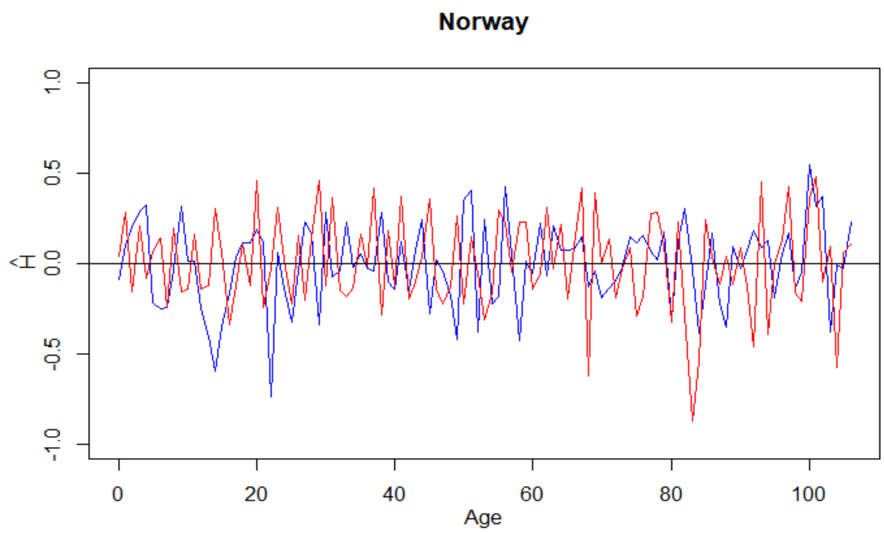

Figure 5.10: Plot showing the evolution of $\widehat{H}$ for men (blue) and women (red). We see that the estimation of the perceived Hurst exponent is problematic as it oscillates around zero which would indicate that Norwegian mortality is not self-similar.

We will in view of this model's poor results in Norway investigate the possibility that fractional mortality is a large population phenomenon.

Remark 5.4.3 (Comparing Hurst parameters of large populations) From [HMD] we have gathered a vast collection of mortality rates from different developed countries in order to compare their underlying Hurst parameter. Moreover, we will divide the estimates based on age, gender and country. Although the database provided is open access, it does require the user to create a free account.

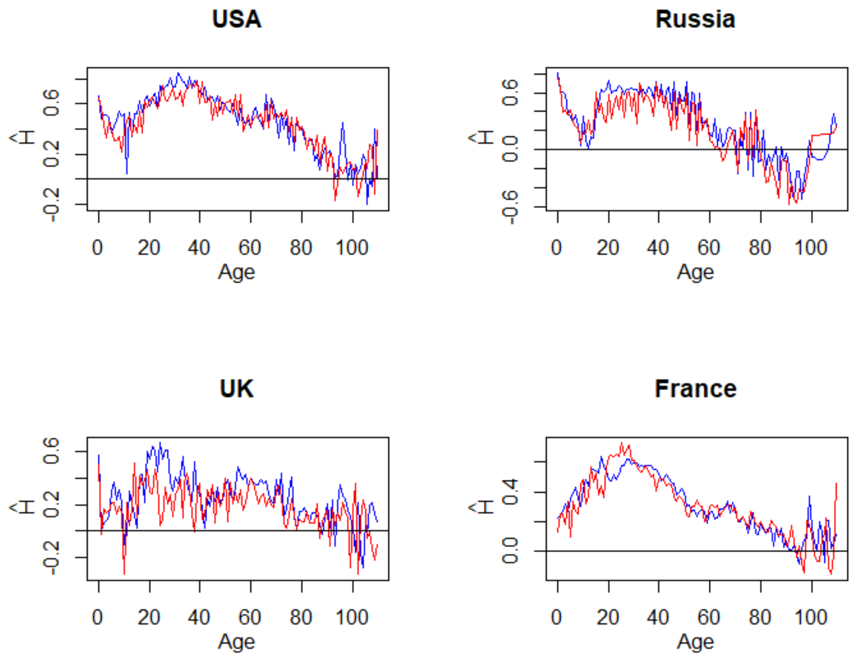

Figure 5.11: Fractionality of large populations. Results for male (blue) and female (red) given an age and a country.

We see that the results of our new model are quite reasonable in large countries within the age group 0 to 80 years. For elderly individuals and small countries the estimation has been shown to be difficult as few people live on, making the data scarce and sensitive to outliers. This is particularly apparent when negative Hurst parameters occur and is a direct consequence of our imperfect Hurst estimator.

Some other difficulties have been in the data itself where some mortality rates where exactly 0. This is problematic as we need to apply the logarithm in the transformation of data and estimation procedure. In order to circumvent this issue, we have replaced zero values within an age group with the mean of the non-zero mortality rates for the same age group. This is not ideal in order to maintain the integrity of our data, but this solution avoids the problem of creating outliers when taking the logarithm of small positive values. The alternative of removing zero valued data in also not appealing as the analysis would require us to work with data of different length.

The difference in fractionality between men and women is surprisingly small within the same country. Men might have a slightly higher Hurst parameter than woman of the same age and country. However, it seems that the choice of country influences fractionality more than the choice of gender given two individuals of same age.

Lastly, we will comment on the shape of the Hurst parameter plots. We can observe an increasing trend in the age group 0 to 30 where it peaks in the latter. Afterwards the Hurst parameter tends to steadily decline until reaching the age of 100. The big variations in the ages 100+ stem from the problems regarding highly sensitive and scarce data. Also note that the data in Russia differs from the rest in the sense that it reaches it peak around age 20 and decreases rapidly. This might be a consequence of the fact that Russia is the country with the lowest life expectancy of the four countries we choose to compare in this remark.

As the Hurst parameter of a large population seems to be close to $H = 0.5$, we will treat this case in more detail under the framework of Itô calculus. But first, we will give a quick introduction on how to calculate prospective reserves in insurance.

CHAPTER 6

Mathematical reserves

The focal point of this chapter will be to develop an explicit formula for insurance mean reserves, that is the expected value of the future liability distribution, in continuous time. This should not be confused with the term reserve or solvency capital which is the capital an insurer is required to set aside by regulators in order to cover future liabilities with high probability. Moreover, we will see that the mean reserving formula can be used to price insurance contracts using the equivalence principle. In what follows, we will adopt the notation and construction of mean reserves given in [Kol12].

6.1 Insurance model in continuous time

This section will introduce the mathematical tools we will need in order to prove the explicit pricing formula.

Let $(\Omega, \mathcal{A}, P)$ be a complete probability space. We emphasize the importance of the completeness as we can not create prefect hedges without the knowledge of nullsets. Let $T > 0$ be a fixed maturity time for when an insurance contract expires. Furthermore, we denote the stochastic process that models the insured's state at any time t by $X = \{X_t, t \in [0, T]\}$. We equip the probability space with the P-augmented and right continuous filtration generated by X, which we will denote by $\mathcal{F}^X = \{\mathcal{F}_t^X, t \in [0, T]\}$. With only this information available, it is reasonable to let $\mathcal{A} = \mathcal{F}_T^X$. In particular, the stochastic process X is an example of a Markov chain with finite state space S described by the insurance contract. Thus all the theory we covered about Markov chains in Section 2.6 can be applied to our setting.

We continue working with the insurance setting described above and note the innately $\mathcal{F}^X$ adapted processes. First of we have the indicator process defined by

$$I_i^X(t) = 1_{\{X_t = i\}}, \qquad i \in S, \qquad t \geq 0.$$

This is a binary process which encodes whether or not the insured is in state i at time t. The second process is called the jump counting process and is defined by

$$N_{ij}^X(t) = \#\{s \in (0, t) : X_{s^-} = i, X_s = j\}, \qquad i, j \in S, \qquad j \neq i,$$

where $\#$ is the counting measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ which keeps track of the number of transitions from state i to state j the insured has in a given time interval $(0, t)$.

Definition 6.1.1 (Stochastic cash flow) Let $A = \{A(t), t \in [0, T]\}$ be a stochastic process. We say that A is a stochastic cash flow if almost all trajectories are of bounded variation.

In particular, we will consider the setting of insurance where the insured's cash flow is determined by policy functions described in the insurance contract.

Definition 6.1.2 (Policy functions) Let $a_i, a_{ij} : [0, \infty) \to \mathbb{R}$ be functions of bounded variation, where $i, j \in S$ and $i \neq j$. We say that the functions are policy functions if

$a_i(t) =$ the accumulated payments from the insurer to the insured up to time t, given that the insured has only been in sate i.

$a_{ij}(t) =$ the punctual payments that must be paid the moment the insured transitions from state i to state j at time t.

For our applications, we will take the policy functions to be deterministic functions. This implies that the cash flow of the insured is completely known and uniquely generated by the policy functions for a fixed state.

Remark 6.1.3 (Modelling cash flow from the insured's perspective) We will follow the convention that payments from the insurer to the insured carries a positive sign, whereas payments from the insured to the insurer constitutes a negative sign. Hence, by the convention it follows that premium payments are of negative sign and insurance benefits are of positive sign.

Example 6.1.4 (Permanent disability insurance) We will continue Example 2.6.3 by determining the relevant policy functions. Recall that in our disability insurance model we have the following state space $S = \{*, \diamond, \dagger\}$, where $*$ means the insured is active, $\diamond$ signifies that the insured is disabled and $\dagger$ means that the insured is dead. The permanent disability insurance contract we will consider states that the insurer must pay the insured a periodic benefit of D monetary units as long as the insured is disabled, this also includes the case where the insured enters the contract disabled. In return, the insurer will receive a periodic premium of π monetary units from the insured while the insured is active. Therefore, the following policy functions uniquely determine the insurance payouts

$$a_*(t) = \begin{cases} -\pi t, & t \in [0, T) \\ -\pi T, & t \in [T, \infty) \end{cases} \qquad a_\diamond(t) = \begin{cases} Dt, & t \in [0, T) \\ DT, & t \in [T, \infty) \end{cases}.$$

As usual, T denotes the maturity date of the insurance contract. It is even possible to consider a contract without maturity date, that is the case when T is infinity. In practice however, most disability contract are no longer than a lifespan which insurance companies set around 114 years.

Definition 6.1.5 (Policy cash flow) Let a_i, a_{ij} be policy functions. Then the (stochastic) policy cash flow of an insurance contract is defined by

$$A(t) = \sum_{i \in S} A_i(t) + \sum_{\substack{i,j \in S \\ j \neq i}} A_{ij}(t),$$

where

$$A_i(t) = \int_0^t I_i^X(s) da_i(s), \qquad A_{ij}(t) = \int_0^t a_{ij}(s) dN_{ij}^X(s).$$

Both integrals make sense as Riemann-Stieltjes integrals since both a_i and N_{ij}^X are of bounded variation. Intuitively, A_i is the accumulated liabilities associated with the insured being in state i and A_{ij} for the case where the insured transitions from state i to state j.

Definition 6.1.6 (Discount factor) Let $v : [0,\infty) \to [0,\infty)$ and $r : [0,\infty) \to \mathbb{R}$ be deterministic functions. If r is an integrable function modelling interest rate, then we define the discount factor v by

$$v(t) = exp\Big(- \int_0^t r(u)du \Big), \qquad t \geq 0.$$

Definition 6.1.7 (Stochastic prospective value of cash flow) Let A be a stochastic cash flow at time t and v an appropriate discount factor. Then we denote the stochastic prospective value of cash flow by $V^+(t,A)$ and define it by

$$V^+(t,A) = \frac{1}{v(t)} \int_t^\infty v(s)dA(s), \qquad t \geq 0.$$

Remark 6.1.8 (Stochastic mean reserve) We can express the stochastic prospective value of cash flow more explicit by applying Definition 6.1.5. Indeed, we can write

$$V^+(t,A) = \frac{1}{v(t)} \Big[\sum_{i\in S} \int_t^\infty v(s)dA_i(s) + \sum_{\substack{i,j\in S\\ j\neq i}} \int_t^\infty v(s)dA_{ij}(s) \Big], \qquad t \geq 0.$$

Furthermore, by the definition of A_i and A_{ij} we have that

$$V^+(t,A) = \frac{1}{v(t)} \Big[\sum_{i\in S} \int_t^\infty v(s)I_i^X(s)da_i(s) + \sum_{\substack{i,j\in S\\ j\neq i}} \int_t^\infty v(s)a_{ij}(s)dN_{ij}^X(s) \Big],$$

where $t \geq 0$.

It is also common to call V^+ the stochastic mean reserve of an insurance policy. The formula we have derived should be intuitive. First, we calculate the future value of an insurance policy by continuously discounting the future payments a_i while the insured is in state i. Secondly, we continuously discount the future payments a_{ij} for all transitions the insured makes from state i to state j. Lastly, we correct the time value to time t by *undiscounting* the sum of all payments over all possible states.

Notice that V^+ is not adapted. Therefore, in order to obtain an estimate of V^+ we must apply the conditional expectation with respect to the available information $\mathcal{F}_t$. In our case, we will only consider the information generated by the insured's state. Using that X is Markovian the expression becomes

$$V_{\mathcal{F}_t}^+(t,A) = E[V^+(t,A)|\mathcal{F}_t] = E[V^+(t,A)|\sigma(X_t)] = H(t,X_t),$$

for some Borel-measurable function H.
Some authors write the previous conditional expectation as

$$V_i^+(t,A) = H(t,i), \qquad i \in S,$$

or simply

$$V_i^+(t,A) = E[V^+(t,A)|X_t = i]. \qquad i \in S.$$

For the sake of keeping the notation simple and clean, we will adopt the latter notation. We now turn to simplifying the mean reserve formula step by step.

