Simulations of a New Model for Stochastic Transition Rates in Life Insurance Based on Generalized Cox Processes
Knut G. M. Aasen
Master’s Thesis, Spring 2016
Abstract

In this thesis we introduce a new model, formulated as a nonlinear filtering problem with a generalized Cox process, for the modeling of transition rates in life insurance. This allows for a time-continuous model which may capture both Gaussian and non-Gaussian noise effects. In addition, for the non-Gaussian noise we introduce a jump component with intensity which is subject to random effects. These elements make for a very flexible model with great opportunities for capturing many different effects such as e.g. regime switching effects or mean reversion. Since we are able to incorporate such elements into our model, we can capture effects from regulatory changes in the insurance market, changes in government activity, and different kinds of "shocks" which affect transition rates, such as natural disasters. In addition, a completely specified model with an in depth discussion on simulations will be given. With our new model we simulate future mortality rates. Where in the simulation stage, the focus has been laid on capturing mean reversion through the intensity of the jump component in the generalized Cox process.

Keywords: Lévy processes, life insurance, nonlinear filtering, stochastic filtering theory, stochastic transition rates.
Acknowledgement

First and foremost, I would like to thank my supervisor Frank Proske for his enthusiasm and wisdom. I have truly appreciated your guidance, Frank.

I would like to thank my family for endless support, and my friends for positive distractions during the thesis work. Also, I would like to thank my good friend Krister Karlsen for proofreading the thesis.
Notation

In this thesis we shall make use of the notation that follows. Some of the notation will be introduced as we go along, but it is good to have a reference to go back to if confusion occurs.

Spaces

- $\mathbb{N} := \{1, 2, 3, \ldots \}$ — The set of all natural numbers.
- $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$ — The set of all natural numbers, with 0 included.
- $\mathbb{R}^d$ — The set of all $d$-dimensional column-vectors with real entries.
- $\mathbb{R}^{d \times m}$ — All $d \times m$ matrices with real entries.
- $\mathbb{R}_{\geq 0} := [0, \infty)$ — The set of all non-negative real numbers.
- $\mathbb{R}^d_{\geq 0} := \mathbb{R}^d \setminus \{0\}$ — All non-negative vectors in $\mathbb{R}^d$.
- $\mathbb{R}^d_0 := \mathbb{R}^d \setminus \{0\}$ — $\mathbb{R}^d$ without the 0-vector in $\mathbb{R}^d$.
- $C(U; H)$ — The set of all continuous functions $f : U \to H$.
- $C_b(U; H)$ — The set of all continuous functions $f : U \to H$, that are bounded.
- $C(U \times V; H)$ — The set of all continuous functions $f : U \times V \to H$.
- $C^{i,j}(U \times V; H)$ — The set of all continuous functions $U \times V \ni (u, v) \mapsto f(u, v)$, that maps into $H$, that are $i$ and $j$ times differentiable in $u$ and $v$, respectively.

Norms

- $\|u\|$, where $u$ is a vector — The vector norm; that is,
  $$\|u\| := \sqrt{u_1^2 + \cdots + u_d^2}, \quad u \in \mathbb{R}^d.$$  

  The norm extends to matrices by putting
  $$\|\sigma\| := \left\{ \sum_{i=1}^d \sum_{j=1}^m (\sigma_{ij})^2 \right\}^{\frac{1}{2}}, \quad \sigma \in \mathbb{R}^{d \times d}.$$
• \( \|u\|_1 \), where \( u \) is a vector — The manhattan norm; that is,
\[
\|u\|_1 := \sum_{j=1}^{d} |u_j|, \quad u \in \mathbb{R}^d.
\]

• \( \|\cdot\|_{\infty,\mathbb{R}^d} \) — the supremum norm; that is, for \( f : \mathbb{R}^d \to \mathbb{R}^m \),
\[
\|f\|_{\infty,\mathbb{R}^d} := \max_{1 \leq j \leq m} \sup_{x \in \mathbb{R}^d} \|f_j(x)\|.
\]

Other notation

• \( \langle x, y \rangle \), where \( x = (x_1, \ldots, x_d)' \in \mathbb{R}^d \) and \( y = (y_1, \ldots, y_d)' \in \mathbb{R}^d \) are vectors — The inner product on \( \mathbb{R}^d \); that is,
\[
\langle x, y \rangle = \sum_{j=1}^{d} x_j y_j.
\]

• \( \langle \mu, f \rangle \), where \( \mu \) is a measure and \( f \) is a function — The integral of \( f \) with respect to \( \mu \).

• \( |\Pi| \), where \( \Pi = (0 = t_0 < t_1 < \cdots < t_n = T) \) is a partitioning of \([0, T]\) — The supremum of the widths of the subintervals; that is,
\[
|\Pi| = \sup_{0 \leq j \leq n-1} |t_{j+1} - t_j|.
\]

• \( \mathcal{B}(\mathbb{R}^d) \) — The Borel \( \sigma \)-algebra on \( \mathbb{R}^d \).

• \( \mathcal{B}^d_0 := \{ U \in \mathcal{B}(\mathbb{R}^d_0) : 0 \notin \overline{U} \} \), where \( \overline{U} \) denotes the closure of \( U \).

• \( ^t \) — Transposition; that is, for a matrix \( \sigma \), \( \sigma' \) is the transpose of \( \sigma \).

• \( \log \{ \cdot \} \) — The natural logarithm.

• \( 1_U \) — The indicator function of the event \( U \); that is, if \( U \) occurs then \( 1_U = 1 \), otherwise \( 1_U = 0 \).

• \( 1_U(x) \) — The indicator function; that is, if \( x \in U \) then \( 1_U(x) = 1 \), otherwise \( 1_U(x) = 0 \).

• \( \mathcal{N} \) — The family of all null sets.

• \( X \overset{d}{=} Y \) — The stochastic variables \( X \) and \( Y \) are equal in distribution.

• \( X \sim \Theta \) — The stochastic variable \( X \) is \( \Theta \)-distributed.

• \( \mu \ll \lambda \) — The probability measure \( \mu \) is absolutely continuous with respect to the probability measure \( \lambda \).

• \( \mu \sim \lambda \) — \( \mu \) and \( \lambda \) are equivalent probability measures.

• ❋ — Marks the end of an example, or the end of a proof inside an example.

• ■ — Marks the end of a proof.

• \( \mathbb{E}^{\mu} [\cdot] \) — The expectation under the probability measure \( \mu \).

• \( f \wedge g := \min \{ f, g \} \).

• \( [x], x \in \mathbb{R} \) — The integer part of \( x \).
Abbreviations and acronyms

- a.s. — Almost surely, with probablity 1.
- i.i.d. — Independent and identically distributed.
- SDE — Stochastic differential equation.
- SLLN — (Kolmogorov’s) strong law of large numbers.
# Contents

1 Introduction 1

2 Preliminaries — probability theory and stochastic analysis 3
   2.1 Measure theory .................................................. 3
   2.2 Probability theory ................................................. 4
      2.2.1 Useful tools ................................................. 6
   2.3 Important processes and their properties ...................... 7
   2.4 Martingale theory ................................................. 10
   2.5 Stochastic analysis ............................................... 13
   2.6 Monte Carlo simulation .......................................... 15

3 Lévy processes 17
   3.1 Lévy processes and their properties ........................... 17
      3.1.1 Jumps and random measures ................................ 19
      3.1.2 Jumps of Lévy processes ................................... 20
      3.1.3 Poisson integration ......................................... 23
      3.1.4 Decomposition of Lévy processes ............................ 23
      3.1.5 Distributional properties of Lévy processes ............... 25
      3.1.6 Lévy processes as martingales .............................. 28

4 Nonlinear stochastic filtering 29
   4.1 An introduction to stochastic filtering ......................... 29
   4.2 Filtering problems ............................................... 30
   4.3 Formulation of the filtering problem ........................... 32
   4.4 The optimal filter ................................................. 35
      4.4.1 Change of measure .......................................... 35
      4.4.2 The unnormalized filter and the optimal filter .......... 38
   4.5 Numerics ............................................................ 38

5 Life insurance 41
   5.1 Risk in life and pension insurance ............................. 42
      5.1.1 Mortality risk .............................................. 42
      5.1.2 Longevity risk .............................................. 43
      5.1.3 Overestimation versus underestimation of risk ............ 44
      5.1.4 Costs ....................................................... 44
   5.2 Markov Chains .................................................... 45
   5.3 Transition rates in life insurance .............................. 48
Chapter 1

Introduction

When performing risk management in relation to life insurance contracts, it is of great importance to capture future uncertainty when modeling transition rates. A great concern when modeling such transition rates, is that it might take half a decade before one sees the full ramification of wrongful estimation. By that time an insurance company might have issued large amounts of contracts, that have been priced based on wrongfully estimated transition rates. Hence the use of deterministic models, such as e.g. the Gompertz-Makeham model, could have catastrophic consequences for anyone making long term guarantees, as such models are significantly limited in the sense of capturing future uncertainty.

As of today, there have been introduced many new models, since the model suggested by Gompertz and Makeham, for the purpose of modeling transition rates by means of stochastic models. For such models, the reader may e.g. consult Lee, Carter [14] or Wang et al. [23], in the case of mortality rates. However, some of the more popular models only allow for Gaussian noise, and do not offer flexibility which would allow for including elements such as e.g. regime switching effects or mean reversion into the model. Though some of the models include perhaps a couple of such elements, they can often not be modified to include any effect the user might want to capture. Our model offers such flexibility as we consider a nonlinear filtering problem with a generalized Cox process for the observation process. In addition, it is up to the user to choose a structure of the actual transition rates. Hence the model may in theory be used to model any sought after transition rate. However, as our model has a lot of flexibility, and thereby many functions and parameters to be either decided on or estimated, one should have a lot of data available to implement the model.

Objective of the thesis

The objective of this thesis has been the following:

- Construct a new model for the modeling of stochastic transition rates in life insurance and simulate future mortality rates.

- Build a solid framework with an introduction to Lévy processes and nonlinear filtering theory.
Outline of the thesis

It has been a goal to structure this thesis to be self-contained. Therefore, for the most part, all the necessary tools needed to be able to read and understand the contents of this thesis, will be developed as we go along. However, some knowledge of probability theory and stochastic analysis will be advantageous.

The thesis has been structured as follows: Chapter 2 is dedicated to giving a foundation in probability theory, stochastic analysis and related theory. Chapter 3 contains an introduction to Lévy processes, Poisson integration and other important concepts to be used later on. Chapter 4 introduces filtering theory, with a focus on filtering problems where we have jump-behaviour in the observation process. Also, we introduce the Monte Carlo techniques that are to be used to estimate the parameters of our model, in Chapter 4. Chapter 5 serves as a motivation into why one is interested in modeling transition rates when considering contracts in life insurance. Also, in Chapter 5, concepts in life insurance mathematics are introduced with some accompanying examples and a discussion on risks to be considered when dealing with life insurance contracts. In Chapter 6, some popular models for the modeling of stochastic transition rates are briefly discussed before we introduce a new model with a detailed simulation procedure. The last chapter is Chapter 7, where we discusses possible extensions and future work.

All probability distributions used in the thesis are given in Appendix A. A mixture of different computations and definitions used in the thesis are given in Appendix B. In addition, all computer programs used to e.g. produce plots or estimate parameters are given in Appendix C.

The programming languages used in this thesis are R [19] and MATLAB [15], and will henceforth not be referenced. All the data that has been used in simulation procedures, have been extracted from the human mortality database [9] and will be referenced as we go along.
Chapter 2

Preliminaries — probability theory and stochastic analysis

Bibliographical notes

The main references of this chapter are: Jacod, Protter [10], Cont, Tankov [22], and Øksendal [26].

In this chapter we will give an introduction to selected parts of probability theory and stochastic analysis. It is beneficial for the reader to be familiar with some probability theory, but we will use this chapter to introduce the concepts necessary for this thesis. Some of the most basic properties and tools in probability theory and stochastic analysis will be covered, and these will be used throughout the thesis. This is done to make the thesis more self contained. To begin with, some basic measure theoretical definitions will be introduced, which can be good to have on hand before we get to probability theory and stochastic analysis.

For the reader who is comfortable with the concepts of probability and stochastic analysis, or is biased towards application, this chapter can be skipped.

2.1 Measure theory

In measure theory we often consider a triplet $(\Omega, \mathcal{H}, \nu)$, where $\Omega$ is a non-empty set, $\mathcal{H}$ is what we call a $\sigma$–algebra on $\Omega$, and $\nu$ is what we call a measure on $(\Omega, \mathcal{H})$. The definitions of these concepts are given below.

**Definition 2.1** ($\sigma$–algebra). Let $\Omega$ be a non-empty set. Let $\mathcal{H}$ be a family of subsets of $\Omega$. We say that $\mathcal{H}$ is a $\sigma$–algebra on $\Omega$ if the following properties hold:

1. $\emptyset \in \mathcal{H}$.
2. If $A \in \mathcal{H}$, then $A^{c} \in \mathcal{H}$.
3. If $(A_{i})_{i \in \mathbb{N}} \subset \mathcal{H}$ is a sequence of pairwise disjoint sets, then $\bigcup_{i \in \mathbb{N}} A_{i} \in \mathcal{H}$.
When we say that \((A_i)_{i \in \mathbb{N}}\) is a sequence of pairwise disjoint sets, we mean that 
\(A_i \cap A_j = \emptyset\), for all \(i \neq j, i,j \in \mathbb{N}\). The pair \((\Omega, \mathcal{H})\), where \(\Omega\) is a non-empty set and 
\(\mathcal{H}\) is a \(\sigma\)-algebra on \(\Omega\), is called a measurable space.

**Definition 2.2** (Measure). Let \((\Omega, \mathcal{H})\) be a measurable space. Let \(\nu : \mathcal{H} \to [0, \infty]\). We say that \(\nu\) is a measure on \((\Omega, \mathcal{H})\) if the following properties are satisfied:

(i) \(\nu(\emptyset) = 0\).

(ii) For a sequence \((A_i)_{i \in \mathbb{N}} \subset \mathcal{H}\) of pairwise disjoint sets, 
\[ \nu \left( \bigcup_{i \in \mathbb{N}} A_i \right) = \sum_{i \in \mathbb{N}} \nu(A_i). \]

Let \((\Omega, \mathcal{H})\) be a measurable space and \(\nu\) a measure on it. Then \((\Omega, \mathcal{H}, \nu)\) is called a measure space. The sets \(A \in \mathcal{H}\) are called measurable, and \(\nu(A)\) is the measure of \(A\).

**Definition 2.3** (Null set). Let \((\Omega, \mathcal{H}, \nu)\) be a measure space. Let \(A \subset \Omega\). If there exists a \(B \in \mathcal{H}\) such that \(A \subset B\) and \(\nu(B) = 0\), we say that \(A\) is a null (or negligible) set, or more specifically a \(\nu\)-null set.

The family of all null sets will usually be denoted by \(\mathcal{N}\).

Let \((\Omega, \mathcal{H})\) be a measurable space. Let \((\mathcal{S}, \mathcal{S})\) be another measurable space. We say that a function \(f : \Omega \to \mathcal{S}\) is \(\mathcal{H}\)-measurable if 
\[ f^{-1}(A) = \{ \omega \in \Omega : f(\omega) \in A \} \in \mathcal{H}, \quad A \in \mathcal{S}. \]

**Definition 2.4** (Absolute continuity of measure). Let \((\Omega, \mathcal{H})\) be a measurable space. Let \(\nu, \pi : \mathcal{H} \to [0, \infty)\) be two measures on it. We say that \(\nu\) is absolutely continuous with respect to \(\pi\) if \(\pi(A) = 0\) implies that \(\nu(A) = 0\), \(A \in \mathcal{H}\). If \(\nu\) is absolutely continuous with respect to \(\pi\) we write \(\nu \ll \pi\).

Let \((\Omega, \mathcal{H})\) be a measurable space. Let \(\nu, \pi\) be two finite measures on \((\Omega, \mathcal{H})\). If \(\nu \ll \pi\) and \(\pi \ll \nu\) we say that \(\pi\) and \(\nu\) are equivalent, which we will denote by \(\pi \sim \nu\).

### 2.2 Probability theory

In probability theory and stochastic analysis we encounter experiments whose outcomes are subject to chance. Hence we can not be completely sure of the outcome of the experiment, before we have actually conducted the experiment.

Throughout this thesis we will be considering a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\); that is:

- \(\Omega\) is the sample space; that is, the set of all outcomes of some experiment.