Theorem 6.1.9 (First step) Let $J = [\alpha,\beta] \subseteq [t,\infty)$ and $i,j,k \in S$. For an integrable function b and a function c of bounded variation we have that

$$(i) \qquad E\Big[\int_J b(s)dN_{jk}^X(s)\Big|X_t = i\Big] = \int_J b(s)p_{ij}(t,s)\mu_{jk}(s)ds.$$

$$(ii) \qquad E\Big[\int_J 1_{\{X_s=j\}}dc(s)\Big|X_t = i\Big] = \int_J p_{ij}(t,s)dc(s).$$

Proof. (i) We know from linear analysis that the step functions are dense in $L^1(\mathbb{R})$. Moreover, by linearity of the integral it suffices to prove the equality for the simple case where $b(t) = 1_{[\alpha,\beta]}$. Define the function $f : [t,\infty) \to \mathbb{Z}_+$ by

$$f(s) = E[N_{jk}^X(s)|X_t = i], \qquad s \geq t.$$

Then for $h > 0$ we have that

$$f(s+h) - f(s) = E[N_{jk}^X(s+h) - N_{jk}^X(s)|X_t = i]$$

$$= \sum_{l\in S} E[1_{\{X_s=l\}}(N_{jk}^X(s+h) - N_{jk}^X(s))|X_t = i]$$

$$= \sum_{l\in S} \frac{P(X_s = l)}{P(X_t = i)} E[1_{\{X_t=i\}}(N_{jk}^X(s+h) - N_{jk}^X(s))|X_s = l]$$

$$= \sum_{l\in S} \frac{P(X_s = l)}{P(X_t = i)} E[1_{\{X_t=i\}}|X_s = l]E[N_{jk}^X(s+h) - N_{jk}^X(s)|X_s = l]$$

$$= \sum_{l\in S} E[N_{jk}^X(s+h) - N_{jk}^X(s)|X_s = l]p_{il}(t,s).$$

Here we have used the Markov property, also note that the last conditional expectation is of order $o(h)$ for all $l \neq j$. In addition we know that the mapping $s \mapsto X_s(\omega)$ is càdlàg with finite state space S, this yields that

$$\lim_{h\to 0^+} \frac{E[N_{jk}^X(s+h) - N_{jk}^X(s)|X_s = l]}{h} = \begin{cases} \mu_{lk}(s), & l = j \\ 0, & l \neq j \end{cases}.$$

Hence the derivative of f is given by

$$f'(s) = \lim_{h\to 0} \frac{f(s+h) - f(s)}{h} = p_{ij}(t,s)\mu_{jk}(s).$$

By the fundamental theorem of calculus we have that

$$f(\beta)-f(\alpha)=\int_\alpha^\beta f'(s)ds=\int_\alpha^\beta p_{ij}(t,s)\mu_{jk}(s)ds=\int_\alpha^\beta b(s)p_{ij}(t,s)\mu_{jk}(s)ds.$$

Using the definition of f we can also compute the increment directly. Indeed,

$$f(\beta)-f(\alpha)=E[N_{jk}^X(\beta)-N_{jk}^X(\alpha)|X_t=i]$$

$$=E\Big[\int_\alpha^\beta dN_{jk}^X(s)|X_t=i\Big]=E\Big[\int_\alpha^\beta b(s)dN_{jk}^X(s)|X_t=i\Big].$$

Ergo,

$$E\Big[\int_\alpha^\beta b(s)dN_{jk}^X(s)\Big|X_t=i\Big]=\int_\alpha^\beta b(s)p_{ij}(t,s)\mu_{jk}(s)ds.$$

(ii) The second statement is a direct application of Tonelli-Fubini's theorem. First, we must check Tonelli-Fubini's integrability assumption. This is easy,

$$\int_\alpha^\beta E\Big[|1_{\{X_s=j\}}|\Big|X_t=i\Big]dc(s)=\int_\alpha^\beta E\Big[1_{\{X_s=j\}}\Big|X_t=i\Big]dc(s)$$

$$=\int_\alpha^\beta p_{ij}(t,s)dc(s)\le\int_\alpha^\beta dc(s)=c(\beta)-c(\alpha)<\infty.$$

Above we have used that the indicator function is nonnegative in the first equality. That $|p_{ij}|\le 1$ in the first inequality. Finally, we have used that c is of bounded variation in the last inequality.

Thus we can safely interchange the integral and expectation such that

$$E\Big[\int_\alpha^\beta 1_{\{X_s=j\}}dc(s)\Big|X_t=i\Big]=\int_\alpha^\beta E\Big[1_{\{X_s=j\}}\Big|X_t=i\Big]dc(s)=\int_\alpha^\beta p_{ij}(t,s)dc(s).$$

■

Corollary 6.1.10 (Second step) Let x be the insured's age when entering the insurance contract and $t\in[0,\infty)$ be the time spent in the contract. If the insured is in state i at time t, then the present value of cash flows A_j and A_{jk} are given by

$$(i)\qquad V_i^+(t,A_j)=\frac{1}{v(t)}\int_t^\infty v(s)p_{ij}(t+x,s+x)da_j(s).$$

$$(ii)\qquad V_i^+(t,A_{jk})=\frac{1}{v(t)}\int_t^\infty v(s)p_{ij}(t+x,s+x)\mu_{jk}(s+x)a_{jk}(s)ds.$$

Proof. The result is an immediate consequence of Theorem 6.1.9 by inserting

$$b(s)=v(s)a_{jk}(s),\qquad c(s)=\int_0^s v(u)da_j(u).$$

■

Finally, by combining the previous results in this section we obtain the following theorem which summarizes the mean reserve formula in an explicit manner.

Theorem 6.1.11 (Explicit formula for prospective mean reserves) Let x be the insured's age at the beginning of an insurance policy. If the insured is in state i at time t, then the value associated with the liability A at time t with respect to policy functions a_i and a_{ij}, $i, j \in S, j \neq i$, is given by

$$V_i^+(t, A) = \frac{1}{v(t)} \Bigg[\sum_{j \in S} \int_t^\infty v(s) p_{ij}(t+x, s+x) da_j(s)$$

$$+ \sum_{\substack{j,k \in S \\ k \neq j}} \int_t^\infty v(s) p_{ij}(t+x, s+x) \mu_{jk}(s+x) a_{jk}(s) ds \Bigg].$$

We end this section by providing some applications of the mean reserve formula to insurance policies.

Example 6.1.12 (Pricing permanent disability) We consider the permanent disability contract established in Example 6.1.4 and shift our attention to pricing the contract. Assuming that the insured is active when entering the contract we have the following prospective mean reserve

$$V_*^+(t, A) = V_*^+(t, A_*) + V_*^+(t, A_\diamond) + V_*^+(t, A_\dagger), \qquad t \in [0, T].$$

Since there is no cash flow associated with with state †, it is trivial that $V_*^+(t, A_\dagger) = 0$ for all $t \in [0, T]$. Furthermore, by Theorem 6.1.11 we can explicitly express the remaining present values as

$$V_*^+(t, A_*) = \frac{1}{v(t)} \int_t^T v(s) p_{**}(t+x, s+x) da_*(s), \qquad t \in [0, T].$$

$$V_*^+(t, A_\diamond) = \frac{1}{v(t)} \int_t^T v(s) p_{*\diamond}(t+x, s+x) da_\diamond(s), \qquad t \in [0, T].$$

In addition, we have that the policy functions are differentiable almost everywhere

$$a_*'(t) = \begin{cases} -\pi, & t \in (0, T) \\ 0, & else \end{cases} \qquad a_\diamond'(t) = \begin{cases} D, & t \in (0, T) \\ 0, & else \end{cases}.$$

Using Remark 2.3.23 for computing Riemann-Stieltjes integrals, noting that the policy functions are continuous and a.e. differentiable and using linearity of integrals, we can simplify the present values to the following Riemann integrals

$$V_*^+(t, A_*) = -\pi \frac{1}{v(t)} \int_t^T v(s) p_{**}(t+x, s+x) ds, \qquad t \in [0, T].$$

$$V_*^+(t, A_\diamond) = D \frac{1}{v(t)} \int_t^T v(s) p_{*\diamond}(t+x, s+x) ds, \qquad t \in [0, T].$$

To find the fair price of this policy we will use the equivalence principle which states that the value of premiums are determined such that the policy has null present value at the onset of the contract. In particular, at the start of the contract we have that

$$V_*^+(0, A) = V_*^+(0, A_*) + V_*^+(0, A_\diamond) = 0.$$

Solving the resulting equation with respect to π we get the fair premium

$$\pi = D\frac{\int_t^T v(s)p_{*\diamond}(t+x, s+x)ds}{\int_t^T v(s)p_{**}(t+x, s+x)ds}.$$

Intuitively, the premium is determined by the size of the disability pension D and whether $p_{*\diamond}$ is large compared to p_{**}.

Example 6.1.13 (Endowment insurance in Norway) We will consider an endowment insurance which has the following state space $S = \{*, \dagger\}$. The policy yields a payoff of B monetary units to the insured in the case of an early death and E monetary units in the case of reaching the maturity date T. Continuing quid pro quo, the insured pays a periodic premium of π monetary units to the insurer. Then the relevant policy functions which completely determine this insurance are given by

$$a_*(t) = \begin{cases} -\pi t, & t \in [0, T) \\ -\pi T + E, & t \in [T, \infty) \end{cases} \qquad a_{*\dagger}(t) = \begin{cases} B, & t \in [0, T) \\ 0, & else \end{cases}.$$

The mean reserve of this policy is given by

$$V_*^+(t, A) = V_*^+(t, A_*) + V_*^+(t, A_{*\dagger}), \qquad t \in [0, T].$$

Using Theorem 6.1.11 we have the following representation for the present values of each benefit

$$V_*^+(t, A_*) = \frac{1}{v(t)}\int_t^T v(s)p_{**}(t+x, s+x)da_*(s), \qquad t \in [0, T].$$

$$V_*^+(t, A_{*\dagger}) = \frac{1}{v(t)}\int_t^T v(s)p_{**}(t+x, s+x)\mu_{*\dagger}(s+x)a_{*\dagger}(s)ds, \qquad t \in [0, T].$$

It is easy to compute the a.e. derivative of the policy function a_*, which is given by

$$a_*'(t) = \begin{cases} -\pi, & t \in (0, T) \\ 0, & else \end{cases}.$$

We observe that a_* is discontinuous at time T and a.e. differentiable. Therefore we can apply Remark 2.3.23 and linearity of the integral to simplify the formula for $V_*^+(t, A_*)$. In the case of $V_*^+(t, A_{*\dagger})$ we just need linearity of the integral. For $t \in [0, T]$ we get the following present values

$$V_*^+(t, A_*) = \frac{1}{v(t)}\Big[-\pi\int_t^T v(s)p_{**}(t+x, s+x)ds + Ev(T)p_{**}(t+x, T+x)\Big].$$

$$V_*^+(t, A_{*\dagger}) = \frac{1}{v(t)} B \int_t^T v(s) p_{**}(t+x, s+x) \mu_{*\dagger}(s+x) ds.$$

The fair premium of this insurance is once again given by the equivalence principle which boils down to solving the following equation with respect to π

$$V_*^+(0, A) = V_*^+(0, A_*) + V_*^+(0, A_{*\dagger}) = 0.$$

After rearranging the terms we arrive at the solution

$$\pi = \frac{Ev(T)p_{**}(x, T+x) + B\int_0^T v(s)p_{**}(x, s+x)\mu_{*\dagger}(s+x)ds}{\int_0^T v(s)p_{**}(x, s+x)ds}.$$

Finally, the present value of benefits, the present value of premiums and the mean reserve of this policy are given by

$$V_*^+(t, A^{benefit}) = \frac{1}{v(t)} \Big[B \int_t^T v(s) p_{**}(t+x, s+x) \mu_{*\dagger}(s+x) ds$$
$$+ Ev(T) p_{**}(t+x, T+x) \Big].$$

$$V_*^+(t, A^{premium}) = \frac{-\pi}{v(t)} \int_t^T v(s) p_{**}(t+x, s+x) ds.$$

$$V_*^+(t, A) = V_*^+(t, A^{benefit}) + V_*^+(t, A^{premium}).$$

We have collected data regarding Norwegian mortality rates from [SSBa]. Moving forward, we proceed by making a plot of the combined mortality rate for both genders in Norway.

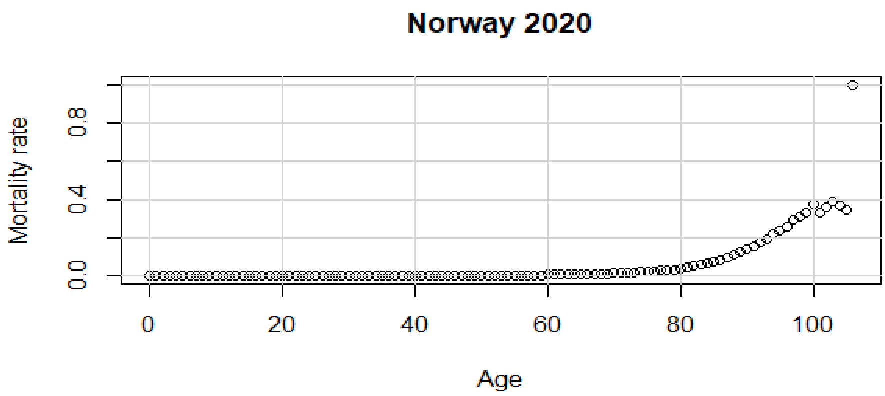

Figure 6.1: Scatter plot of Norwegian mortality rates from 2020 for both genders combined.

The mortality rate seems to grow exponentially with age as one could expect. Therefore a suitable model would be Gompertz-Makeham's law of mortality which states that the rate of transitioning from alive to dead is given by the function

$$\mu(t) = \alpha_0 + \alpha_1 e^{\alpha_2 t}, \qquad \alpha_0, \alpha_1, \alpha_2 > 0, \qquad t \geq 0.$$

We will use the software package "mosaic" in R to estimate the parameters of the curve that best fit the data. After doing so we obtain the following estimates

$$\widehat{\alpha}_0 = 2.962978 \cdot 10^{-4}, \qquad \widehat{\alpha}_1 = 1.178166 \cdot 10^{-5}, \qquad \widehat{\alpha}_2 = 1.028398 \cdot 10^{-1}.$$

Now we plot the estimated curve with the data.

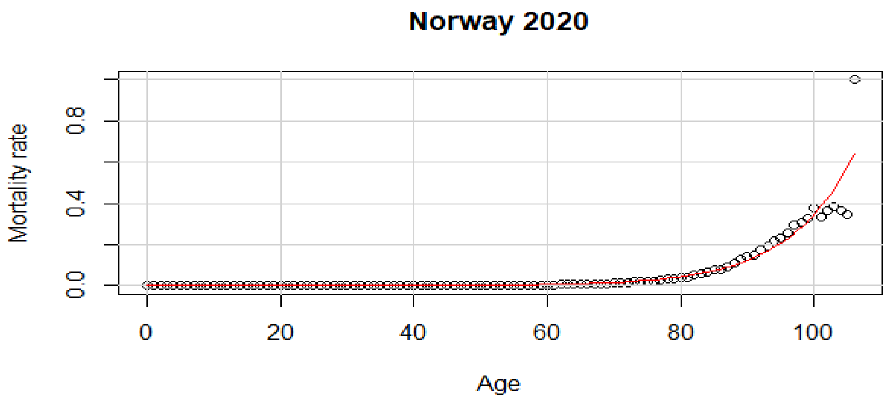

Figure 6.2: Data together with the best fitted Gompertz-Makeham curve.

We see in Figure 6.2 that the curve fits the data well up to age 80. After 80 the curve tends to underestimate the empirical mortality. Moreover, the data suggests that the empirical mortality rate increases at a decreasing rate. This phenomenon is known as late-life mortality deceleration in the field of gerontology and was first proposed in human aging by Gompertz himself in the paper [Gom25]. Although this phenomenon has become a disputed topic in recent years, for instance see the following paper [GG11].

Finally, it is worth commenting that we see a clear outlier at the age of 106. The outlier might stem from lack of reliable data at such high ages and it would therefore be reasonable to exclude it when performing statistical analysis.

Proceeding forward, we have that the survival probability in our model is computed by Theorem 2.6.14 as

$$p_{**}(t,s) = exp\Big(-\int_t^s \mu(u)du\Big) = exp\Big(-\alpha_0(s-t) - \frac{\alpha_1}{\alpha_2}(e^{\alpha_2 s} - e^{\alpha_2 t})\Big), \qquad t \leq s.$$

We finish this example by calculating the yearly premium and plotting the present values with the mean reserve in the case where the insured enters the endowment contract at age 30 with maturity date in 40 years. Moreover, we will assume that the interest rate remains fixed at 2% during the whole contract. Finally, we set the endowment to be 100 000 NOK and the death benefit to 200 000 NOK. Using the expressions we have derived and computing the integrals numerically by Simpson's rule we obtain the yearly premium

$$\pi \approx 2\ 062\ NOK.$$

We summarize the present values in one final plot given the parameters above.

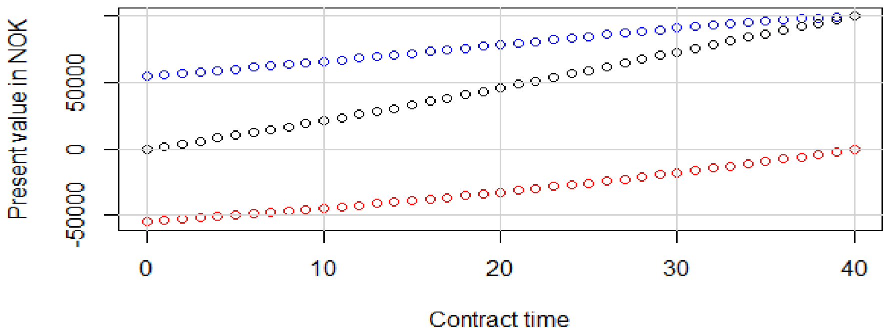

Figure 6.3: Plot showing the present value of benefits (blue), mean reserve (black) and present value of premiums (red).