- \(\mathcal{F}\) is a \(\sigma\)-algebra on \(\Omega\), which contains all events that might occur. We assume \(\mathcal{F}\) to be complete; that is, we assume that \(\mathcal{F}\) contains all \(\mathbb{P}\)-null sets, or more mathematically we require that 
\[ \{ A \subset \Omega : \exists B \in \mathcal{F}, A \subset B, \mathbb{P}[B] = 0 \} \subset \mathcal{F}. \]
• $\mathbb{P}$ is a probability measure; that is, a measure on $(\Omega, \mathcal{F})$ such that $\mathbb{P}[\Omega] = 1$, $\mathbb{P}[\emptyset] = 0$ and, for all $A \in \mathcal{F}$, $0 \leq \mathbb{P}[A] \leq 1$.

We say that a property holds $\mathbb{P}$-almost surely, or $\mathbb{P}$-a.s. for short, if it holds for all events that are outside the family $\mathcal{N}$ of all $\mathbb{P}$-null sets. Note that this is equivalent to saying that a property holds $\mathbb{P}$-a.s., if there exists a $\Omega_0 \in \mathcal{F}$ such that $\mathbb{P}[\Omega_0] = 1$, and the property holds for all $\omega \in \Omega_0$.

In practice, one usually thinks of the family $\mathcal{N}$ as the collection of all "impossible" events.

**Definition 2.5 (Stochastic variable).** Let $(\mathcal{S}, \mathcal{S})$ be a measurable space. A mapping $X : \Omega \rightarrow \mathcal{S}$ is called a stochastic variable on $(\Omega, \mathcal{F}, \mathbb{P})$, with values in $\mathcal{S}$, if it is $\mathcal{F}$-measurable; that is,

$$\forall A \in \mathcal{S}, \quad X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\} \in \mathcal{F}.$$

The following gives an example of a stochastic variable.

**Example 2.1.** Define the function $X : \Omega \rightarrow \mathbb{R}$ to be given by, for all $\omega \in \Omega$, $X(\omega) := A$, where $A \in \mathbb{R}$ is a constant.

**Claim:** $X$ is a stochastic variable on $(\Omega, \mathcal{F}, \mathbb{P})$.

**Proof.** We have, for all $B \in \mathcal{B}(\mathbb{R})$,

$$X^{-1}(B) = \begin{cases} \Omega, & \text{if } A \in B, \\ \emptyset, & \text{if } A \notin B. \end{cases}$$

Since $\emptyset$ and $\Omega$ is in $\mathcal{F}$, we have shown that any constant is a stochastic variable. ♣

Let $X$ be a stochastic variable on $(\Omega, \mathcal{F}, \mathbb{P})$, which takes values in $\mathcal{S}$; that is $X : \Omega \rightarrow \mathcal{S}$. Then $\mathcal{S}$ is what we call the state space of $X$. Typically the state space $\mathcal{S}$ will either be $\mathbb{N}_0$ or $\mathbb{R}^d$.

Formally we define the distribution of a stochastic variable $X$ with values in $\mathcal{S}$, as the probability measure

$$\mathbb{P}^X[A] = \mathbb{P}\{\omega \in \Omega : X(\omega) \in A\} = \mathbb{P}[X \in A], \quad A \in \mathcal{S}.$$ 

Also, we will sometimes speak of almost sure (a.s. for short) convergence, defined below, which should not be confused with convergence in probability*.

**Definition 2.6 (Almost sure convergence).** Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of stochastic variables on $(\Omega, \mathcal{F}, \mathbb{P})$. We say that $(X_n)_{n \in \mathbb{N}}$ converges almost surely to a stochastic variable $Y$ on $(\Omega, \mathcal{F}, \mathbb{P})$ if

$$\mathbb{P}\left[\lim_{n \rightarrow \infty} X_n = Y\right] = 1.$$  

(2.1)

For a sequence $(X_n)_{n \in \mathbb{N}}$ of stochastic variables which converges almost surely to a stochastic variable $Y$, we will often write

$$X_n \xrightarrow[n \rightarrow \infty]{} Y, \quad \text{a.s.}$$  

(2.2)

*See e.g. Definition 2.8 in Cont, Tankov [22], for the definition of convergence in probability.
2.2. PROBABILITY THEORY

2.2.1 Useful tools

When considering a problem in mathematics in general, it can be very useful to perform some kind of transformation to study the problem under different circumstances. An example of this in probability theory is the characteristic function.

Definition 2.7 (Characteristic Function). Let $X$ be a stochastic variable with values in $\mathbb{R}^d$. Then the characteristic function $\varphi_X : \mathbb{R}^d \to \mathbb{C}$ of $X$, is defined by

$$\varphi_X(u) = \mathbb{E}[\exp\{i\langle u, X \rangle\}], \quad u \in \mathbb{R}^d. \quad (2.3)$$

An important property with regard to the characteristic function, is that there is a 1–1 relation between the characteristic function of a stochastic variable and its distribution. In other words; if $X$ and $Y$ are two stochastic variables with the same characteristic function, then $X \overset{d}{=} Y$. The characteristic function always exists (see e.g. argument on page 103 in Jacod, Protter [10]), which makes it a very popular tool for analyzing distributional properties of stochastic variables.

In application of probability theory, it is very useful to be able to change the probability measure under which we examine a problem. For such a change of measure the theorem that follows is central.

The following theorem is borrowed from Jacod, Protter [10] (Theorem 28.3).

Theorem 2.1 (Radon-Nikodym Theorem). Let $Q$ be a finite measure on $(\Omega, \mathcal{F})$, such that that $Q \ll P$. Then there exists a unique integrable positive stochastic variable $\Lambda$ such that

$$Q[A] = \mathbb{E}[1_A \Lambda], \quad A \in \mathcal{F}.$$

We will often write

$$\Lambda = \frac{dQ}{dP},$$

and refer to this as the Radon-Nikodym derivative.

Next we state Bayes’ theorem for conditional expectation. Later on in Chapter 4, when we tackle what is called nonlinear filtering problems, we will rely heavily on a formula which is a direct consequence of the Bayes’ theorem.

Theorem 2.2 (Bayes’ Theorem). Let $X$ be an integrable stochastic variable on $(\Omega, \mathcal{F}, \mathbb{P})$. Let $\mathcal{H}$ be a sub-$\sigma$-algebra of $\mathcal{F}$. Let $Q$ be a probability measure on $(\Omega, \mathcal{F})$ such that $Q \gg \mathbb{P}$ and

$$\Lambda = \frac{d\mathbb{P}}{dQ}.$$

Then

$$\mathbb{E}[X|\mathcal{H}] \mathbb{E}^Q[\Lambda|\mathcal{H}] = \mathbb{E}^Q[X\Lambda|\mathcal{H}], \quad a.s.$$  

Proof. See e.g. Lemma 8.6.2 in Øksendal [26] or Theorem 3.22 in Xiong [25].
In practice when we perform a change of measure, we might sometimes want to do so with a restriction. Let $\mathcal{H}$ be a sub-$\sigma$-algebra of $\mathcal{F}$. Then we could define a restricted version of the probability measure $\mathbb{P}$, restricted to $\mathcal{H}$, as $\tilde{\mathbb{P}} := \mathbb{P}|_{\mathcal{H}}$. Now $\tilde{\mathbb{P}}$ is defined by

$$\tilde{\mathbb{P}}[A] = \mathbb{P}[A], \quad A \in \mathcal{H},$$

and is a probability measure on $(\Omega, \mathcal{H})$.

**Definition 2.8 (Stochastic process).** A stochastic process $(X_t)_{t \in \mathbb{T}}$ is a family of stochastic variables parametrized by $t \in \mathbb{T}$, where $\mathbb{T} \neq \emptyset$ is called the parameter space.

Let $X = (X_t)_{t \in \mathbb{T}}$ be a stochastic process. We have the following:

- For a fixed $t \in \mathbb{T}$, $\Omega \ni \omega \mapsto X_t(\omega)$ is a stochastic variable.
- For a fixed $\omega \in \Omega$, the mapping $\mathbb{T} \ni t \mapsto X_t(\omega)$ is called the path of $X$.

Here the parameter space $\mathbb{T}$ could for example be $\mathbb{N}_0$, $\mathbb{R}_{\geq 0}$ or some interval $(a, b]$, $b > a > 0$. However, from now on we will only consider parameter spaces that are finite time intervals, for example $[0, T]$, where $T > 0$ is a finite time horizon. One example of the time horizon $T$, in the setting of finance, could be the time when a given call option reaches maturity.

Also, when considering stochastic processes in this thesis we will always define them on the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$, unless we specifically state otherwise.

### 2.3 Important processes and their properties

Two properties that many stochastic processes considered in literature today possess, are: the property of independent increments and the property of stationary increments.

Let $X = (X_t)_{0 \leq t \leq T}$ be a stochastic process.

- We say that $X$ has independent increments if, for all $t, h > 0$, the increment $X_{t+h} - X_t$ is independent of the process $(X_s)_{0 \leq s \leq t}$.
- We say that $X$ has stationary increments if, for all $t, h > 0$, the increment $X_{t+h} - X_t$ is equal to $X_h$ in distribution.

A very important process, which is used in a wide range of fields, and which also will be used in all problems encountered in this thesis, is the Brownian motion. The Brownian motion is typically used as the noisy part of a model, when we are trying to model a phenomenon that we can not be certain of how evolves over time. An example of such a phenomenon could be the behavior of a stock price.

**Definition 2.9 (Brownian Motion).** A stochastic process $B = (B_t)_{0 \leq t \leq T}$ with values in $\mathbb{R}$, is called a (1-dimensional) Brownian motion, if the following properties hold:

1. $B_0 = 0$, $\mathbb{P}$-a.s.
(ii) $B$ has independent increments.

(iii) $B$ has stationary increments.

(iv) $B$ has Gaussian increments, that is, $\forall t, h > 0$, $(B_{t+h} - B_t) \sim N(0, h)$.

A $d$-dimensional Brownian motion is simply a process $(B_t)_{0 \leq t \leq T}$, taking values in $\mathbb{R}^d$, given by $B_t = (B^{(1)}_t, B^{(2)}_t, \ldots, B^{(d)}_t)'$, $0 \leq t \leq T$, where the components $(B^{(i)}_t)_{0 \leq t \leq T}$, $i = 1, \ldots, d$, are mutually independent (1-dimensional) Brownian motions. Also, if we define a Brownian motion without specifying its dimension it will always be 1-dimensional.

Remark 1. It is important to note that even though the Brownian motion is continuous there exists a $\Omega_0 \in \mathcal{F}$ such that $\mathbb{P}[\Omega_0] = 1$, and for all $\omega \in \Omega_0$, $t \mapsto B_t(\omega)$ is nowhere differentiable. Hence we can not speak of $\frac{d}{ds} B(s)$.

The Brownian motion has also the advantage that it has paths that are quite easy to simulate. To perform a simulation procedure for the paths of the Brownian motion, we do the following. Let $B = (B_t)_{0 \leq t \leq T} \subset \mathbb{R}$ be a Brownian motion. Let $\Pi = (0 = t_0 < t_1 < \cdots < t_n = T)$ be an equidistant partitioning of $[0, T]$, with $|\Pi| = \Delta t$; that is, $t_j := j \Delta t$, $j = 0, \ldots, n$, where $\Delta t := T/n$. Now let $\xi$ be a standard Gaussian stochastic variable. By the Gaussian increments of the Brownian motion, we have

$$B_{t_{j+1}} - B_{t_j} \overset{d}{=} \xi \sqrt{\Delta t}, \quad j = 0, \ldots, n - 1,$$

such that

$$B_{t_{j+1}} = B_{t_j} + (B_{t_{j+1}} - B_{t_j}) \overset{d}{=} B_{t_j} + \xi \sqrt{\Delta t}, \quad j = 0, \ldots, n - 1.$$

We now arrive at Algorithm 2.1, which we will utilize on many occasions when we handle problems numerically in this thesis.

---

**Algorithm 2.1 Path of Brownian motion**

**Input:** Time horizon $T$; $n$ for partitioning.

1. $\Delta t \leftarrow T/n$
2. generate $\xi_j \sim N(0, 1)$, $j = 0, \ldots, n - 1$
3. $B_0 \leftarrow 0$
4. for $j = 0, \ldots, n - 1$ do
   5. $B_{t_{j+1}} \leftarrow B_{t_j} + \xi_j \sqrt{\Delta t}$
5. end for
7. return $(B_{t_{j}})_{j=0}^n$

---

The Brownian motion provides us with an easy way of modeling general noise in a model. However, the Brownian motion is continuous and sometimes the phenomena we wish to model experience discontinuities; that is, there are jumps in the behaviour of what we are attempting to model.

For modeling of jumps or counting the Poisson process plays a central role.

---

$^1$Notation: $|\Pi| = \sup_{0 \leq j \leq n-1} |t_{j+1} - t_j|$
Definition 2.10 (Poisson Process). Let $N = (N_t)_{0 \leq t \leq T}$ be a stochastic process with values in $\mathbb{N}_0$. We say that $N$ is a Poisson process with intensity $\lambda > 0$ if the following properties hold:

(i) $N_0 = 0$, $\mathbb{P}$-a.s.

(ii) $N$ has independent increments.

(iii) $N$ has stationary increments.

(iv) $N$ has increments that are Poisson distributed with parameter $\lambda$; that is, for all $t, h > 0$, $(N_{t+h} - N_t) \sim \text{Poisson}(\lambda h)$.

Above we have defined the Poisson process by its properties. Some authors define the Poisson process with a more specific expression. For such an alternative definition of the Poisson process, see e.g. Definition 2.17 in Cont, Tankov [22].

Typically one uses the Poisson process to count the number of jumps when modeling a phenomena which experiences jumps, and then chooses some distribution for the sizes of the jumps. The resulting process is what we call a compound Poisson process $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$, which is defined by

$$X_t = \sum_{j=1}^{N_t} \zeta_j, \quad 0 \leq t \leq T,$$

where $N = (N_t)_{0 \leq t \leq T}$ is a Poisson process and $(\zeta_j)_{j \in \mathbb{N}} \subset \mathbb{R}^d$ is a sequence of i.i.d. stochastic variables.

As mentioned above, the Poisson process is very popular for modeling discontinuous behaviour. However, it is possible to define a process which is more general than the classical Poisson process. This generalization can be defined by introducing a stochastic intensity instead of the deterministic intensity of the classical Poisson process. The introduction of a stochastic intensity will allow for much greater flexibility. These $\mathbb{N}_0$-valued processes with stochastic intensities are called Cox processes.

We define the Cox process as in Bening, Korolev [3].

Definition 2.11 (Cox Process). Let $N_1 = (N_1(t))_{0 \leq t \leq T}$ be a Poisson process with intensity equal to 1. Let $\mu = (\mu_t)_{0 \leq t \leq T}$ be a stochastic process, independent of $N_1$, with values in $\mathbb{R}_{\geq 0}$ and non-decreasing paths. In addition, assume that $\mu$ satisfies the conditions

$$\mu_0 = 0 \quad \text{and} \quad \mathbb{P}[\mu_t < \infty] = 1, \ 0 \leq t \leq T.$$

Then the time-changed $\mathbb{N}_0$-valued process

$$N^\mu = (N_t^\mu)_{0 \leq t \leq T} = (N_1(\mu_t))_{0 \leq t \leq T},$$

is a Cox process.

We define a compound Cox process similarly to how we defined the compound Poisson process.
We say that a process $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ is a compound Cox process if it is given by

$$X_t = \sum_{j=1}^{N^\mu_t} \zeta_j, \quad 0 \leq t \leq T,$$

where $N^\mu = (N^\mu_t)_{0 \leq t \leq T} \subset \mathbb{N}_0$ is a Cox process, and $(\zeta_j)_{j \in \mathbb{N}} \subset \mathbb{R}^d$ is a sequence of i.i.d. stochastic variables.

### 2.4 Martingale theory

In martingale theory collecting information over time is central and is translated into theory through the following mathematical definition.

**Definition 2.12 (Filtration).** A filtration $\mathcal{F} := (\mathcal{F}_t)_{0 \leq t \leq T}$ on $(\Omega, \mathcal{F})$, is a family of $\sigma$-algebras which satisfies the following property:

$$\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}, \quad 0 \leq s \leq t \leq T.$$

A probability space $(\Omega, \mathcal{F}, \mathbb{P})$ which is equipped with a filtration $\mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}$ is said to be filtered, and for such a probability space we write $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$. We will assume the usual hypotheses for $\mathcal{F}$; that is,

1. $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets.
2. $\mathcal{F}$ is right-continuous; that is,

$$\mathcal{F}_t = \mathcal{F}_{t+} := \bigcap_{u>t} \mathcal{F}_u, \quad 0 \leq t \leq T. \quad (2.5)$$

Property (i) in the above definition helps us avoid the situation where two stochastic variables are equal a.s., but only one is measurable with respect $\mathcal{F}_t$, $0 \leq t \leq T$.