All values obtained are reasonable. Note that the computations were made well inside the age window where Gompertz-Makeham's mortality rate fits the data fairly well.

6.2 Thiele's ordinary differential equation

We will end this chapter with a powerful result known as Thiele's ordinary differential equation. It is a linear differential equation for finding V_i^+, at each time t, without invoking transition probabilities, but rather transition rates.

Theorem 6.2.1 (Thiele's ODE) Let x be the insured's age at the beginning of the contract, t the time passed since the contract started and T the contract's maturity date. Assume that $a_i, a_{ij}, i, j \in S$ are policy functions. If a_i is almost everywhere differentiable with at most one discontinuity at time $t = T$, then the mean reserve $V_i^+(t)$ is given by

$$\frac{d}{dt}V_i^+(t) = r(t)V_i^+(t) - a_i'(t) - \sum_{\substack{j\in S\\ j\neq i}} \mu_{ij}(x+t)\Big(a_{ij}(t) + V_j^+(t) - V_i^+(t)\Big),$$

with terminal condition

$$V_i^+(T) = a_i(T) - a_i(T-).$$

Proof. First, we use that a_i is a.e. differentiable with at most one discontinuity at time $t = T$. By using Remark 2.3.23 in conjunction with Theorem 6.1.11 we can write the explicit formula for the prospective mean reserve as

$$\begin{aligned} V_i^+(t) = \frac{1}{v(t)}\Bigg[&\sum_{j\in S} v(T)p_{ij}(t+x, T+x)\Big(a_j(T) - a_j(T-)\Big) \\ &+ \sum_{j\in S}\int_t^T v(s)p_{ij}(t+x, s+x)\theta_j^x(s)ds\Bigg], \end{aligned}$$

where

$$\theta_j^x(s) = a_j'(s) + \sum_{\substack{k\in S\\ k\neq j}} \mu_{jk}(s+x)a_{jk}(s).$$

Next, we rewrite the equation for V_i^+ in the following compressed form

$$v(t)V_i^+(t) = G_i^T(t) + \int_t^T F_i^s(t)ds, \tag{6.2.1}$$

where

$$G_i^T(t) = v(T)\sum_{j\in S} p_{ij}(t+x, T+x)\Big(a_j(T) - a_j(T-)\Big),$$

$$F_i^s(t) = v(s)\sum_{j\in S} p_{ij}(t+x, s+x)\theta_j^x(s).$$

Since each sum in G_i^T and F_i^s is finite, we can differentiate with respect to t yielding

$$\frac{d}{dt}G_i^T(t) = v(T)\sum_{j\in S}\frac{d}{dt}p_{ij}(t+x,T+x)\Big(a_j(T)-a_j(T-)\Big),$$

$$\frac{d}{dt}F_i^s(t) = v(s)\sum_{j\in S}\frac{d}{dt}p_{ij}(t+x,s+x)\theta_j^x(s).$$

Kolmogorov's backward equation implies that

$$\begin{aligned}\frac{d}{dt}p_{ij}(t+x,s+x) &= \mu_i(t+x)p_{ij}(t+x,s+x) - \sum_{\substack{k\in S\\k\neq i}}\mu_{ik}(t+x)p_{kj}(t+x,s+x)\\ &= \sum_{\substack{k\in S\\k\neq i}}\mu_{ik}(t+x)\Big(p_{ij}(t+x,s+x)-p_{kj}(t+x,s+x)\Big).\end{aligned}$$

Inserting the former identity into $\frac{d}{dt}F_i^s$ we get that

$$\begin{aligned}\frac{d}{dt}F_i^s(t) &= v(s)\sum_{j\in S}\left(\sum_{\substack{k\in S\\k\neq i}}\mu_{ik}(t+x)\Big(p_{ij}(t+x,s+x)-p_{kj}(t+x,s+x)\Big)\right)\theta_j^x(s)\\ &= \sum_{\substack{k\in S\\k\neq i}}\mu_{ik}(t+x)\Big(F_i^s(t)-F_k^s(t)\Big).\end{aligned} \tag{6.2.2}$$

A similar computation gives that

$$\frac{d}{dt}G_i^T(t) = \sum_{\substack{k\in S\\k\neq i}}\mu_{ik}(t+x)\Big(G_i^T(t)-G_k^T(t)\Big). \tag{6.2.3}$$

The next step is to differentiate (6.2.1). On the left hand side we have that

$$\frac{d}{dt}\big[v(t)V_i^+(t)\big] = -r(t)v(t)V_i^+(t) + v(t)\frac{d}{dt}V_i^+(t). \tag{6.2.4}$$

Define the following integral function by

$$I(z,w) = \int_z^T F_i^s(w)ds.$$

Computing the partial derivatives of I yield that

$$\partial_z I(z,w) = -F_i^z(w), \qquad \partial_w I(z,w) = \int_z^T \partial_w F_i^s(w)ds.$$

Combining the chain rule with the fundamental theorem of calculus, allows us to evaluate

$$\frac{d}{dt}I(t,t) = \partial_z I(z,w)\Big|_{(z,w)=(t,t)} + \partial_w I(z,w)\Big|_{(z,w)=(t,t)} = -F_i^t(t) + \int_t^T \partial_t F_i^s(t)ds.$$

We see that

$$F_i^t(t) = \lim_{s\to t^+} F_i^s(t) = v(t)\sum_{j\in S} p_{ij}(t+x,t+x)\theta_j^x(t)$$
$$= v(t)\theta_i^x(t) = v(t)\Big(a_i'(t) + \sum_{\substack{k\in S\\ k\neq i}} \mu_{ik}(t+x)a_{ik}(t)\Big).$$

Finally, we differentiate the right-hand side of (6.2.1) and obtain that

$$\partial_t\Big(G_i^T + \int_t^T F_i^s(t)ds\Big) = \partial_t G_i^T(t) - v(t)a_i'(t) - v(t)\sum_{\substack{k\in S\\ k\neq i}} \mu_{ik}(t+x)a_{ik}(t) \quad (6.2.5)$$
$$+ \int_t^T \partial_t F_i^s(t)ds.$$

Setting (6.2.4) equal to (6.2.5) and isolating the integral term gives that

$$\int_t^T \partial_t F_i^s(t)ds = -r(t)v(t)V_i^+(t) + v(t)\frac{d}{dt}V_i^+(t) - \partial_t G_i^T(t) \quad (6.2.6)$$
$$+v(t)a_i'(t) + v(t)\sum_{\substack{k\in S\\ k\neq i}} \mu_{ik}(t+x)a_{ik}(t).$$

To find another expression for the integral above, we will integrate (6.2.2) on both sides with respect to s on the interval $[t,T]$. Also by interchanging the order of integration and summation together with the relation (6.2.1), we can write

$$\int_t^T \partial_t F_i^s(t)ds = v(t)\sum_{\substack{k\in S\\ k\neq i}} \mu_{ik}(t+x)\Big(V_i^+(t+x) - V_k^+(t+x)\Big)$$
$$-\sum_{\substack{k\in S\\ k\neq i}} \mu_{ik}(t+x)\Big(G_i^T(t) - G_k^T(t)\Big).$$

However, last term above can be rewritten using (6.2.3) such that

$$\int_t^T \partial_t F_i^s(t)ds = v(t)\sum_{\substack{k\in S\\ k\neq i}} \mu_{ik}(t+x)\Big(V_i^+(t+x) - V_k^+(t+x)\Big) - \partial_t G_i^T(t). \quad (6.2.7)$$

Lastly, we set (6.2.6) equal to (6.2.7), cancel redundant terms and isolate the term $\frac{d}{dt}V_i^+$. This gives the desired result

$$\frac{d}{dt}V_i^+(t) = r(t)V_i^+(t) - a_i'(t) - \sum_{\substack{j\in S\\ j\neq i}} \mu_{ij}(x+t)\Big(a_{ij}(t) + V_j^+(t) - V_i^+(t)\Big).$$

■

CHAPTER 7

Insurance policies with stochastic mortality

Finally we have reached the final chapter of this thesis. We will now deal with actuarial computations regarding reserving and pricing under stochastic mortality. For our applications, we will limit the analysis of insurance contracts to the simple case where $S = \{*, \dagger\}$. This simplifies the formula for the mean reserve to the following expression

$$V_*^+(t) = \frac{1}{v(t)}\Bigg[\int_t^\infty v(s)p_{**}(x+t, x+s)da_*(s) \\ + \int_t^\infty v(s)p_{**}(x+t, x+s)\mu_{*\dagger}(x+s)a_{*\dagger}(s)ds\Bigg],$$

where the survival probability is given by

$$p_{**}(x+t, x+s) = exp\bigg(-\int_t^s \mu_{*\dagger}(x+u)du\bigg), \qquad t \le s.$$

In our context, the mortality $\mu_{*\dagger}$ is replaced by the stochastic mortality μ. Thus, we need to apply conditional expectation in order to obtain adapted reserves. That is,

$$V_*^+(t, \mu_t) = \int_t^T \frac{v(s)}{v(t)} E\Big[e^{-\int_t^s \mu_u du}\Big|\mathcal{F}_t\Big]da_*(s) \\ + \int_t^T \frac{v(s)}{v(t)} E\Big[e^{-\int_t^s \mu_u du}\mu_s\Big|\mathcal{F}_t\Big]a_{*\dagger}(s)ds.$$

Computing the survival probability is a demanding task in general, which requires numerical integration. For that reason it is more practical to use Thiele's differential equation, which only requires $\mu(t)$. In our case Thiele's differential equation becomes

$$\frac{d}{dt}V_*^+(t) = r(t)V_*^+(t) - a_*'(t) - \mu_{*\dagger}(x+t)\Big(a_{*\dagger}(t) + V_\dagger^+(t) - V_*^+(t)\Big),$$

with terminal condition

$$V_*^+(T) = a_*(T) - a_*(T-).$$

7.1 General stochastic mortality model

In this section, we will consider the case of a general Markov mortality in order to derive an analytic formula for the prospective reserve. Before proceeding with the analysis however, we wish to caution the reader against two important details. Firstly, what we denote by μ in this chapter is not the generational mortality of a given age, but rather the mortality of an individual of a given age as time passes. Secondly, that we are working with a Wiener process which satisfies the Markov property and the framework of Itô calculus.

Let μ be the insured's mortality rate. We will assume that μ can be modeled by the following stochastic differential equation

$$d\mu_t = b(t, \mu_t)dt + \sigma(t, \mu_t)dW_t,$$

where b and σ are measurable functions for which a unique strong solution exists.

We introduce some examples of mortality rate models which fit our framework and also restrict μ to positive values.

Example 7.1.1 (Geometric Brownian motion) The solution of the following SDE is known as the geometric Brownian motion

$$d\mu_t = b\mu_t \, dt + \sigma\mu_t \, dW_t, \qquad b, \sigma \in \mathbb{R}, \qquad \mu_0 > 0.$$

The geometric Brownian motion is a well known SDE which is useful in predicting the prices of future commodities.

Example 7.1.2 (Cox-Ingersoll-Ross) The solution of the following SDE is known as the CIR model

$$d\mu_t = a(b - \mu_t)dt + \sigma\sqrt{\mu_t}dW_t, \qquad a, b, \sigma > 0.$$

The CIR model is often used to describe the evolution of interest rates and is an extension of the Vasicek model.

Example 7.1.3 (Log-Ornstein-Uhlenbeck) The solution of the following SDE is known as the log-Ornstein-Uhlenbeck process

$$d\mu_t = \mu_t\Big(\alpha + \frac{1}{2}\sigma^2 + \lambda ln(\mu_0) + \alpha\lambda t - \lambda ln(\mu_t)\Big)dt + \sigma\mu_t dW_t, \ \mu_0, \lambda, \sigma > 0, \ \alpha \in \mathbb{R}.$$

This SDE will be the topic of the final section in this thesis which explores applications to insurance contracts. In particular computing the fair premium and prospective reserves given the log-Ornstein-Uhlenbeck mortality rate model.

7.2 Thiele's PDE for stochastic mortality models

Next theorem is the corresponding Thiele's differential equation when the mortality is seen as a Markovian Itô process and it is one of the main contributions of this thesis.

Theorem 7.2.1 (Thiele's PDE) Let $V_*^+(t,x)$ be the value of the insurance contract at time t with mortality level x, given that the insured is alive. Assume that the insured's mortality rate μ satisfy the following stochastic differential equation

$$d\mu_t = b(t,\mu_t)dt + \sigma(t,\mu_t)dW_t,$$

where b and σ are measurable functions for which a unique strong solution exists. If $a_*, a_{*\dagger}$ are policy functions, where a_* is almost everywhere differentiable, and r is a measurable function modeling the interest rate, then V_*^+ solves the following PDE

$$\partial_t V_*^+(t,x) = r(t)V_*^+(t,x) - a_*'(t) + \Big[V_*^+(t,x) - a_{*\dagger}(t)\Big]x$$

$$-b(t,x)\partial_x V_*^+(t,x) - \frac{1}{2}\sigma^2(t,x)\partial_x^2 V_*^+(t,x),$$

with terminal condition

$$V_*^+(T,x) = a_*(T) - a_*(T-).$$

Proof. Fix $T > 0$ and let $b, \sigma : [0,T]\times\mathbb{R}\to\mathbb{R}$ be Borel measurable functions. Consider the following stochastic differential equation

$$d\mu_t = b(t,\mu_t)dt + \sigma(t,\mu_t)dW_t.$$

Furthermore, we define the function $u : [0,T]\times\mathbb{R}\to\mathbb{R}$ by

$$u_T^\Psi(t,x) = E\Big[e^{-\int_t^T \mu_s ds}\Psi(\mu_T)\Big|\mu_t = x\Big],$$

where $\Psi : \mathbb{R}\to\mathbb{R}$ is a deterministic and Borel measurable function. By Markovianity we have that

$$u_T^\Psi(t,x) = E\Big[e^{-\int_t^T \mu_s ds}\Psi(\mu_T)\Big|\sigma\big(\mu_s, 0\le s\le t\big)\Big].$$

Let L be the differential operator defined by

$$L(u) = b(t,x)\partial_x u + \frac{1}{2}\sigma^2(t,x)\partial_x^2 u, \qquad u\in C^{1,2}([0,T]\times\mathbb{R}).$$

Then, by the Feynman-Kac formula, the function $(t,x)\mapsto u_T^\Psi(t,x)$ solves the following partial differential equation

$$\partial_t u + L(u) = xu,$$

with terminal condition

$$u(T,x) = \Psi(x).$$

Combining our expression of the adapted reserve with the Feynman Kac PDE we obtain a new expression for the prospective reserve with respect to μ_t. Indeed,

$$V_*^+(t,\mu_t) = \int_t^T \frac{v(s)}{v(t)} u_s^{\Psi_1}(t,\mu_t)da_*(s) + \int_t^T \frac{v(s)}{v(t)} u_s^{\Psi_2}(t,\mu_t)a_{*\dagger}(s)ds,$$

where Ψ_1 is the constant function 1 and Ψ_2 is the identity function. Since the prospective reserve $V_*^+ \in C^{1,2}([0,T]\times\mathbb{R})$ we can apply Itô's formula and evaluate at $(t,x) = (t,\mu_t)$. This yields

$$dV_*^+ = \partial_t V_*^+(t,\mu_t)dt + \partial_x V_*^+(t,\mu_t)d\mu_t + \frac{1}{2}\partial_x^2 V_*^+(t,\mu_t)d[\mu,\mu]_t.$$