The usual hypotheses is unproblematic to impose on our filtration $\mathbb{F}$, and is mostly a safety measure.

An important type of filtration which one often encounters in practical problems formulated by means of stochastic processes, is the natural filtration of a stochastic process. This filtration will also be encountered on many occasions in this thesis.

Let $X = (X_t)_{0 \leq t \leq T}$ be a stochastic process taking values in $\mathbb{R}^d$. Then the natural filtration $\mathbb{F}^X := (\mathcal{F}^X_t)_{0 \leq t \leq T}$ of $X$, is the $\mathbb{P}$-augmented $\sigma$-algebra generated by $X$; that is,

$$\mathcal{F}^X_t := \sigma\{(X_s)_{0 \leq s \leq t} \cup \mathcal{N}\}, \quad 0 \leq t \leq T,$$

where $\mathcal{N}$ is the family of all $\mathbb{P}$-null sets,

$$\mathcal{N} := \{A \subset \Omega : \exists B \in \mathcal{F}, A \subset B, \mathbb{P}[B] = 0\}.$$

When we say that a filtration $\mathcal{F} := (\mathcal{F}_t)_{0 \leq t \leq T}$ is $\mathbb{P}$-augmented, we mean that $\mathcal{F}_0$ contains all $\mathbb{P}$-null sets. The notation $\mathcal{A} := \sigma\{A\}$, for a set $A \in \mathcal{F}$, means that $\mathcal{A}$ is the smallest $\sigma$-algebra generated by $A$. 

In application, usually \( \mathcal{F}_t^X, 0 \leq t \leq T \), contains all the information that the process \((X_u)_{0 \leq u \leq t}, 0 \leq t \leq T \) provides us with, along with the knowledge of "impossible" events, which we have from \( \mathcal{N} \). By adding the family of null sets, \( \mathcal{N} \), we have that \( \mathcal{N} \subset \mathcal{F}_0 \). Hence if an outcome is impossible, we already know of its impossibility at time \( t = 0 \). The following gives an example of such an impossible event.

Example 2.2 (Impossible event). Let \( S = (S_t)_{0 \leq t \leq T} \) be a stochastic process, where \( S_t \) is the price of a given stock at time \( t \leq T \). Let \( H \) be the event that "\( S_t = \infty \)", for a time \( t \leq T \). Then \( H \) is an impossible event, such that \( H \in \mathcal{N} \). ♠

Another important concept in relation to filtrations and the collecting of information, is that of adaptedness.

Definition 2.13 (Adaptedness). Let \( X = (X_t)_{0 \leq t \leq T} \) be a stochastic process on a filtered probability space \((\Omega, \mathcal{F}, \mathbb{P} := (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})\). We say that \( X \) is \( \mathbb{F} \)-adapted if, for each \( t \leq T \), the random variable \( X_t \) is \( \mathcal{F}_t \)-measurable.

In practice one interprets adaptedness in the following way: Say that \( \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T} \) is a filtration which collects information for us over time. If \( X = (X_t)_{0 \leq t \leq T} \) is a \( \mathbb{F} \)-adapted process, then the value of \( X_t \), for some \( t \leq T \), will be revealed to us at time \( t \) and is contained in \( \mathcal{F}_t \).

Definition 2.14 (Martingale). Let \( M = (M_t)_{0 \leq t \leq T} \) be a stochastic process with values in \( \mathbb{R} \). \( M \) is a martingale (with respect to \( \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T} \), or \( \mathbb{F} \)-martingale), if the following properties hold:

(i) \( \forall 0 \leq t \leq T, \ E[|M_t|] < \infty \).

(ii) \( M \) is \( \mathbb{F} \)-adapted.

(iii) \( \forall 0 \leq s \leq t \leq T, \ E[M_t | \mathcal{F}_s] = M_s \).

Hence, for a martingale \( M = (M_t)_{0 \leq t \leq T} \), if we have observed \( M \) up to a time \( s < T \), our best prediction of the future value \( M_t, s < t \leq T \), is given by \( M_s \). The following gives an example of a well known martingale.

Example 2.3. Let \( B = (B_t)_{0 \leq t \leq T} \subset \mathbb{R} \) be a Brownian motion on a filtered probability space \((\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})\), where \( \mathbb{F} \) is the natural filtration of \( B \). Then \( B \) is a martingale with respect to \( \mathcal{F} \).

Proof. (i): By the definition of the Brownian motion, \( B_t \sim \mathcal{N}(0, t) \). Hence, for all \( 0 \leq t \leq T \), \( B_t \) is integrable.

(ii): A stochastic process is always measurable with respect to its natural filtration at any point in time. Hence \( B \) is \( \mathbb{F} \)-adapted.

(iii): For all \( 0 \leq s \leq t \leq T \),

\[
E[B_t | \mathcal{F}_s] = E[B_t - B_s + B_s | \mathcal{F}_s] \overset{(*)}{=} E[B_t - B_s] + B_s = B_s,
\]

where we have used:

(*) Since the Brownian motion has independent increments, \( B_t - B_s \) is independent of the process \((B_u)_{0 \leq u \leq s} \), and then \( \mathcal{F}_s \).
2.4. MARTINGALE THEORY

We have now arrived at the desired result.

Another important process is the compensated Poisson process, which is a modification of the Poisson process given in Definition 2.10. Let \( N = (N_t)_{0 \leq t \leq T} \) be a Poisson process with intensity \( \lambda > 0 \). Let \( \tilde{N} = (\tilde{N}_t)_{0 \leq t \leq T} \) be the process given by

\[
\tilde{N}_t = N_t - \lambda t, \quad 0 \leq t \leq T.
\]

Then the process \( \tilde{N} \) is called the compensated Poisson process. In similar fashion to the argument carried out in Example 2.3, it is easily verified that the compensated Poisson process is a martingale.

**Definition 2.15** (Stopping time). Let \( \tau \) be a stochastic variable with values in \( \mathbb{R}_{\geq 0} \). We say that \( \tau \) is a \( \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T} \)-stopping time, if

\[
\{ \tau \leq t \} = \{ \omega \in \Omega : \tau(\omega) \leq t \} \in \mathcal{F}_t, \quad 0 \leq t \leq T.
\]

(2.6)

Stopping times can be encountered in a wide variety of applications and definitions. An example of such definitions is in the definition of local martingales, given below. In addition they offer a way of defining random partitionings, which could be good for underlining that, for some problems, the choice of partitioning is not important. This will be the case when we define stochastic integrals with respect to the Brownian motion, in the next section.

**Definition 2.16** (Local martingale). Let \( X = (X_t)_{0 \leq t \leq T} \) be a stochastic process on \( (\Omega, \mathcal{F}, \mathbb{P}) \), and assume that \( X_0 \) is \( \mathcal{F}_0 \)-measurable. Assume that there exists an increasing sequence of \( \mathbb{F} \)-stopping times \( (\tau_n)_{n \in \mathbb{N}} \), such that \( \tau_n \to \infty \) when \( n \to \infty \). Then if the process \( X^{\tau_n} = (X_{t \wedge \tau_n} - X_0)_{0 \leq t \leq T}, \quad n \in \mathbb{N} \), is a \( \mathbb{F} \)-adapted martingale, we say that \( X \) is a local martingale.

It is easily verified that any martingale is also a local martingale. This is argued in the example below, where the key step in our simple argument is to note that stopping times take values in the extended halfline \( [0, \infty] \).

**Example 2.4** (Any martingale is a local martingale). Let \( M = (M_t)_{0 \leq t \leq T} \) be a martingale. Define a sequence of stopping times \( (\tau_n)_{n \in \mathbb{N}} \) such that, for all \( n \in \mathbb{N} \), \( \tau_n = \infty \). Now for all \( (t, n) \in [0, T] \times \mathbb{N} \),

\[
M_{t \wedge \tau_n} := M_{t \wedge \tau_n} - M_0 = M_t - M_0.
\]

Then we have that for all \( (t, n) \in [0, T] \times \mathbb{N} \),

\[
\mathbb{E}[M_{t \wedge \tau_n}] = \mathbb{E}[M_t - M_0 | \mathcal{F}_s] = M_s - M_0 = M_{s \wedge \tau_n},
\]

and we have arrived at the desired result.