Inserting the dynamics of μ we get

$$dV_*^+ = \partial_t V_*^+(t,\mu_t)dt + \partial_x V_*^+(t,\mu_t)\Big(b(t,\mu_t)dt + \sigma(t,\mu_t)dW_t\Big)$$

$$+\frac{1}{2}\partial_x^2 V_*^+(t,\mu_t)\sigma^2(t,\mu_t)dt$$

$$= \Big[\partial_t V_*^+(t,\mu_t) + b(t,\mu_t)\partial_x V_*^+(t,\mu_t) + \frac{1}{2}\partial_x^2 V_*^+(t,\mu_t)\sigma^2(t,\mu_t)\Big]dt + \sigma(t,\mu_t)dW_t.$$

The next step, is to calculate the partial derivatives. Differentiating once in space gives

$$\partial_x V_*^+ = \int_t^T \frac{v(s)}{v(t)}\partial_x u_s^{\Psi_1}(t,x)da_*(s) + \int_t^T \frac{v(s)}{v(t)}\partial_x u_s^{\Psi_2}(t,x)a_{*\dagger}(s)ds.$$

Similarly after differentiating twice in space we obtain

$$\partial_x^2 V_*^+ = \int_t^T \frac{v(s)}{v(t)}\partial_x^2 u_s^{\Psi_1}(t,x)da_*(s) + \int_t^T \frac{v(s)}{v(t)}\partial_x^2 u_s^{\Psi_2}(t,x)a_{*\dagger}(s)ds.$$

Before computing the partial derivative in time, we will for convenience work under the assumption that a_* is a.e. differentiable with at most one discontinuity at $t = T$. This assumption allows us to write

$$da_*(s) = \Delta a_*(s)\delta_T(s)ds + a_*'(s)ds,$$

where $\delta_T(s)$ is the Dirac measure on T such that

$$\int_t^T \Delta a_*(s)\delta_T(s)ds = \begin{cases} a_*(T) - a_*(T-), & s = T \\ 0, & else. \end{cases}$$

Thus when a_* is a.e. differentiable with a discontinuity at $t = T$ we can burn the Lebesgue-Stieltjes integral appearing in the expression of $V_*^+(t,x)$ into a Lebesgue integral. Using linearity of the Lebesgue integral we arrive at

$$V_*^+(t,x) = \frac{v(T)}{v(t)} u_T^{\Psi_1}(t,x)\Delta a_*(T) + \int_t^T \frac{v(s)}{v(t)} \Big[u_s^{\Psi_1}(t,x) a_*'(s) + u_s^{\Psi_2}(t,x) a_{*\dagger}(s) \Big] ds$$

$$= \frac{v(T)}{v(t)} u_T^{\Psi_1}(t,x)\Delta a_*(T) + \int_t^T \frac{v(s)}{v(t)} u_s^{\Psi^s}(t,x) ds,$$

where we have introduces the following notations

$$u_s^{\Psi^s}(t,x) = u_s^{\Psi_1}(t,x) a_*'(s) + u_s^{\Psi_2}(t,x) a_{*\dagger}(s),$$

$$\Psi^s(x) = a_*'(s)\Psi_1(s) + a_{*\dagger}(s)\Psi_2(s) = a_*'(s) + a_{*\dagger}(s)x.$$

We call attention to the notation Ψ^s in order to remember the dependence on s, however we recall that s is fixed when solving the PDE with respect to u. That is $u_s^{\Psi^s}$ is a function of (t,x) which solves the PDE

$$\partial_t u + L(u) = xu,$$

with terminal condition

$$u(s,x) = \Psi^s(x).$$

We now turn our attention back to finding the partial derivative in time for the prospective reserve. To simplify the computation, we send the discount factor $v(t)$ to the left side of the equation yielding

$$V_*^+(t,x)v(t) = v(T) u_T^{\Psi_1}(t,x)\Delta a_*(T) + \int_t^T v(s) u_s^{\Psi^s}(t,x) ds.$$

Finally, we use the product rule on the left side and the fundamental theorem of analysis on the right side to compute the partial derivative in time. This gives

$$\partial_t V_*^+(t,x)v(t) - V_*^+(t,x) r(t) v(t)$$

$$= v(T)\Delta a_*(T) \partial_t u_T^{\Psi_1}(t,x) - v(t) u_t^{\Psi^t}(t,x) + \int_t^T v(s) \partial_t u_s^{\Psi^s}(t,x) ds,$$

where we have used that

$$\partial_t v(t) = \partial_t e^{-\int_0^t r(u)du} = -r(t)v(t).$$

Note that we can exchange $\partial_t u$ by rearranging the Feynman-Kac formula as

$$\partial_t u = xu - L(u).$$

Furthermore, we will make use of the relation

$$u_t^{\Psi^t}(t,x) = \Psi^t(x) = a_*'(t) + a_{*\dagger}(t)x.$$

After inserting the two identities, we have that

$$\partial_t V_*^+(t,x)v(t) - V_*^+(t,x)r(t)v(t) = v(T)\Delta a_*(T)\Big[xu_T^{\Psi_1}(t,x) - L\Big(u_T^{\Psi_1}(t,x)\Big)\Big]$$

$$-v(t)\Big(a_*'(t) + a_{*\dagger}(t)x\Big) + \int_t^T v(s)\Big[xu_s^{\Psi^s}(t,x) - L\Big(u_s^{\Psi^s}(t,x)\Big)\Big]ds.$$

Collecting terms containing x and L gives

$$\partial_t V_*^+(t,x)v(t) - V_*^+(t,x)r(t)v(t)$$

$$= \left[v(T)\Delta a_*(T)u_T^{\Psi_1}(t,x) + \int_t^T v(s)u_s^{\Psi^s}(t,x)ds - v(t)a_{*\dagger}(t)\right]x$$

$$-L\left(v(T)\Delta a_*(T)u_T^{\Psi_1}(t,\cdot) + \int_t^T v(s)u_s^{\Psi^s}(t,\cdot)ds\right)$$

$$-v(t)a_*'(t),$$

where we have used linearity of the operator L.
We observe that

$$v(t)V_*^+(t,x) = v(T)\Delta a_*(T)u_T^{\Psi_1}(t,x) + \int_t^T v(s)u_s^{\Psi^s}(t,x).$$

Inserting the previous observation we achieve the formula

$$\partial_t V_*^+(t,x)v(t) - V_*^+(t,x)r(t)v(t)$$

$$= \Big[v(t)V_*^+(t,x) - v(t)a_{*\dagger}(t)\Big]x - L\Big(v(t)V_*^+(t,\cdot)\Big) - v(t)a_*'(t).$$

Since $v(t)$ appears in all terms and L is a linear operator, we can remove $v(t)$ from the equation. Hence, we get the PDE

$$\partial_t V_*^+(t,x) = r(t)V_*^+(t,x) + \Big[V_*^+(t,x) - a_{*\dagger}(t)\Big]x - L\Big(V_*^+(t,\cdot)\Big) - a_*'(t).$$

Ultimately, we apply the differential operator L on V_*^+ resulting in the second order parabolic linear PDE

$$\partial_t V_*^+(t,x) = r(t)V_*^+(t,x) + \Big[V_*^+(t,x) - a_{*\dagger}(t)\Big]x - a_*'(t)$$

$$-b(t,x)\partial_x V_*^+(t,x) - \frac{1}{2}\sigma^2(t,x)\partial_x^2 V_*^+(t,x).$$

■

7.3 Numerical methods for solving Thiele's PDE

The goal of this section will be to explain how we can approximate the solution surface of Thiele's PDE along a grid of points. Recall that Thiele's PDE in the context of stochastic mortality is given by

$$\partial_t V_*^+(t,x) = r(t)V_*^+(t,x) + \Big[V_*^+(t,x) - a_{*\dagger}(t)\Big]x - a_*'(t)$$
$$-b(t,x)\partial_x V_*^+(t,x) - \frac{1}{2}\sigma^2(t,x)\partial_x^2 V_*^+(t,x),$$

with terminal condition

$$V_*^+(T,x) = a_*(T) - a_*(T-).$$

Let $\{(t_i,x_j)\}_{i=0,\dots,n}^{j=1,\dots,m} \subseteq [0,T]\times[0,\infty)$ be a partition of time and space where we wish to approximate $(t,x)\mapsto V_*^+(t,x)$. We will denote by Δt a fixed increment in time and similarly Δx for a fixed increment in space.

The information at our disposal is the terminal condition, which states a payoff at the end of the contract at time T. Although it is essential information, it is not enough to solve Thiele's PDE numerically without also imposing boundary conditions. These conditions are not given, but must be chosen with reasonable economical considerations. For instance, we can choose the mortality level $\mu_t = x_0 = 0$ as the lower boundary condition and for the upper boundary condition, a fixed maximum mortality level $\mu_t = x_m$.

We are now ready to explain the general idea behind the numerical schemes associated with Thiele's PDE.

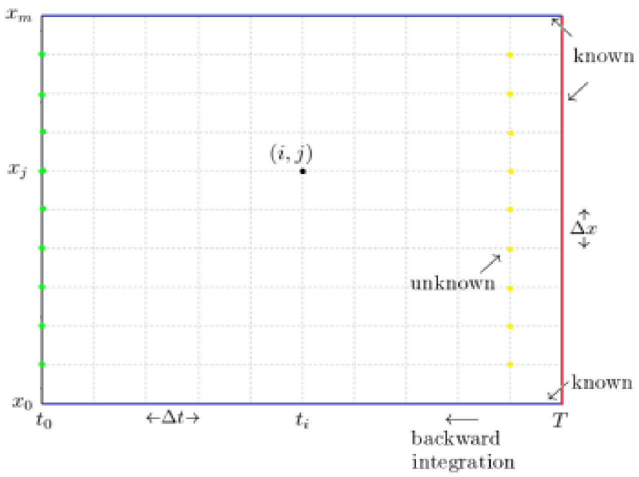

Figure 7.1: Grid of point where we approximate V_*^+. In blue, we have the boundary conditions. In red, we see the terminal condition. In green, we find the initial value of the contract for different mortality levels. In yellow, we initiate the first iteration of unknown points using the points from the red and blue line. Also note that the integration is backwards in time.

Example 7.3.1 (Boundary conditions in a pure endowment insurance) In the case of a pure endowment with payoff E at time $t = T$, the policy function is given by

$$a_*(t) = \begin{cases} 0, & t \in [0, T) \\ E, & t \in [T, \infty) \end{cases} \implies a_*'(t) = \begin{cases} 0, & t \in (0, T) \\ E, & else \end{cases}.$$

The contract's terminal condition is given by

$$V_*^+(T, x) = \begin{cases} 0, & x = x_m \\ E, & else \end{cases}.$$

The discontinuity at $x = x_m$ happens for a large enough mortality level due to the fact that the insured must be alive in order to claim the endowment.

Next, we get that the lower boundary condition for the mortality level $x = 0$ is given by the explicit pricing formula for the prospective reserve. Hence

$$V_*^+(t, 0) = \frac{1}{v(t)} \int_t^\infty v(s) p_{**}(x_0 + t, x_0 + s) da_*(s) = \frac{1}{v(t)} \int_t^\infty v(s) a_*'(s) ds$$

$$= \frac{1}{v(t)} \left[\int_t^\infty v(s) \cdot 0 \, ds + v(T)(E - 0) \right] = \frac{v(T)}{v(t)} E, \qquad t \in [0, T].$$

This should be intuitive as $V_*^+(t, 0)$ mimics the behavior of a zero coupon bond with face value E being paid at time $t = T$.

The upper boundary condition is simpler since

$$\begin{aligned} V_*^+(t, x_m) &= \frac{1}{v(t)} \int_t^\infty v(s) p_{**}(x_0 + t, x_0 + s) da_*(s) \\ &= \frac{1}{v(t)} \int_t^\infty v(s) \cdot 0 \, da_*(s) = 0, \qquad t \in [0, T]. \end{aligned}$$

Above we have used that the insured's survival probability is zero for a large enough mortality level.

7.3.1 Explicit method

As previously stated, we will not attempt to solve Thiele's PDE analytically. Instead we will focus on approximating V_*^+ on a grid of points. That is

$$V_*^+(t,x) \approx V_*^+(t_i,x_j), \qquad i=\{0,...,n\}, \qquad j=\{0,...,m\},$$

where $\{t_i\}_{i=0}^n \subseteq [0,T]$ and $\{x_j\}_{j=0}^m \subseteq [0,\infty)$ are partitions.

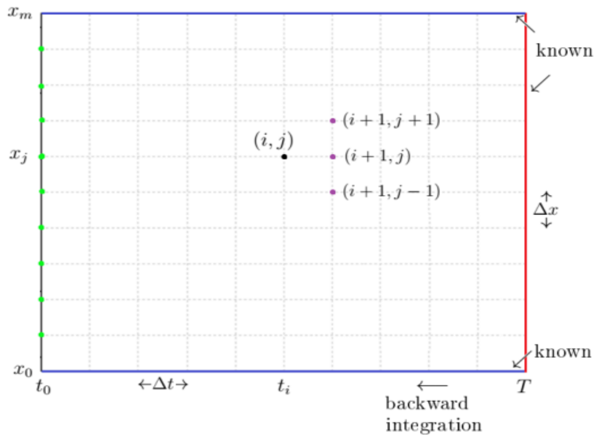

Figure 7.2: Explicit method to compute V_*^+. Given $V_*^+(t_{i+1},x_{j-1}), V_*^+(t_{i+1},x_j)$ and $V_*^+(t_{i+1},x_{j+1})$ we can estimate $V_*^+(t_i,x_j)$.

We will approximate the partial derivatives by the following incremental procedure

$$\partial_t V_*^+(t,x) \approx \frac{V_*^+(t+\Delta t,x) - V_*^+(t,x)}{\Delta t},$$

$$\partial_x V_*^+(t,x) \approx \frac{V_*^+(t,x+\Delta x) - V_*^+(t,x)}{\Delta x},$$

$$\partial_x^2 V_*^+(t,x) \approx \frac{V_*^+(t,x+\Delta x) - 2V_*^+(t,x) + V_*^+(t,x-\Delta x)}{(\Delta x)^2},$$

where $\Delta t, \Delta x > 0$ are suitably small.