**Definition 2.17** (Semimartingale). Let \( S = (S_t)_{0 \leq t \leq T} \) be a \( \mathbb{F} \)-adapted process. Then \( S \) is a semimartingale if it can be decomposed in the following way:

\[
S_t = S_0 + M_t + V_t, \quad 0 \leq t \leq T,
\]

(2.7)

where \( M = (M_t)_{0 \leq t \leq T} \) is a \( \mathbb{F} \)-adapted local martingale and \( V = (V_t)_{0 \leq t \leq T} \) is a \( \mathbb{F} \)-adapted process with finite variation over \( [0, T] \).
Later on when we consider our filtering problems we will apply a theorem called
the Girsanov theorem, on a semimartingale to perform a change of measure. This
change of measure will be an important step in our search for stochastic transition
rates.

In the next section we will introduce some basic concepts in stochastic analysis.

2.5 Stochastic analysis

In what follows, we will consider a complete filtered probability space \((\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})\).

This section will be a rather short introduction to some basic concepts in stochas-
tic analysis. We only introduce the concepts that we consider to be relevant for the
theory presented later on in the thesis. This section is heavily based on Øksendal
[26] and Cont, Tankov [22], and all definitions and propositions can be found in
these references. Also, for a more thorough treatment of stochastic analysis these
references can be consulted.

First we give some definitions.

**Definition 2.18** (Stopping time \(\sigma\)-algebra). Let \(\tau\) be a stopping time. Then we
define the stopping time \(\sigma\)-algebra \(\mathcal{F}_\tau\) by
\[
\mathcal{F}_\tau := \{A \in \mathcal{F} : A \cap \{\tau \leq t\} \in \mathcal{F}_t, 0 \leq t \leq T\}.
\]

With Definition 2.18 in mind, we introduce the following class of processes.

**Definition 2.19** (Simple predictable process). A stochastic process \((f_t)_{0 \leq t \leq T} \subset \mathbb{R}\)
is called a simple predictable process if it can be represented as
\[
f_t = \varphi_0 1_{\{t=0\}} + \sum_{j=0}^{n} \varphi_j 1_{(\tau_j, \tau_{j+1}]}(t), \quad 0 \leq t \leq T,
\]
where \((0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_n = T)\) is a sequence of stopping times and, for all
\(j = 0, \ldots, n, \varphi_j\) is \(\mathcal{F}_{\tau_j}\)-measurable.

Henceforth we will denote the class of all simple predictable processes \(f = (f_t)_{0 \leq t \leq T} \subset \mathbb{R}\) by \(\mathbb{L}\).

Finally we define the stochastic integral with respect to the Brownian motion (as
in Cont, Tankov [22]).

**Definition 2.20** (Stochastic integrals with respect to Brownian motion). Let \(B = (B_t)_{0 \leq t \leq T} \subset \mathbb{R}\) be a Brownian motion. Let \(f = (f_t)_{0 \leq t \leq T} \subset \mathbb{R}\) be a simple predictable
process with the representation
\[
f_t = \varphi_0 1_{\{t=0\}} + \sum_{j=0}^{n} \varphi_j 1_{(\tau_j, \tau_{j+1}]}(t), \quad 0 \leq t \leq T.
\]

Then we define the stochastic integral with respect to the Brownian motion \(B\) as
\[
\int_0^t f_s dB_s = \sum_{j=0}^{n} \varphi_j (B_{\tau_{j+1} \wedge t} - B_{\tau_j \wedge t}), \quad 0 \leq t \leq T.
\]
The use of stopping times in the definition of the stochastic integral, implies that which partitioning of \([0,T]\) we choose does not matter.

In addition we have the following properties for stochastic integrals of simple predictable processes, with respect to the Brownian motion:

**Proposition 2.3.** For a Brownian motion \(B = (B_t)_{0 \leq t \leq T} \subset \mathbb{R}\), \(f, g \in \mathbb{L}\) and constants \(x, y \in \mathbb{R}\), we have the following:

(i) \((\int_0^t f_s dB_s)_{0 \leq t \leq T}\), is a \(\mathbb{F}\)-martingale if \(\mathbb{E}\left[\int_0^T (f_s)^2 \, ds\right] < \infty\).

(ii) \(\forall 0 \leq t \leq T, \mathbb{E}\left[\int_0^t f_s dB_s\right] = 0\) — Mean zero.

(iii) \(\forall 0 \leq t \leq T, \mathbb{E}\left[\int_0^t (f_s dB_s)^2\right] = \mathbb{E}\left[\int_0^t (f_s)^2 \, ds\right]\) — Itô isometry.

(iv) \(\forall 0 \leq t \leq T, \int_0^t (xf_s + yg_s) \, dB_s = x \int_0^t f_s \, dB_s + y \int_0^t g_s \, dB_s\) — Linearity.

Using Proposition 2.3 one can extend the concept of a stochastic integral of integrand processes in \(\mathbb{L}\) to the case of integrand processes, which are \(\mathbb{F}\)-adapted processes \((f_t)_{0 \leq t \leq T}\) such that

\[ \mathbb{E}\left[\int_0^T (f_s)^2 \, ds\right] < \infty. \]

For this class of integrand processes the properties (i),(ii),(iii),(iv) in Proposition 2.3 still hold.

As mentioned above, the reader may consult Øksendal [26] for a more thorough discussion on integration with respect to the Brownian motion.

Having introduced basic concepts for integration with respect to the Brownian motion, we will now introduce the Girsanov theorem. Later on we will use a version of the Girsanov theorem to impose a change of measure, and construct a new Brownian motion under the new measure. The version we will use later can be found in Jacod, Shiryaev [11], a version which do not state in this thesis. However we state a simpler version below, which gives a good intuition of how the Girsanov theorem can be applied.

**Theorem 2.4** (The Girsanov theorem). Let \(X = (X_t)_{0 \leq t \leq T}\) be a \(\mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}\)-adapted process with values in \(\mathbb{R}^d\). Let \(B = (B_t)_{0 \leq t \leq T}\) be a \(d\)-dimensional Brownian motion. Let \(\theta = (\theta_t)_{0 \leq t \leq T}\) be a \(\mathbb{F}\)-adapted process with values in \(\mathbb{R}^d\), which satisfies

\[ \int_0^T \|\theta_s\|^2 \, ds < \infty, \quad a.s., \quad (2.8) \]

and

\[ \mathbb{E}\left[\exp\left\{\sum_{j=1}^d \int_0^T \theta_s^{(j)} \, dB_s^{(j)} - \frac{1}{2} \int_0^T \|\theta_s\|^2 \, ds\right\}\right] = 1. \quad (2.9) \]

Let \(\mathbb{Q}\) be a probability measure on \((\Omega, \mathcal{F}, \mathbb{P})\), given by

\[ \mathbb{Q}[A] = \mathbb{E}[1_A M_T], \quad A \in \mathcal{F}, \quad (2.10) \]
where $M = (M_t)_{0 \leq t \leq T}$ is the process

$$M_t := \exp \left\{ \sum_{j=1}^{d} \int_{0}^{t} \theta_s^{(j)} \, dB_s^{(j)} - \frac{1}{2} \int_{0}^{t} \|\theta_s\|^2 \, ds \right\}, \quad 0 \leq t \leq T. \quad (2.11)$$

Then $\mathbb{P} \sim \mathbb{Q}$ and the process $\tilde{B} = (\tilde{B}_t)_{0 \leq t \leq T}$, with components

$$\tilde{B}_t^{(i)} := B_t^{(i)} - \int_{0}^{t} \theta_s^{(i)} \, ds, \quad 0 \leq t \leq T, \quad i = 1, \ldots, d, \quad (2.12)$$

is a Brownian motion under $\mathbb{Q}$.

When we say that $\tilde{B}$ is a Brownian motion under $\mathbb{Q}$, in Theorem 2.4, we mean that $\tilde{B}$ is a stochastic process on $(\Omega, \mathcal{F}, \mathcal{F}, \mathbb{Q})$ which is defined according to Definition 2.9.

It is important to note that, even though the new Brownian motion $\tilde{B}$ defined by (2.12) is a Brownian motion under a new equivalent probability measure, $\tilde{B}$ is generally not a Brownian motion under $\mathbb{P}$.

**Example 2.5** (Brownian motion and the Girsanov theorem). Let $B = (B_t)_{0 \leq t \leq T}$ be a 1-dimensional Brownian motion. Let $\alpha > 0$ be a given constant. Then we may construct a probability measure $\mathbb{Q}$ by use of the Girsanov theorem (Theorem 2.4) under which the process $W = (W_t)_{0 \leq t \leq T}$, given by

$$W_t = B_t - \int_{0}^{t} \alpha \, ds, \quad 0 \leq t \leq T,$$

is a Brownian motion.

**Claim:** $W$ is not a Brownian motion under $\mathbb{P}$.

**Proof.** We have for all $0 \leq t \leq T$,

$$\mathbb{E}[W_t] = \mathbb{E} \left[ B_t - \int_{0}^{t} \alpha \, ds \right] = \mathbb{E}[B_t] - t\alpha = -t\alpha,$$

where we have used that $B_t$, $0 \leq t \leq T$, has mean zero. Hence $W$ is not a Brownian motion under $\mathbb{P}$.

---

### 2.6 Monte Carlo simulation

When we tackle practical problems later on, Monte Carlo techniques will be a central concept. Monte Carlo techniques are ways of approximating the solution to some problem by associating it with a probability distribution, and then sampling repeatedly from that distribution. Then by having generated stochastic variables from a chosen distribution, one relates the problem to the mean of, possibly with some transformation applied to, those stochastic variables. The theorem that follows is the foundation of Monte Carlo techniques.
Theorem 2.5 (Kolmogorov’s Strong Law of Large Numbers). Let \((X_n)_{n \in \mathbb{N}}\) be a sequence of i.i.d. stochastic variables with values in \(\mathbb{R}\). Assume that

\[
\mathbb{E}[|X_1|] < \infty.
\]

Then

\[
\frac{1}{\ell} \sum_{n=1}^{\ell} X_n \xrightarrow{\ell \to \infty} \mathbb{E}[X_1], \quad \text{a.s.}
\]

Kolmogorov’s strong law of large numbers (alternatively, SLLN for short) is one of the most central results when we attempt solving stochastic problems, numerically. Consider the case where we are interested in computing some quantity which is subject to chance. Say we have been able to relate that problem to the mean, of a transformation \(\varphi(\cdot)\), of a stochastic variable \(\xi\), where \(\xi\) has some distribution \(\Theta\). Then we have the following by Kolmogorov’s SLLN. Let \((\xi_j)_{j \in \mathbb{N}}\) be a sequence of independent and identically distributed stochastic variables with distribution \(\Theta\), then we may approximate the mean of \(\varphi(\xi)\) by

\[
\mathbb{E}[\varphi(\xi)] \approx \varphi := \frac{1}{\ell} \sum_{j=1}^{\ell} \varphi(\xi_j),
\]

where \(\ell \in \mathbb{N}\) is large. The procedure is written in Algorithm 2.2.

Often we can solve rather difficult problems, by reducing the problems to be expressed by an expected value, and then in turn apply Monte Carlo techniques.

**Algorithm 2.2 Monte Carlo**

**Input:** Function \(\varphi(\cdot)\); distribution \(\Theta\); a chosen \(\ell \in \mathbb{N}\).

1: Generate \(\xi_j \sim \Theta, \quad j = 1, \ldots, \ell\)

2: \(\varphi \leftarrow \frac{1}{\ell} \sum_{j=1}^{\ell} \varphi(\xi_j)\)

3: return \(\varphi\)
Chapter 3

Lévy processes

Bibliographical notes

The main references of this chapter are: Applebaum [1], Çinlar [5], Sato [21] and Cont, Tankov [22].

Lévy processes are being applied to a wide range of fields these days. These process has gotten special attention, especially in the field of finance and insurance. Many phenomena which in practice one discovers to move discontinuously, have earlier been assumed to move continuously over time in the models, and as a consequence have been modeled by strictly continuous processes. An example of such a process is the Brownian motion with drift. Processes like the Brownian motion, are now typically replaced by more general Lévy processes in order to account for sudden surprising movements. A good example of a setting which the above applies to, is the movement of stock prices. The most famous way of describing stock price movement is through the classical Black-Scholes framework, where one assumes the stock prices to move continuously with Gaussian noise. Hence, in the Black-Scholes framework one can not be surprised by sudden big changes in the price of the stock. However, in reality one observes that stock prices sometimes exhibit sudden large movements. Such sudden large movements can not be modeled through a strictly continuous model, without setting some of the parameters in the model to unrealistically high values. It is when we model such behaviour, that we may turn to Lévy processes for more flexibility.

3.1 Lévy processes and their properties

We start this section with the definition of a Lévy process.

**Definition 3.1** (Lévy process). Let $\eta = (\eta_t)_{0 \leq t \leq T}$ be a stochastic process with values in $\mathbb{R}^d$. $\eta$ is said to be a Lévy process if it satisfies the following properties:

(i) $\eta_0 = 0$, $\mathbb{P}$-a.s.

(ii) $\eta$ has independent increments.
(iii) $\eta$ has stationary increments.

(iv) $\eta$ is stochastically continuous; that is, for all $h, \varepsilon > 0$, we have

$$\lim_{h \downarrow 0} \mathbb{P}[\|\eta_{t+h} - \eta_t\| > \varepsilon] = 0, \quad 0 \leq t \leq T.$$ 

Here the property of stochastic continuity does not imply in any way that Lévy processes are continuous. However, the property of stochastic continuity implies that for a Lévy processes which can move discontinuously, the probability of a discontinuity occurring at a given point in time is always zero. This means that all discontinuities of Lévy processes occur at completely random times. Such discontinuous movements will henceforth often be referred to as jumps.

It can be shown that every Lévy process has a modification which is càdlàg; that is, every Lévy process has a modification that is right continuous with existing left limits. Where by modification we mean that a stochastic process $Y = (Y_t)_{0 \leq t \leq T}$ is a modification of another stochastic process $X = (X_t)_{0 \leq t \leq T}$ if

$$\mathbb{P}[X_t = Y_t] = 1, \quad 0 \leq t \leq T.$$ 

When considering Lévy processes in this thesis, we will always consider the càdlàg modification of the process, such that when we define a Lévy process it will always have càdlàg paths.

**Remark 2.** As in Applebaum [1], we make a remark of that the càdàg property of Lévy processes holds almost surely. More specifically (by Sato [21]): For any Lévy process $\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$, we have the following. There exists a $\Omega_0 \in \mathcal{F}$, $\mathbb{P}[\Omega_0] = 1$, such that, for all $\omega \in \Omega_0$, the paths $[0, T] \ni t \mapsto \eta_t(\omega)$ are càdlàg; that is,

- For all $\omega \in \Omega_0$, $[0, T] \ni t \mapsto \eta_t(\omega)$ is right-continuous.

- For all $\omega \in \Omega_0$, $(0, T] \ni t \mapsto \eta_t(\omega)$ has existing left limits.

More generally, when we speak of càdlàg in the setting of stochastic processes, one should always read "a.s. càdlàg".

By Definition 3.1 it is clear that both the Brownian motion and the Poisson process are Lévy processes. The only difference in the properties of these two processes, the Brownian motion and the Poisson process, are their incremental distributions. As a consequence of their distributional differences, they move very differently and might be applied to completely different problems. The Poisson process is an example of process which only moves by jumps; more specifically, a Poisson process has piecewise constant paths that are càdlàg, hence the only change in the value of a Poisson process occur by discontinuities. Whereas the Brownian motion is continuous in its movement. The two processes are often used together to capture different aspects of the movement of some phenomenon. Typically one uses the Poisson process to capture sudden unexpected movements, and the Brownian motion to capture more general noise. Two examples of the paths of a Brownian motion and a Poisson process, are shown in the plots in Figure 3.1.

The example of the Brownian motion and the Poisson process underlines how rich a class of processes, the class of Lévy processes is.
3.1.1 Jumps and random measures

Here in this subsection we will look closer at the jump behaviour of Lévy processes. Also, we will introduce important concepts such as Poisson integration and the Lévy measure.

First we give a formal definition of a random measure and a Poisson random measure, as these will help better understand the theory which we will introduce in the sequel. These two definitions are be borrowed from Çinlar [5].

We start with the definition of a random measure.

Definition 3.2 (Random Measure). Let $(\mathcal{K}, \mathcal{H})$ be a measurable space. A mapping $M : \Omega \times \mathcal{H} \rightarrow \mathbb{R}_{\geq 0}$ is a random measure if the following hold:

(i) $\forall A \in \mathcal{H}$, $\omega \mapsto M(\omega, A)$ is a stochastic variable.