Substituting our approximations in Thiele's PDE we get

$$\frac{V_*^+(t+\Delta t,x) - V_*^+(t,x)}{\Delta t} \approx r(t)V_*^+(t,x) + \Big[V_*^+(t,x) - a_{*\dagger}(t)\Big]x - a_*'(t)$$

$$-b(t,x)\frac{V_*^+(t,x+\Delta x) - V_*^+(t,x)}{\Delta x}$$

$$-\frac{1}{2}\sigma^2(t,x)\frac{V_*^+(t,x+\Delta x) - 2V_*^+(t,x) + V_*^+(t,x-\Delta x)}{(\Delta x)^2}.$$

Moreover, we evaluate V_*^+ at the points (t_{i+1}, x_j) on the right hand side of the equation and make use of the fact that

$$t_{i+1} = t_i + \Delta t, \qquad x_{j+1} = x_j + \Delta x.$$

This gives the numerical scheme

$$\frac{V_*^+(t_{i+1}, x_j) - V_*^+(t_i, x_j)}{\Delta t} \approx r(t_{i+1})V_*^+(t_{i+1}, x_j) + \Big[V_*^+(t_{i+1}, x_j) - a_{*\dagger}(t_{i+1})\Big]x_j$$

$$-a_*'(t_{i+1}) - b(t_{i+1}, x_j)\frac{V_*^+(t_{i+1}, x_{j+1}) - V_*^+(t_{i+1}, x_j)}{\Delta x}$$

$$-\frac{1}{2}\sigma^2(t_{i+1}, x_j)\frac{V_*^+(t_{i+1}, x_{j+1}) - 2V_*^+(t_{i+1}, x_j) + V_*^+(t_{i+1}, x_{j-1})}{(\Delta x)^2}.$$

$V_*^+(t_n, x_j)$ is known for all j, assuming that we have boundary conditions and the terminal condition. Hence, we can for a fixed timeline (t_{i+1}, x_j) compute $V_*^+(t_i, x_j)$ explicitly, as the name of the method suggests, for all i and j recursively. Isolating $V_*^+(t_i, x_j)$ gives

$$V_*^+(t_i, x_j) \approx V_*^+(t_{i+1}, x_j) - \Delta t\Bigg[r(t_{i+1})V_*^+(t_{i+1}, x_j) - a_{*\dagger}(t_{i+1})x_j$$

$$+V_*^+(t_{i+1}, x_j)x_j - a_*'(t_{i+1}) - b(t_{i+1}, x_j)\frac{V_*^+(t_{i+1}, x_{j+1}) - V_*^+(t_{i+1}, x_j)}{\Delta x}$$

$$-\frac{1}{2}\sigma^2(t_{i+1}, x_j)\frac{V_*^+(t_{i+1}, x_{j+1}) - 2V_*^+(t_{i+1}, x_j) + V_*^+(t_{i+1}, x_{j-1})}{(\Delta x)^2}\Bigg].$$

Thus the explicit method can simply be implemented as

$$V_*^+(t_i, x_j) \approx f\Big(V_*^+(t_{i+1}, x_{j-1}), V_*^+(t_{i+1}, x_j), V_*^+(t_{i+1}, x_{j+1})\Big).$$

Unfortunately, this method is not stable. That is to say that we may not get a numerical solution for arbitrary choices of Δt and Δx, but only for very small and specific choices.

In the specific case of solving the heat equation, it is shown in [Cra75] that the explicit method is both numerically stable and convergent whenever

$$\frac{\Delta t}{(\Delta x)^2} \leq \frac{1}{2}.$$

Moreover, the explicit method has a numerical error

$$\epsilon = O\Big(\Delta t\Big) + O\Big((\Delta x)^2\Big).$$

Since Thiele's PDE is in the same family as the heat equation, we could expect similar results depending on the complexity of the policy functions.

7.3.2 Implicit method

Since the explicit method is not stable for arbitrary choices of Δt and Δx, we will also overview the implicit method which is always stable. The main idea of the implicit method is to simultaneously solve a system of linear equations which consists of as many unknowns as spacial points in the grid.

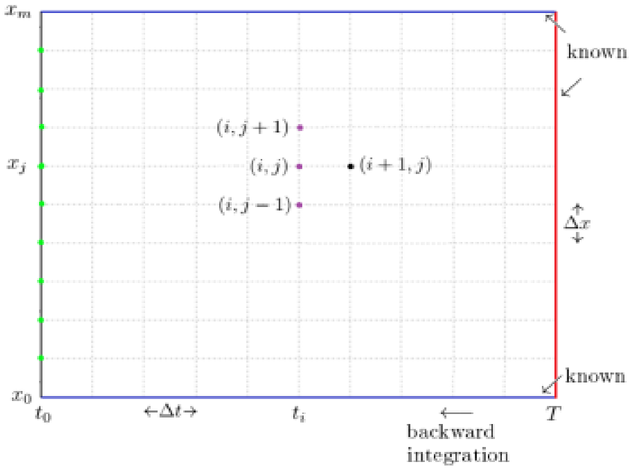

Figure 7.3: Implicit method to evaluate V_*^+. Given the point $V_*^+(t_{i+1}, x_j)$ we can solve a system of equations involving three unknown points $V_*^+(t_i, x_{j-1})$, $V_*^+(t_i, x_j)$ and $V_*^+(t_i, x_{j+1})$.

The implicit method is obtained by evaluating the right hand side of the approximation to the PDE at the points (t_i, x_j), instead of the points (t_{i+1}, x_j) which was the case with the explicit method. We get

$$\frac{V_*^+(t_{i+1}, x_j) - V_*^+(t_i, x_j)}{\Delta t} \approx r(t_i)V_*^+(t_i, x_j) + \Big[V_*^+(t_i, x_j) - a_{*\dagger}(t_i)\Big]x_j$$

$$-a_*'(t_i) - b(t_i, x_j)\frac{V_*^+(t_i, x_{j+1}) - V_*^+(t_i, x_j)}{\Delta x}$$

$$-\frac{1}{2}\sigma^2(t_i, x_j)\frac{V_*^+(t_i, x_{j+1}) - 2V_*^+(t_i, x_j) + V_*^+(t_i, x_{j-1})}{(\Delta x)^2}.$$

The next step is to group all terms with V_*^+ on the left hand side of the equation and put the remaining terms on the right hand side. After simplifying the notation a little bit, we arrive at an equation of the following form

$$A_i^j V_i^{j-1} + B_i^j V_i^j + C_i^j V_i^{j+1} = D_i^j,$$

where

$$V_i^j = V_*^+(t_i, x_j),$$

$$A_i^j = \frac{\Delta t}{2}\sigma^2(t_i, x_j),$$

$$B_i^j = -(\Delta x)^2 - (\Delta x)^2 \Delta t\, r(t_i) - (\Delta x)^2 \Delta t\, x_j - \Delta x \Delta t\, b(t_i, x_j) - \Delta t\, \sigma^2(t_i, x_j),$$

$$C_i^j = \Delta x \Delta t\, b(t_i, x_j) + \frac{\Delta t}{2}\sigma^2(t_i, x_j),$$

$$D_i^j = (\Delta x)^2 \Delta t \Big(- a_{*\dagger}(t_i) x_j - a'_*(t_i) \Big) - (\Delta x)^2 V_{i+1}^j,$$

It is important to see that the system of equations is only well-defined for $j = \{1, ..., m-1\}$ as for $j = 0$ and $j = m$ we have the nonsensical terms V_i^{-1} and V_i^{m+1} respectively. Therefore it would at a first glance appear that we have a system of linear equations consisting of $m-1$ equations and $m+1$ unknowns, $V_i^0, ..., V_i^m$.

Fortunately, that is not the case after a closer inspection. The two missing equations which stem from the boundary conditions also reduce the amount of unknowns by two. Indeed, the variables V_i^0 and V_i^m are known from the boundary conditions for all i. Writing our system of linear equations in matrix form, we get

$$M_i V_i + E_i = D_i, \qquad i = \{n, ..., 0\},$$

where

$$M_i = \begin{pmatrix} B_i^1 & C_i^1 & 0 & 0 & \cdots & 0 & 0 \\ A_i^2 & B_i^2 & C_i^2 & 0 & \cdots & 0 & 0 \\ 0 & A_i^3 & B_i^3 & C_i^3 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & C_i^{m-3} & 0 \\ 0 & 0 & 0 & 0 & \cdots & B_i^{m-2} & C_i^{m-2} \\ 0 & 0 & 0 & 0 & \cdots & A_i^{m-1} & B_i^{m-1} \end{pmatrix},$$

$$V_i = \begin{pmatrix} V_i^1 \\ V_i^2 \\ \vdots \\ V_i^{m-2} \\ V_i^{m-1} \end{pmatrix}, \qquad E_i = \begin{pmatrix} A_i^1 V_i^0 \\ 0 \\ \vdots \\ 0 \\ C_i^{m-1} V_i^m \end{pmatrix}, \qquad D_i = \begin{pmatrix} D_i^1 \\ D_i^2 \\ \vdots \\ D_i^{m-2} \\ D_i^{m-1} \end{pmatrix}.$$

Finally, by simple linear algebra we can compute the vector of unknowns V_i as

$$V_i = M_i^{-1}(D_i - E_i), \qquad i = \{n, ..., 0\}.$$

The implicit method is always numerically stable and convergent which is a big advantage. However, it is also more costly than the explicit method in terms of computational power. In the case of the heat equation, it is known that the numerical error of the implicit method is also of size

$$\epsilon = O\Big(\Delta t\Big) + O\Big((\Delta x)^2\Big).$$

7.4 Applications to insurance contracts

In this section, we will focus on calculating the mean reserve with stochastic mortality.

We will in the following examples assume that mortality follows a log-Ornstein-Uhlenbeck process of the form

$$\mu_t = \mu_0 e^{\alpha t + X_t}, \qquad t \in [0, T],$$

where X solves the SDE

$$dX_t = -\lambda X_t dt + \sigma dW_t, \qquad X_0 = 0, \qquad t \in [0, T],$$

and W is the standard Brownian motion. In order to apply our numerical methods, we need to derive the associated SDE for μ. The way forward is a simple application of Itô's formula using the function

$$f(t, x) = \mu_0 e^{\alpha t + x}.$$

Computing the partial derivatives gives

$$\partial_t f = \alpha f, \qquad \partial_x f = f, \qquad \partial_x^2 f = f.$$

Inserting in Itô's formula and evaluating at $(t, x) = (t, \mu_t)$ we have

$$d\mu_t = \alpha \mu_t dt + \mu_t[-\lambda X_t dt + \sigma dW_t] + \frac{1}{2} d[X, X]_t.$$

Calculating the quadratic variation and collecting terms yield

$$d\mu_t = \mu_t \Big[\Big(\alpha + \frac{1}{2}\sigma^2 - \lambda X_t \Big) dt + \sigma dW_t \Big].$$

Finally, we rewrite the expression by adding terms and using the formula of μ to obtain

$$d\mu_t = b(t, \mu_t) dt + \sigma(t, \mu_t) dW_t.$$

where the drift and volatility are given by

$$b(t, \mu_t) = \mu_t \Big(\alpha + \frac{1}{2}\sigma^2 + \lambda ln(\mu_0) + \alpha\lambda t - \lambda ln(\mu_t) \Big),$$

$$\sigma(t, \mu_t) = \sigma \mu_t.$$

We must of course remember to incorporate the information about the insured's age x_0, in the parameter μ_0 which appears in the drift.

Example 7.4.1 (Stochastic pension insurance) Assume for simplicity a constant risk-free interest rate such that

$$v(t) = e^{-\int_0^t r(u) du} = e^{-rt}.$$

For a pension insurance without premiums, the only policy function that contributes is given by

$$a_*(t) = \begin{cases} 0, & t \in [0, T_0) \\ P(t - T_0), & t \in [T_0, T) \\ P(T - T_0), & t \in [T, \infty) \end{cases} \implies a'_*(t) = \begin{cases} P, & t \in (T_0, T) \\ 0, & else \end{cases}.$$

The last order of business is to specify terminal and boundary conditions for the mean reserve. For our pension insurance, the terminal condition is

$$V_*^+(T, x) = 0, \qquad \forall x.$$

This should make sense as the contract expires and no more payments are made regardless of the insured's health.
In the case where the insured's health is frail and the mortality level is high, we get the upper boundary condition

$$V_*^+(t, x_m) = 0, \qquad \forall t.$$

Practically speaking this means that the insurer knows that the insured will die instantaneously after signing the pension policy. Hence all pension payouts happen with probability zero and there is no reason to set aside money. This assumption requires x_m to be large enough to mimic the limit

$$\lim_{x \to \infty} V_*^+(t, x) = 0$$

.
On the other extreme, if the insured is in shape and the mortality level is low, we get the lower boundary condition

$$\begin{aligned} V_*^+(t, 0) &= \frac{1}{v(t)} \int_t^\infty v(s) p_{**}(x_0 + t, x_0 + s) da_*(s) = \frac{1}{v(t)} \int_t^\infty v(s) a'_*(s) ds \\ &= \frac{P}{v(t)} \left[\delta_{[0,T_0)}(t) \int_{T_0}^T v(s) ds + \delta_{[T_0,T]}(t) \int_t^T v(s) ds \right] \\ &= \begin{cases} \frac{-P}{re^{-rt}} \left(e^{-rT} - e^{-rT_0} \right), & t \in [0, T_0) \\ \frac{-P}{re^{-rt}} \left(e^{-rT} - e^{-rt} \right), & t \in [T_0, T] \end{cases}. \end{aligned}$$

Hence the insured will obtain full pension with probability one.

Without loss of generality, we will use the data collected from [HMD] in order to illustrate the shape of the prospective reserve surface. We will consider a 30 year old male in 2019 with a pension policy paying $P = 100\ 000$ NOK starting from $T_0 = 40$ years with maturity date $T = 70$ years. Note that the data we have used does not aim to forecast future mortality based on past observations. That is our 71 data points consist of the mortality of a 30 year old male in 2019, mortality of a 31 year old male in 2019,..., mortality of a 100 year old male in 2019.

Recall that we are working with a mortality rate that follows a log-Ornstein-Uhlenbeck process. Thus we can once again use the estimation procedure developed in Section 4.2 to calibrate α, σ and λ. Since we are not dealing with generation mortality any more, but rather individual mortality of a given age as time passes by, we should expect that $\alpha > 0$.

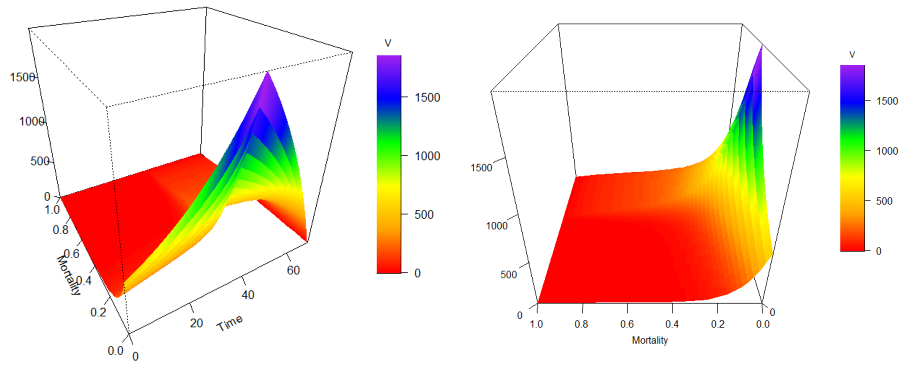

Figure 7.4: V_*^+ using the explicit method. Time units in years and monetary units in 1 000 NOK. $P = 100$, $r = 3\%$, $x_0 = 30$, $T_0 = 40$, $T = 70$, $\alpha = 0.0693$, $\lambda = 1.1129 \cdot 10^{-5}$, $\sigma = 0.03031$, $\mu_0 = 0.001837$, $\Delta t = 0.0002$ and $\Delta x = 0.02$.

In Figure 7.4 we observe that V_*^+ decreases with the mortality level. For a fixed mortality, we see the typical shape of a pension present value. First we have to put aside more and more money in order to meet the liabilities starting at time $t = T_0$ when we start paying out pensions. At time $t = T_0$ we obtain the maximum, since pension payments begin, and the value there is the expected cost for the remaining time of the contract. Then it decreases to 0 as time goes by since the maturity date approaches.

Lastly, the resolution of V_*^+ could have been better by choosing Δx and Δt smaller. However, the computational cost and runtime of such a program is considerably higher.

Example 7.4.2 (Deterministic pension insurance) We are interested in comparing the behaviour of the mean reserves when applying Thiele's PDE versus Thiele's ODE. The first case deals with stochastic mortality while the second is for deterministic mortality. In order to obtain a deterministic model, we simply apply the expected value. Using the same mortality as the previous example we get that

$$\tilde{\mu}_t = E\big[\mu_t\big] = \mu_0 e^{\alpha t} E\big[e^{X_t}\big] = \mu_0 exp\left(\alpha t + \frac{\sigma^2(1-e^{-\lambda t})}{4\lambda}\right), \qquad t \in [0, T],$$

where we have used that X_t is normally distributed.

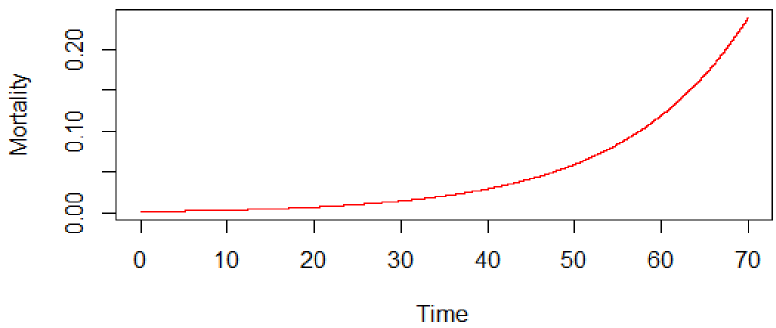

Figure 7.5: $\tilde{\mu}$ using the same estimates as in Figure 7.3.

Applying Thiele's ODE to our pension policy we get that

$$\frac{d}{dt} V_*^+(t) = rV_*^+(t) - a_*'(t) + \tilde{\mu}_t V_*^+(t), \qquad t \in [0, T]$$

with terminal condition

$$V_*^+(T) = 0.$$

Finally, we use Euler's method to iterate backwards in time for a partition $t_i \in [0, T]$.

$$\frac{V_*^+(t_i) - V_*^+(t_{i-1})}{\Delta t} = rV_*^+(t_i) - a_*'(t_i) + \tilde{\mu}_{t_i} V_*^+(t_i), \qquad i \in \{0, 1, ..., n\}.$$

Isolating $V_*^+(t_{i-1})$ we get the numerical scheme

$$V_*^+(t_{i-1}) = V_*^+(t_i) - \Delta t\Big[rV_*^+(t_i) - a_*'(t_i) + \tilde{\mu}_{t_i} V_*^+(t_i)\Big],$$

$$V_*^+(t_n) = V_*^+(T) = 0.$$

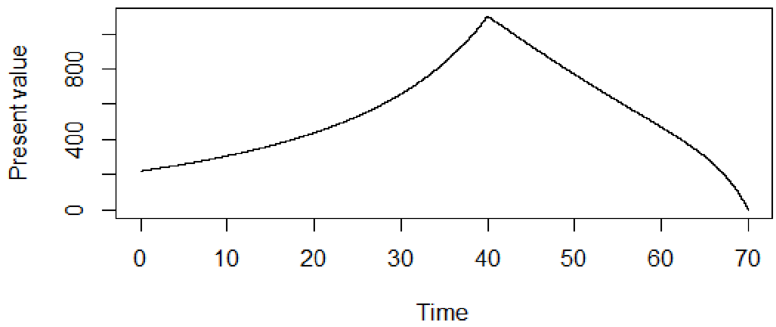

Figure 7.6: V_*^+ using Thiele's ODE and the same estimates as in Figure 7.3. Time units in years and monetary units in 1 000 NOK. $V_*^+(0, \tilde{\mu}_0) = 222.0283$.

We see in Figure 7.6 that $V_*^+(t, \tilde{\mu})$ is a smooth curve. Moreover, this method has the added benefits of being easy to implement and quick execution time.

On the other hand, we could also estimate the prospective reserve from Thiele's PDE. Indeed, the way forward would be to trace the curve $V_*^+(t, \tilde{\mu}_t)$ along the surface. However, this method has some notable problems. The first obstacle is the fact that the surface generated is only a finite approximation along a grid of points. For each partition point in time, it would be extremely unlikely that $\tilde{\mu}_{t_i}$ would hit one of the fixed partition points in mortality. The natural way of assigning meaningful values to the prospective reserve in these cases, would be to extrapolate by means of convex combinations. Mathematically, we calculate the following weighted average

$$V_*^+(t_i, \tilde{\mu}_{t_i}) \approx wV_*^+(t_i, x_{i+1}) + (1-w)V_*^+(t_i, x_i), \qquad \forall t_i \in [0, T],$$

where

$$w = \frac{\tilde{\mu}_{t_i} - x_i}{x_{i+1} - x_i}, \qquad \tilde{\mu}_{t_i} \in [x_i, x_{i+1}].$$

The second problem lies with truncation. In general $\mu : [0, T] \to [0, \infty]$, but in order to numerically compute a finite approximate to the solution of Thiele's PDE, we have to consider a bounded set for the codomain of mortality. In our case, we made the arbitrary choice of indexing the non-negative real numbers by the interval $[0, 1]$. We know that such a bijection is possible, for instance by a stereographic projection, since both sets have the same cardinality which is uncountably infinite.

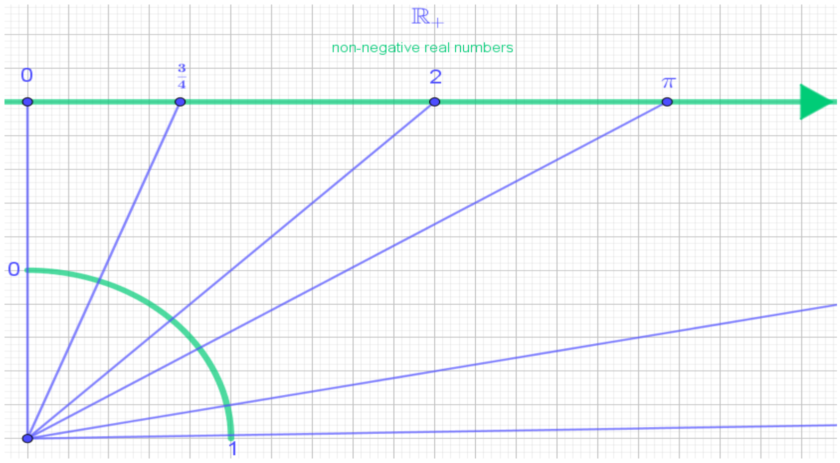

Figure 7.7: Stereographic projection of the interval $[0, 1]$ onto the non-negative real number line.

We note that the enumeration we have chosen has an inescapable downside of being highly sensitive as we approach one from bellow. We will however argue that this is not a problem as the high mortality rates will yield prospective reserves close to zero in the case of our pension example. On the other hand, we see from Figure 7.5 that the mortality rate $\tilde{\mu}_t$ takes values approximately in the interval $[0, 0.25]$ which have indexes near zero when the sensitivity is low. In fact, we will for simplicity work under the assumption that the sensitivity is so low close to zero that the inverse map of the stereographic projection is negligible.

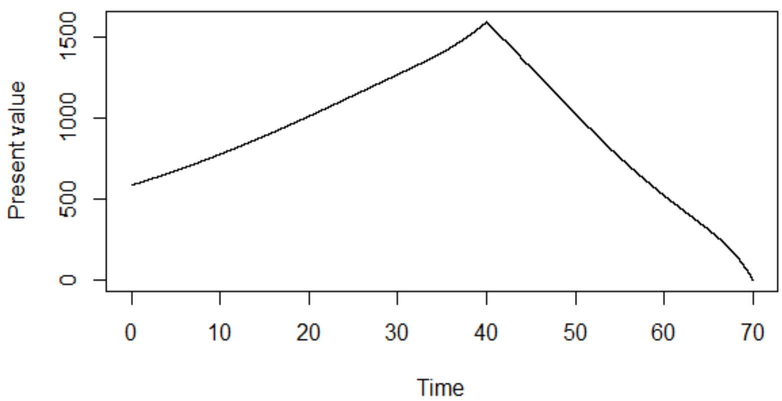

Figure 7.8: $V_*^+(t, \tilde{\mu}_t)$ traced on the surface we got in Figure 7.4. Time units in years and monetary units in 1 000 NOK. $V_*^+(0, \tilde{\mu}_0) = 585.4846$.

The results that we got clearly do not align with Figure 7.6. The most likely reason for this discrepancy is that the partition on the mortality axis is not refined enough. Indeed the mortality rates used in the estimation of parameters were of the scale of 10^{-6} while the partition we chose was of scale 10^{-2}. Unfortunately, the computational power and time it would take to run a program with such a refined partition is extremely high regardless of the numerical method we use to solve the PDE. Thus, for real life applications with deterministic mortality we conclude that Thiele's ODE should be preferred over Thiele's PDE.

Example 7.4.3 (Monte Carlo simulations of pension insurance) We will explore the possibility of simulating random mortality paths following a log Ornstein-Uhlenbeck process. To do so, we introduce a new probability space $(\Omega', \mathcal{A}', P')$ where $\mu_t(\omega')$ lives. This is necessary in order for us to treat the map $\omega' \mapsto \mu(\omega')$ deterministic for each fixed $\omega' \in \Omega'$. The main idea being that we can then for each $\mu_t(\omega')$ compute $V_*^+(t, \mu_t(\omega'))$ by means of Thiele's ODE.

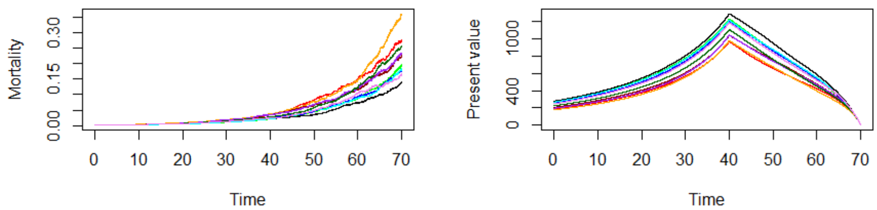

Figure 7.9: Using the same parameters as in Figure 7.3, we have simulated $m = 10$ paths of $\mu_t(\omega')$ with the associated prospective reserves $V_*^+(t, \mu_t(\omega'))$.

The small number of simulations were purely chosen for visibility purposes. For a larger simulation, we can expect more reliable data which we can use to assess risk in the pension contract.

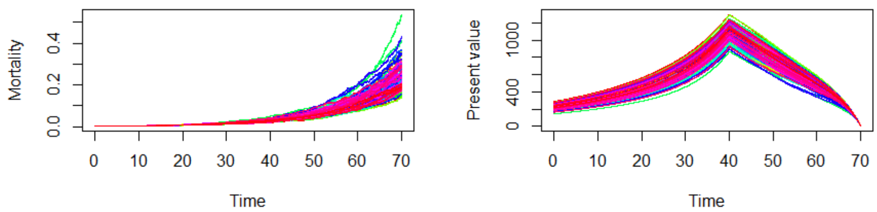

Figure 7.10: Using the same parameters as in Figure 7.3, we have simulated $m = 100$ paths of $\mu_t(\omega')$ with the associated prospective reserves $V_*^+(t, \mu_t(\omega'))$.

It is easy to spot the trend of increasing uncertainty in the mortality rate as time passes.

On the other hand, the uncertainty in the prospective reserve exhibits two distinct behaviours. On the interval $[0, T_0]$ we see an increasing trend of uncertainty due to the increasing risk of longevity which peaks at $t = T_0$. After making the first payment, the uncertainty diminishes quickly on the interval $(T_0, T]$. The reason being that we are approaching the maturity date of the contract where we know with 100% certainty that the prospective reserve is zero.

Finally, we plot the mean prospective reserve $E[V_*^+(t, \mu_t)]$.

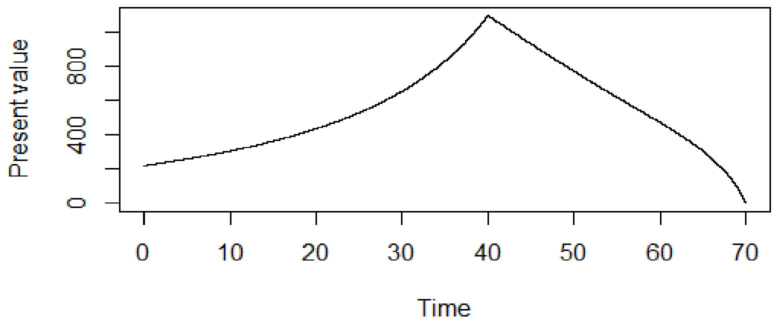

Figure 7.11: $E[V_*^+(t, \mu_t)]$ using Thiele's ODE and the same estimates as in Figure 7.3. Time units in years and monetary units in 1 000 NOK. $E[V_*^+(0, \mu_0)] = 221.409$.

Using the simulations we can approximate the 99% reserve of the policy by taking the empirical 99% percentile of the distribution $V_*^+(0, \mu_0)$. In our simulation, the 99% solvency capital turned out to be 278.3506 thousand NOK.

CHAPTER 8

Conclusions and further work

To conclude, we have provided a general framework for the computation of prospective reserves under deterministic and stochastic mortality. The fractional mortality model did not bear fruit in our research as Norwegian mortality did not seem to exhibit the self-similarity property, which is crucial for fractional noise models. In addition, the estimation of the Hurst parameter turned out to be a non trivial matter in the case of a fractional Ornstein-Uhlenbeck process where we must distinguish the underlying Hurst parameter from the perceived Hurst parameter.

On the other hand, we had great success deriving the theoretical result on Thiele's PDE with a general Markovian stochastic mortality. Two algorithms were constructed for solving Thiele's PDE and we implemented the explicit method in the case of a pension insurance with a geometric Ornstein-Uhlenbeck process. Since the implicit method proved to be too computational expensive and time consuming, we excluded its implementation.

Lastly, we ran Monte Carlo simulations in order to grasp the risk involved with reserving. The results we obtained are highly useful as the insurance company is obliged to maintain enough liquidity in order to meet its future liabilities under the Solvency II directive.

A natural extension of this theory would be to include financial risk. For instance, we could have considered unit-linked insurance policies where the payouts depend on the performance of a fund or some other financial derivative. Computing analytical prospective reserves, or at least simulating them, could be a model of high utility for the insurance sector.

Moreover, we have worked under the assumption that the company's return on investment is risk free and constant under the whole contract period. The assumption on risk-free interest rate is not unrealistic as the company can invest in the money market with risk-free interest rate or buy treasury bonds. However, the interest rate tends to be dynamic over larger time periods and stochastic models such as the Vasicek model could be incorporated to fit our setting.

A final though would be to expand the stochastic mortality from the class of Itô processes to the more general class of Lévy processes. The main benefit of this generalization would be that it allows us to consider jumps in the mortality rate which seems reasonable after a global pandemic.

APPENDIX A

Appendix

Code used in Example 3.2.4

```
# Import package
library(MASS)

# Initiate parameters
T = 1 # Final time
n = 1200 # Amount of time steps
lambda = 0 # Mean reversion rate
sigma = 1 # Volatility parameter
H = 0.5 # Hurst parameter

# Initiate vectors
time_vec = rep(0,n+1)
fOU_vec = rep(0,n+1)

# Fill time vector
for (i in 1:(n+1)){
  time_vec[i] = (i-1)*T/n
}

# Initiate covariance function
R_H = function(s,t){
  return(1/2*(abs(t)^(2*H)+abs(s)^(2*H)-abs(t-s)^(2*H)))
}

# Initiate covariance matrix
C = matrix(data = 0 ,nrow = n+1, ncol = n+1)

# Fill covariance matrix
for (i in 1:(n+1)){
  for (j in 1:(n+1)){
    C[i,j] = R_H(time_vec[i],time_vec[j])
  }
}

# Draw realisations of fBm
set.seed(100) # Fix omega to generate same path
fBm_vec = mvrnorm(n = 1, mu = rep(0,n+1), Sigma = C) # Remember to
      import MASS package


# Simulate paths of fOU-process using the Euler-Maruyama method
for (i in 1:n){
  fOU_vec[i+1] = fOU_vec[i]-lambda*fOU_vec[i]*1/n+sigma*(fBm_vec[
    i+1]-fBm_vec[i])
```

```
}

# Plot a path of the fOU-process for a given H
plot(time_vec, fOU_vec, type = "l", xlab = "Time", ylab = "Value",
     col = "blue", main = paste("H = ", toString(H)))
abline(h = 0)
```

Code used in Example 4.2.10

```
# Import MASS package
library(MASS)
library(fitdistrplus)

# Initiate parameters
T = 10 # Final time
n = 1000 # Amount of time steps
lambda = 0.4 # Mean reversion rate
sigma = 0.01 # Volatility parameter
H = 0.6 # Hurst parameter
m = 1000 # Simulations
fOU_start = 0 # Initial condition

# Initiate vectors
time_vec = rep(0,n+1)
fOU_vec = rep(0,n+1)
fOU_vec[1] = fOU_start

H_estimate_vec = rep(0,m)
sigma_estimate_vec = rep(0,m)
lambda_estimate_vec = rep(0,m)

# Fill time vector
for (i in 1:(n+1)){
  time_vec[i] = (i-1)*T/n
}

# Initiate covariance function
R_H = function(s,t){
  return(1/2*(abs(t)^(2*H)+abs(s)^(2*H)-abs(t-s)^(2*H)))
}

# Initiate covariance matrix
C = matrix(data = 0 ,nrow = n+1, ncol = n+1)

# Fill covariance matrix
for (i in 1:(n+1)){
  for (j in 1:(n+1)){
    C[i,j] = R_H(time_vec[i],time_vec[j])
  }
}

# Make m simulations
for (k in 1:m){

  # Draw realisations of fBm
  set.seed(k) # Fix omega to generate same path
  fBm_vec = mvrnorm(n = 1, mu = rep(0,n+1), Sigma = C)


  # Simulate paths of fOU-process using Euler's method
  for (i in 1:n){
```

```
    fOU_vec[i+1] =  fOU_vec[i]-lambda*fOU_vec[i]*1/n+sigma*(fBm_
    vec[i+1]-fBm_vec[i])
  }

  # Estimators
  H_hat = function(X) {
    X2n = X
    dX2n = X2n[-1] - X2n[-length(X2n)]
    Xn = X2n[-c(seq(2,length(X2n),2))]
    dXn = Xn[-1] - Xn[-length(Xn)]
    V2n = sum(dX2n^2)
    Vn = sum(dXn^2)
    return(1/2 - 1/(2*log(2))*log(V2n/Vn))
  }

  sigma_hat = function(X){
    V = 0

    for (i in 1:n){
      V = V + (X[i+1]-X[i])^2
    }

    return(sqrt(n^(2*H-1)*V/T))
  }

  lambda_hat = function(X){
    return((sum(X^2)/(n*sigma^2*H*gamma(2*H)))^(-1/(2*H)))
  }

  # Estimates
  H_estimate_vec[k] = H_hat(fOU_vec)
  sigma_estimate_vec[k] = sigma_hat(fOU_vec)
  lambda_estimate_vec[k] = lambda_hat(fOU_vec)
  print(k)
}

# Make plots
descdist(H_estimate_vec, discrete = FALSE)
plot(fitdist(H_estimate_vec, "norm"))

descdist(sigma_estimate_vec, discrete = FALSE)
plot(fitdist(sigma_estimate_vec, "norm", lower = c(0,0)))

descdist(lambda_estimate_vec, discrete = FALSE)
plot(fitdist(log(lambda_estimate_vec), "norm"))
```

Code used in Remark 5.1.2

```
# Import packages
library(readxl)
library(pracma)
library(MASS)

# Norwegian mortality data
dodlighetsdata_menn_og_kvinner = read_excel("dodlighetsdata_menn_
    og_kvinner.xlsx",
    col_types = c("text", "text", "numeric", "skip", "numeric"))

# We assume that mortality is strictly positive to avoid -Inf
    problems
```

```r
zero_elements = which(dodlighetsdata_menn_og_kvinner$Mortalitet %in% 0)

# Extrapolate data
for (i in zero_elements){
  blokk = ceiling(i/55)
  dodlighetsdata_menn_og_kvinner$Mortalitet[i] = mean(dodlighetsdata_menn_og_kvinner$Mortalitet[(55*(blokk-1)+1):(55*blokk)])
}

# Prepare data collection
time_vec = 0:54 # Year 1966-2020
age_vec = 0:106
n = length(time_vec)
m = length(age_vec)

alpha_estimate_matrix = matrix(0, ncol = 2, nrow = m)
sigma_estimate_matrix = matrix(0, ncol = 2, nrow = m)
lambda_estimate_matrix = matrix(0, ncol = 2, nrow = m)
mu_0_estimate_matrix = matrix(0, ncol = 2, nrow = m)

# Estimators
H_hat = function(fOU_vec){
  return(0.5) # Fixed value
}

sigma_hat = function(fOU_vec){
  V = 0

  for (i in 1:(n-1)){
    V = V + (fOU_vec[i+1]-fOU_vec[i])^2
  }

  return(sqrt(n^(2*H_hat(fOU_vec)-1)*V/n))
}

lambda_hat = function(fOU_vec){
  return((sum(fOU_vec^2)/(n*sigma_hat(fOU_vec)^2*H_hat(fOU_vec)*gamma(2*H_hat(fOU_vec))))^(-1/(2*H_hat(fOU_vec))))
}

alpha_hat = function(mortality_vec){
  sum_1 = sum((log(mortality_vec)-log(mortality_vec[1]))*time_vec)
  sum_2 = sum(time_vec^2)
  return(sum_1/sum_2)
}

# Estimate alpha
for (i in 1:(2*m)){
  dod_vec = dodlighetsdata_menn_og_kvinner$Mortalitet[(55*(i-1)+1):(55*i)] # Group data by year

  if (i <= m){
    alpha_estimate_matrix[i,1] = alpha_hat(dod_vec) # Men
    fOU_data_vec = log(dod_vec/dod_vec[1])-alpha_estimate_matrix[i,1]*time_vec # Transform data
    sigma_estimate_matrix[i,1] = sigma_hat(fOU_data_vec)
    lambda_estimate_matrix[i,1] = lambda_hat(fOU_data_vec)
    mu_0_estimate_matrix[i,1] = dod_vec[1]
  }

```

```r
  if (i > m){
    alpha_estimate_matrix[i-m,2] = alpha_hat(dod_vec) # Women
    fOU_data_vec = log(dod_vec/dod_vec[1])-alpha_estimate_matrix[i-m,2]*time_vec # Transform data
    sigma_estimate_matrix[i-m,2] = sigma_hat(fOU_data_vec)
    lambda_estimate_matrix[i-m,2] = lambda_hat(fOU_data_vec)
    mu_0_estimate_matrix[i-m,2] = dod_vec[1]
  }
}

# Plot parameters against age for men and women together
plot(age_vec, alpha_estimate_matrix[,1], type = "l", col = "blue", ylim = c(-0.06,0.15) , xlab = "Age", ylab = expression(hat(alpha)[0]))
lines(age_vec, alpha_estimate_matrix[,2], type = "l", col = "red")
abline(h = 0)

plot(age_vec, sigma_estimate_matrix[,1], type = "l", col = "blue", ylim = c(0,2.5) , xlab = "Age", ylab = expression(hat(sigma)))
lines(age_vec, sigma_estimate_matrix[,2], type = "l", col = "red")
abline(h = 0)

plot(age_vec, lambda_estimate_matrix[,1], type = "l", col = "blue", ylim = c(0, 2.5) , xlab = "Age", ylab = expression(hat(lambda)))
lines(age_vec, lambda_estimate_matrix[,2], type = "l", col = "red")
abline(h = 0)



# Simulations
age_index = 15 # 0,15,30,50,70,90
m = 10000 # Number of simulations
n = 55 # Number of years
gender = 0 # {male,female} = {0,1}

# Prepare data collection
fOU_simulation = rep(0,n)
fOU_matrix = matrix(0, nrow = n, ncol = m)
mean_mortality_vec = rep(0,n)

# Initiate covariance function
R_H = function(s,t){
  return(1/2*(abs(t)^(2*H_hat(1))+abs(s)^(2*H_hat(1))-abs(t-s)^(2*H_hat(1))))
}

# Initiate covariance matrix
C = matrix(data = 0 ,nrow = n, ncol = n)

# Fill covariance matrix
for (i in 1:n){
  for (j in 1:n){
    C[i,j] = R_H(time_vec[i],time_vec[j])
  }
}

# Initiate simulations
for (k in 1:m){

```

```
    # Draw realisations of fBm
    set.seed(k) # Fix omega to generate same path
    fBm_vec = mvrnorm(n = 1, mu = rep(0,n), Sigma = C) # Requires
      the MASS package


    # Simulate paths of fOU-process using Euler's method
    # Adjust for age and gender in estimate_vec
    for (i in 1:(n-1)){
      fOU_simulation[i+1] = fOU_simulation[i]-lambda_estimate_
      matrix[age_index+1,gender+1]*fOU_simulation[i]*1/n+sigma_
      estimate_matrix[age_index+1,gender+1]*(fBm_vec[i+1]-fBm_vec[i
      ])
    }

    fOU_matrix[,k] = fOU_simulation
}

row_sorted_fOU_matrix = t(apply(fOU_matrix, 1, sort))

# Extract 95% confidence interval
lower_percentile = row_sorted_fOU_matrix[,2.5/100*m]
upper_percentile = row_sorted_fOU_matrix[,97.5/100*m]

# Make plots
for (t in 1:n){
  mean_mortality_vec[t] = mu_0_estimate_matrix[age_index+1,gender
    +1]*exp(alpha_estimate_matrix[age_index+1,gender+1]*(t-1)+
    rowMeans(fOU_matrix)[t])
  lower_percentile[t] = mu_0_estimate_matrix[age_index+1,gender+1]
    *exp(alpha_estimate_matrix[age_index+1,gender+1]*(t-1)+lower_
    percentile[t])
  upper_percentile[t] = mu_0_estimate_matrix[age_index+1,gender+1]
    *exp(alpha_estimate_matrix[age_index+1,gender+1]*(t-1)+upper_
    percentile[t])
}

plot(0:(n-1), mean_mortality_vec, type = "l", xlab = "Time", ylab
     = "Mortality", ylim = c(0,max(upper_percentile)), col = "red",
      main = paste("Age ", toString(age_index)) ) # Simulated mean
      mortality
lines(0:(n-1), dodlighetsdata_menn_og_kvinner$Mortalitet[(55*age_
      index+1+5885*gender):(55*age_index+55+5885*gender)], type = "l
      ", col = "blue") # Historical mortality
lines(0:(n-1), lower_percentile, type = "l", col = "black") # 2.5
      % percentile
lines(0:(n-1), upper_percentile, type = "l", col = "black") # 97.5
       % percentile
```

Code used in Example 6.1.13

```
# Import package
library(mosaic)

# Norwegian mortality year 2020
# Initiate data
age_vec = 0:106

# Sum mortality of men and women
# Mortality data SSB 2020
dead_vec =
```

```
c(170,16,14,3,5,3,8,11,2,3,6,6,8,17,12,17,25,37,38,26,39,40,
25,35,30,52,55,36,38,49,36,48,47,52,54,51,58,58,81,64,71,
100,95,102,77,92,122,141,160,175,163,171,215,213,244,267,
327,335,335,384,450,450,536,606,615,681,814,903,930,1020,
1122,1259,1398,1530,1687,1780,1870,2105,2315,2466,2871,
2849,3349,3524,3872,4157,3933,4270,4503,4737,4493,4206,
4120,3759,3402,2866,2401,2072,1519,1114,859,476,344,
234,135,82,152)

alive_vec =
c(100000,99830,99813,99799,99796,99791,99788,99780,99768,99767,
99764,99758,99752,99744,99727,99715,99697,99672,
99636,99598,99572,99533,99492,99467,99433,99403,99351,
99296,99259,99221,99172,99137,99089,99042,98990,
98936,98885,98827,98769,98688,98624,98553,98453,98358,
98256,98179,98087,97965,97825,97665,97490,97327,97157,
96941,96728,96484,96218,95891,95556,95221,94837,94387,
93937,93401,92795,92180,91500,90686,89783,88853,87833,
86711,85452,84054,82524,80837,79057,77188,75082,72767,
70301,67430,64581,61232,57708,53836,49679,45745,41475,
36972,32235,27743,23537,19417,15658,12256,9390,6989,
4917,3398,2284,1424,948,604,370,235,152)

mortality_vec = dead_vec/alive_vec

# Examine plot
plot(age_vec, mortality_vec, xlab = "Age", ylab = "Mortality rate", title("Norway 2020"))
grid(lty = 1)

# Create dataframe
mortality_df = data.frame(age_vec, mortality_vec)

# Use mosaic package to fit the curve to our data
f = fitModel(mortality_vec ~ a+b*exp(c*age_vec), data = mortality_df, start=list(a=0.00127529, b=2.51137*10^-6,c=0.1271853))

print(coef(f))

# Estimated curve
mu = function(t){
  return(coef(f)[1]+coef(f)[2]*exp(coef(f)[3]*t))
}

# Plot curve with datapoints
curve(mu, from = min(age_vec), to = max(age_vec), col = "red", add = TRUE)


# Estimated survival probability
survival_probability = function(t,s){
  return(exp(-coef(f)[1]*(s-t)-coef(f)[2]/coef(f)[3]*(exp(coef(f)[3]*s)-exp(coef(f)[3]*t))))
}

# Insurance data
E = 100000 # Endowment benefit
B = 200000 # Death benefit
T = 40 # Maturity date
x = 30 # Insured's age
r = 0.02 # Interest rate

```

```
# Discount factor
v = function(t){
  return(exp(-r*t))
}

# Integrands
integrand_benefit = function(t,s){
  return(v(s)*survival_probability(t+x,s+x)*mu(s+x))
}

integrand_premium = function(t,s){
  return(v(s)*survival_probability(t+x,s+x))
}

# Numeric integration
simpsons_rule = function(f, t, a, b) {
  if (is.function(f) == FALSE) {
    stop('f must be a function with one parameter (variable)')
  }

  n = 120000 # Higher n gives better accuracy
  h = (b - a) / n

  xj = seq.int(a, b, length.out = n + 1)
  xj = xj[-1]
  xj = xj[-length(xj)]
  approx = (h / 3) * (f(t,a) + 2 * sum(f(t,xj[seq.int(2, length(xj), 2)])) + 4 * sum(f(t,xj[seq.int(1, length(xj), 2)])) + f(t,b))

  return(approx)
}

# Calculate yearly premium
yearly_premium = (E*v(T)*survival_probability(x,T+x)+B*simpsons_rule(integrand_benefit,0,0,T))/(simpsons_rule(integrand_premium,0,0,T))

# Make plot
PV_benefit = rep(0,T+1)
PV_premium = rep(0,T+1)
mean_reserve = rep(0,T+1)

for (i in 0:T){
  PV_benefit[i+1] = 1/v(i)*(B*simpsons_rule(integrand_benefit,i,i,T)+E*v(T)*survival_probability(i+x,T+x))
  PV_premium[i+1] = -yearly_premium/v(i)*simpsons_rule(integrand_premium,i,i,T)
}

mean_reserve = PV_benefit + PV_premium

plot(0:T,PV_benefit, xlab = "Contract time", ylab = "Present value in NOK", col = "blue", ylim = c(min(PV_premium),E))
lines(0:T, PV_premium, type = "p", col = "red")
lines(0:T, mean_reserve, type = "p", col = "black")
grid(lty = 1)
```

Code used in Example 7.4.1

```
# Import package
```

```r
library(readxl)

# Read mortality data of a 30 year old male in 2019
dod_30 = read_excel("Jobb/STK/dod_30.xlsx",
                    col_types = c("numeric"))

# Initiate parameters
n = length(dod_30$data)
time_vec = 0:(n-1)

# Define estimators
H_hat = function(Z_vec){
  return(1/2)
}

alpha_hat = function(mortality_vec){
  sum_1 = sum((log(mortality_vec)-log(mortality_vec[1]))*time_vec)
  sum_2 = sum(time_vec^2)
  return(sum_1/sum_2)
}

sigma_hat = function(Z_vec){
  V = 0

  for (i in 1:(n-1)){
    V = V + (Z_vec[i+1]-Z_vec[i])^2
  }

  return(sqrt(n^(2*H_hat(Z_vec)-1)*V/n))
}

lambda_hat = function(Z_vec){
  return((sum(Z_vec^2)/(n*sigma_hat(Z_vec)^2*H_hat(Z_vec)*gamma(2*
    H_hat(Z_vec))))^(-1/(2*H_hat(Z_vec))))
}

# Transform data
alpha = alpha_hat(dod_30$data)

OU_vec = rep(0,n)

for (i in 1:n){
  OU_vec[i] = log(dod_30$data[i]/dod_30$data[1])-alpha*time_vec[i]
}

# Initiate parameter estimation
mu_0 = dod_30$data[1]
sigma = sigma_hat(OU_vec)
lambda = lambda_hat(OU_vec)
```

Code used in Example 7.4.1 and Example 7.4.2

```r
# Solving Thiele's PDE for prospective reserves

# Define geometric OU parameters
alpha = 0.0692813492
lambda = 1.112907144*10^(-5)
sig = 0.0303133478
mu_0 = 0.001837 # Insured's mortality at age 30 at contract start

# Define contract information for pension
```

```r
P = 100 # Pension
T_0 = 40 # Retirement date
T = 70 # Maturity date
r = 0.03 # Constant interest rate

# For stability we need that dt/dx^2 <= 0.5
# Initiate space and time vectors
dt = 0.00005
time_vec = seq(0,T,dt)

dx = 0.01
space_vec = seq(0,1,dx)

# Define policy function
a_prime = function(t){
  if (time_vec[t] >= 0 & time_vec[t] < T_0){
    return(0)
  }
  if (time_vec[t] >= T_0 & time_vec[t] < T){
    return(P)
  }
  if (time_vec[t] >= T){
    return(0)
  }
}

# Define drift and volatility
b = function(t,x){
  return(space_vec[x]*(alpha + 0.5*sig^2 + lambda*log(mu_0) +
    alpha*lambda*time_vec[t] - lambda*log(space_vec[x])))
}

sigma = function(t,x){
  return(sig*space_vec[x])
}

# Reserve matrix
V = matrix(NA, nrow = length(time_vec) , ncol = length(space_vec))

# Impose lower boundary condition
lower_boundary = function(t){
  if (time_vec[t] >= 0 & time_vec[t] < T_0){
    return(-P/(r*exp(-r*time_vec[t]))*(exp(-r*T)-exp(-r*T_0)))
  }

  if (time_vec[t] >= T_0 & time_vec[t] <= T){
    return(-P/(r*exp(-r*time_vec[t]))*(exp(-r*T)-exp(-r*time_vec[t
    ])))
  }
}

# Fill in lower boundary condition
for (i in 1:length(time_vec)){
  V[i,1] = lower_boundary(i)
}

# Fill in upper boundary condition
for (i in 1:length(time_vec)){
  V[i,length(space_vec)] = 0
}

# Fill in terminal condition
```

```r
for (j in 1:length(space_vec)){
  V[length(time_vec),j] = 0
}

# Explcit method
m = length(space_vec)
n = length(time_vec)

for (i in (n-1):1){

  for (j in 2:(m-1)){
    partial_1 = (V[i+1,j+1]- V[i+1,j])/dx
    partial_2 = (V[i+1,j+1]- 2*V[i+1,j] + V[i+1,j-1])/dx^2

    V[i,j] = V[i+1,j] - dt*(r*V[i+1,j]+V[i+1,j]*space_vec[j]-a_prime(i+1)-b(i+1,j)*partial_1-0.5*sigma(i+1,j)*partial_2)
  }
  print(i)
}

# Plot surface with color
library(fields)

par(bg = "white")
nrz = nrow(V)
ncz = ncol(V)

# Create a function interpolating colors in the range of specified colors
jet.colors = colorRampPalette( c("red", "orange", "yellow", "green", "blue", "purple") )

# Generate the desired number of colors from this palette
nbcol = 100
color = jet.colors(nbcol)

# Compute the z-value at the facet centres
zfacet = (V[-1, -1] + V[-1, -ncz] + V[-nrz, -1] + V[-nrz, -ncz])/4

# Recode facet z-values into color indices
facetcol = cut(zfacet, nbcol)
persp(time_vec, space_vec, V, col = color[facetcol], phi = 30, theta = -30, axes=T, ticktype = "detailed", border = NA, xlab = "Time", ylab = "Mortality", zlab = " ")
## add color bar
image.plot(legend.only = TRUE, zlim=range(zfacet), col=color, legend.args = list( text = "V",cex = .8,side = 3,line = .5))

# Expected mortality
mu = function(t){
  return(mu_0*exp(alpha*t+(sig^2-sig^2*exp(-lambda*t))/(4*lambda)))
}

mu_vec = rep(0,length(time_vec))

for (i in 1:(length(time_vec))){
  mu_vec[i] = mu(time_vec[i])
}

plot(time_vec,mu_vec,type = "l")
```

```r

# Project expected mortality under surface and extrapolate V by
    convex combinations
V_vec = rep(0,length(time_vec))


for (i in 1:length(time_vec)){

  if (0 <= mu_vec[i] & mu_vec[i] < 0.02){
    w = (mu_vec[i]-0)/dx
    V_vec[i] = w*V[i,2]+(1-w)*V[i,1]
  }

  if (0.02 <= mu_vec[i] & mu_vec[i] < 0.04){
    w = (mu_vec[i]-0.02)/dx
    V_vec[i] = w*V[i,3]+(1-w)*V[i,2]
  }

  if (0.04 <= mu_vec[i] & mu_vec[i] < 0.06){
    w = (mu_vec[i]-0.04)/dx
    V_vec[i] = w*V[i,4]+(1-w)*V[i,3]
  }

  if (0.06 <= mu_vec[i] & mu_vec[i] < 0.08){
    w = (mu_vec[i]-0.06)/dx
    V_vec[i] = w*V[i,5]+(1-w)*V[i,4]
  }

  if (0.08 <= mu_vec[i] & mu_vec[i] < 0.10){
    w = (mu_vec[i]-0.08)/dx
    V_vec[i] = w*V[i,6]+(1-w)*V[i,5]
  }

  if (0.10 <= mu_vec[i] & mu_vec[i] < 0.12){
    w = (mu_vec[i]-0.10)/dx
    V_vec[i] = w*V[i,7]+(1-w)*V[i,6]
  }

  if (0.12 <= mu_vec[i] & mu_vec[i] < 0.14){
    w = (mu_vec[i]-0.12)/dx
    V_vec[i] = w*V[i,8]+(1-w)*V[i,7]
  }

  if (0.14 <= mu_vec[i] & mu_vec[i] < 0.16){
    w = (mu_vec[i]-0.14)/dx
    V_vec[i] = w*V[i,9]+(1-w)*V[i,8]
  }

  if (0.16 <= mu_vec[i] & mu_vec[i] < 0.18){
    w = (mu_vec[i]-0.16)/dx
    V_vec[i] = w*V[i,10]+(1-w)*V[i,9]
  }

  if (0.18 <= mu_vec[i] & mu_vec[i] < 0.20){
    w = (mu_vec[i]-0.18)/dx
    V_vec[i] = w*V[i,11]+(1-w)*V[i,10]
  }

  if (0.20 <= mu_vec[i] & mu_vec[i] < 0.22){
    w = (mu_vec[i]-0.20)/dx
    V_vec[i] = w*V[i,12]+(1-w)*V[i,11]
  }
```

```

  if (0.22 <= mu_vec[i] & mu_vec[i] < 0.24){
    w = (mu_vec[i]-0.22)/dx
    V_vec[i] = w*V[i,13]+(1-w)*V[i,12]
  }

}

plot(time_vec,V_vec, type = "l", xlab = "Time", ylab = "Present
    value")
```

Code used in Example 7.4.3

```
# Monte Carlo simulations of prospective reserve
# Define parameters
alpha = 0.0692813492
lambda = 1.112907144*10^(-5)
sigma = 0.0303133478

P = 100 # Pension payout
T_0 = 40 # First pension payout
T = 70 # Maturity date
r = 0.03 # Risk free interest rate
mu_0 = 0.001837 # Mortality of 30 year old male

# Define policy function
a_prime = function(t){
  if (t >= 0 & t < T_0){
    return(0)
  }
  if (t >= T_0 & t < T){
    return(P)
  }
  if (t >= T){
    return(0)
  }
}

# Simulate Ornstein Uhlenbeck process
m = 100 # Number of simulations
dt = 0.0002
time_vec = seq(0,T,dt)
n = length(time_vec)

bm_matrix = matrix(0, ncol = m, nrow = n)
ou_matrix = matrix(0, ncol = m, nrow = n)
mu_matrix = matrix(0, ncol = m, nrow = n)
V_matrix = matrix(0, ncol = m, nrow = n)

for (j in 1:m){
  ou_vec = rep(0,n)
  mu_vec = rep(0,n)

  ou_vec[1] = log(mu_0)

  # Generate Brownian motion starting at 0
  bm_vec = cumsum(c(0,rnorm(n-1, mean = 0, sd = sqrt(dt))))

  # Generate Orstein Uhlenbeck process starting in log(mu_0)
  for (i in 1:(n-1)){
```

```
    ou_vec[i+1] = ou_vec[i] - lambda*ou_vec[i]*dt + sigma*(bm_vec[i+1]-bm_vec[i])
  }

  # Generate log Ornstein Uhlenbeck process starting at mu_0
  for (i in 1:n){
    mu_vec[i] = exp(alpha*time_vec[i]+ou_vec[i])
  }

  # Solve Thieles ODE for each realisation
  V = rep(0, n)

  for (i in (n-1):1){
    V[i] = V[i+1] - dt*(r*V[i+1]-a_prime(time_vec[i+1])+mu_vec[i+1]*V[i+1])
  }

  # Store data
  bm_matrix[,j] = bm_vec
  ou_matrix[,j] = ou_vec
  mu_matrix[,j] = mu_vec
  V_matrix[,j] = V
}

# Find mean reserve
expected_V = rowMeans(V_matrix)

# Make plot of all paths and mean reserve
cols = rainbow(m)

matplot(time_vec, bm_matrix, type = "l", col = cols, lty = 1, xlab = "Time", ylab = "Brownian motion")
matplot(time_vec, ou_matrix, type = "l", col = cols, lty = 1, xlab = "Time", ylab = "Ornstein Uhlenbeck")
matplot(time_vec, mu_matrix, type = "l", col = cols, lty = 1, xlab = "Time", ylab = "Mortality")
matplot(time_vec, V_matrix, type = "l", col = cols, lty = 1, xlab = "Time", ylab = "Present value")

plot(time_vec, expected_V, type = "l", col = "black", xlab = "Time", ylab = "Present value")

# Calculate expected premium and 99 % reserve
expected_premium = expected_V[1] # 221.409
reserve_99 = sort(V_matrix[1,])[99] # 278.3506
```

Bibliography

[Bal17] Baldi, P. *Stochastic Calculus An Introduction Through Theory and Exercises*. Springer, 2017.

[CKM03] Cheridito, P., Kawaguchi, H. and Maejima, M. 'FRACTIONAL ORNSTEIN-UHLENBECK PROCESSES'. In: *Electronic Journal of Probability* vol. 8, no. 3 (2003), p. 14.

[Cra75] Crank, J. *THE MATHEMATICS OF DIFFUSION*. OXFORD UNIVERSITY PRESS, 1975.

[DO19] Delgado-Vences, F. and Ornelas, A. 'MODELLING ITALIAN MORTALITY RATES WITH A GEOMETRIC-TYPE FRACTIONAL ORNSTEIN-UHLENBECK PROCESS.' In: (2019), p. 19.

[Gar18] Garcin, M. 'HURST EXPONENTS AND DELAMPERTIZED FRACTIONAL BROWNIAN MOTIONS'. In: *International Journal of Theoretical and Applied Finance* vol. 22 (2018), p. 26.

[GG11] Gavrilov, L. A. and Gavrilova, N. S. 'Mortality Measurement at Advanced Ages: A Study of the Social Security Administration Death Master File'. In: *North American Actuarial Journal* vol. 15 (2011), pp. 432–447.

[Gom25] Gompertz, B. 'On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies'. In: *Philosophical Transactions of the Royal Society of London* vol. 115 (1825), pp. 513–585.

[HMD] HMD. *The Human Mortality Database*. URL: https://www.mortality.org/. (accessed: 12.01.2022).

[HN10] Hu, Y. and Nualart, D. 'Parameter estimation for fractional Ornstein-Uhlenbeck processes'. In: *Statistics Probability Letters* vol. 80 (2010), pp. 1030–1038.

[Kol12] Koller, M. *Stochastic Models in Life Insurance*. Springer, 2012.

[Lin17] Lindstrøm, T. L. *Spaces : an introduction to real analysis*. American Mathematical Society, 2017.

[Mel11] Melichov, D. 'ON ESTIMATION OF THE HURST INDEX OF SOLUTIONS OF STOCHASTIC DIFFERENTIAL EQUATIONS'. In: (2011), p. 78.

[MP01] Milevsky, M. A. and Promislow, S. D. 'Mortality Derivatives and the Option to Annuitize'. In: *Insurance: Mathematics and Economics* vol. 29 (2001), pp. 299–318.

[Nua95] Nualart, D. *The Malliavin Calculus and Related Topics*. Springer, 1995.

[PT00] Pipiras, V. and Taqqu, M. S. 'Integration questions related to fractional Brownian motion'. In: *Probability Theory and Related Fields* (2000), pp. 251–291.

[SSBa] SSB. *Aldersavhengige dødsfallsrater for menn og kvinner (Tabell 5)*. URL: https://www.ssb.no/befolkning/fodte-og-dode/statistikk/dode. (accessed: 08.10.2021).

[SSBb] SSB. *Dødelighetstabeller, etter kjønn og alder 1966 - 2020*. URL: https://www.ssb.no/statbank/table/07902/. (accessed: 09.11.2021).

[Wal12] Walsh, J. *Knowing the Odds An Introduction to Probability*. American Mathematical Society, 2012.

[Yer+14] Yerlikaya-Özkurt, F. et al. 'Estimation of the Hurst parameter for fractional Brownian motion using the CMARS method'. In: *Journal of Computational and Applied Mathematics* vol. 259 (2014), pp. 843–850.

[ZCY12] Zeng, C., Chen, Y. Q. and Yang, Q. 'THE FBM-DRIVEN ORNSTEIN-UHLENBECK PROCESS:PROBABILITY DENSITY FUNCTION ANDANOMALOUS DIFFUSION'. In: *Fractional Calculus and Applied Analysis* vol. 15, no. 3 (2012), pp. 479–492.

[Øks03] Øksendal, B. *Stochastic Differential Equations An Introduction with Applications*. Springer, 2003.