(ii) $\forall \omega \in \Omega$, $A \mapsto M(\omega, A)$ is a measure on $(\mathcal{K}, \mathcal{H})$.

In connection with the above definition, we now introduce the Poisson random measure.

Definition 3.3 (Poisson random measure). Let $(\mathcal{S}, \mathcal{S})$ be a measurable space and let $\mu$ be a measure on it. A random measure $N$ on $(\mathcal{S}, \mathcal{S})$ is said to be a Poisson random measure with intensity measure $\mu$ if the following hold:

(i) For each $A \in \mathcal{S}$, $N(A) \sim \text{Poisson}(\mu(A))$.

(ii) If $(A_j)_{j=1}^n \subset \mathcal{S}$ is a sequence of disjoint sets, $(N(A_j))_{j=1}^n$ are independent.

Poisson random measures will often be encountered in this thesis, and will be one of the main tools for modeling jump behaviour.
3.1.2 Jumps of Lévy processes

Let $\eta = (\eta_t)_{0 \leq t \leq T}$ be a Lévy process with values in $\mathbb{R}^d$. We define the jump sizes of $\eta$ in the following way:

$$\Delta \eta_t := \eta_t - \eta_t^-, \quad 0 \leq t \leq T,$$

where $\eta_t^-$ is the left limit of $\eta_t$; that is,

$$\eta_t^- := \lim_{s \to t^-} \eta_s, \quad 0 \leq t \leq T.$$

Also, we will use the convention that for $t = 0$, $\eta_t^- = 0$.

Now we define the jump measure of $\eta$ by

$$N(t, A) = \sum_{0 \leq s \leq t} 1_A(\Delta \eta_s), \quad 0 \leq t \leq T, \quad (3.1)$$

where $A \subset \mathbb{R}^d_0$ is a Borel set with $0 \notin \bar{A}$. Here $\bar{A}$ denotes the closure of $A$.

By (3.1) we see that for any pair $(t, \omega) \in [0, T] \times \Omega$, $N(t, \cdot, \omega)$ is a $\mathbb{N}_0$-valued counting measure on $(\mathbb{R}^d_0, \mathcal{B}(\mathbb{R}^d_0))$ which counts the jumps of $\eta$, up to time $t \leq T$, with sizes in $A$. For any $A \in \mathcal{B}(\mathbb{R}^d_0)$, $[0, T] \times \Omega \ni (t, \omega) \mapsto N(t, A, \omega)$ is a jump process, and the paths only change their value at their discontinuities.

We will often write the jump measure and other similar measures in their differential form $N(dt, dz)$, $0 \leq t \leq T$, $z \in \mathbb{R}^d$. The following proposition gives a result which is important to be aware of.

**Proposition 3.1.** Let $N : \mathcal{B}([0, T]) \otimes \mathcal{B}(\mathbb{R}^d_0) \times \Omega \to \mathbb{N}_0$ be the jump measure of a Lévy process $\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$. For any Borel set $A \subset \mathbb{R}^d_0$ with $0 \notin \bar{A}$, we have

$$N(t, A) < \infty, \quad 0 \leq t \leq T.$$

**Proof.** See e.g. Lemma 2.3.4 in Applebaum [1].

By putting $A \subset \mathbb{R}^d_0$ to be Borel and requiring $0 \notin \bar{A}$, Proposition 3.1 states that, for any $t \leq T$, any Lévy processes experiences a finite number of jumps in $A$ over the time interval $[0, t]$.

The reason that we require the closure of $A$ to not contain zero, is that if we allow for zero to be included in the closure of $A$, we might end up with an explosion of jumps with jump size close to zero. Here by explosion, we mean an accumulation of an infinite number of jumps. Hence, if we allow for 0 to be included in the closure of $A$, the sum in (3.1) could potentially not be well defined.

For less cumbersome notation we will denote the family of all Borel sets $A \subset \mathbb{R}^d_0$ with $0 \notin \bar{A}$, by $\mathcal{B}_0^d$.

We will discuss the jump measure further after we have given an introduction to the Lévy measure.

**Definition 3.4** (Lévy measure of Lévy process). Let $\eta = (\eta_t)_{0 \leq t \leq T}$ be a Lévy process with values in $\mathbb{R}^d$. Let $N(dt, dz)$, $0 \leq t \leq T$, $z \in \mathbb{R}^d_0$, be the jump measure of $\eta$. Then the Lévy measure $\nu$ of $\eta$ is defined by

$$\nu(A) = \mathbb{E}[\# \{0 \leq t \leq 1 : \Delta \eta_t \in A\}] = \mathbb{E}[N(1, A)], \quad A \in \mathcal{B}(\mathbb{R}^d_0). \quad (3.2)$$

*Notation: $\mathbb{R}^d_0 := \mathbb{R}^d \setminus \{0\}$*
The above Lévy measure is not necessarily finite. However, it will be finite if $0$ is not included in the closure of $A$, by the previous reasoning.

**Example 3.1 (Lévy measure of Compound Poisson process).** Let $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ be a compound Poisson process with intensity $\lambda > 0$ and jump sizes $(Y_n)_{n \in \mathbb{N}} \subset \mathbb{R}^d$. Let $N_1 = (N_1(t))_{0 \leq t \leq T} \subset \mathbb{N}_0$ be a Poisson process, independent of $(Y_n)_{n \in \mathbb{N}}$, with intensity equal to 1. Let $N(dt, dz)$ be the jump measure of $X$. Then we find the Lévy measure of $X$ by

$$
\nu(A) = \mathbb{E}[N(1, A)] = \mathbb{E} \left[ \sum_{0 \leq s \leq 1} \mathbf{1}_A(\Delta X_s) \right] = \mathbb{E} \left[ N_1(\lambda) \sum_{n=1}^{N_1(\lambda)} \mathbf{1}_A(Y_n) \right] = \mathbb{E} \left[ N_1(\lambda) \mathbb{E}[\mathbf{1}_A(Y_1)] \right] = \lambda \mathbb{P}[Y_1 \in A].
$$

It is important to be aware of that even though the above definition concerns the Lévy measure of a process, we do not have to define the Lévy measure to be associated with a Lévy process. More generally we define the Lévy measure $\nu$, to be a measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ which satisfies the following:

$$
\nu(\{0\}) = 0 \quad \text{and} \quad \int_{\mathbb{R}^d} (\|z\|^2 \wedge 1) \nu(dz) < \infty. \quad (3.3)
$$

The following gives an example of such a Lévy measure.

**Example 3.2 (Lévy measure).** Let $\xi \sim \text{U}(-\varepsilon, \varepsilon)$ for some $\varepsilon > 0$. Then we could define a Lévy measure

$$
\nu(A) = \mathbb{P}[\xi \in A], \quad A \in \mathcal{B}(\mathbb{R}),
$$

which obviously is a measure that satisfies (3.3).

As mentioned above, the Lévy measure of a Lévy process $\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$, is not necessarily finite. Even though the integrability condition in (3.3) holds, we could have

$$
\int_{\mathbb{R}^d} (\|z\| \wedge 1) \nu(dz) = \infty,
$$

which would imply an aggregation of infinitely many jumps with sizes around zero, a so called explosion around zero.

**Proposition 3.2.** A Lévy process with values in $\mathbb{R}^d$ and Lévy measure $\nu$, has jumps of finite variation if and only if

$$
\int_{\mathbb{R}^d} \mathbf{1}_{\{\|z\| \leq 1\}} \nu(dz) < \infty.
$$

**Proof.** See e.g. Theorem 21.9 in Sato [21] or Proposition 3.9 in Cont, Tankov [22].
By finite variation we mean the following: Let \( f = (f(t))_{0 \leq t \leq T} \) be a stochastic process with values in \( \mathbb{R}^d \). Let \( (\Pi^{(m)})_{m \in \mathbb{N}} = (0 = t_0^{(m)} < t_1^{(m)} < \cdots < t_{n_m}^{(m)} = T)_{m \in \mathbb{N}} \) be a sequence of partitions of \([0, T]\). Then if

\[
\sup_{\Pi \in (\Pi^{(m)})_{m \in \mathbb{N}}} \sum_{j=0}^{n_m-1} \|f(t_{j+1}) - f(t_j)\| < \infty, \quad \text{a.s.,}
\]

we say that \( f \) has finite variation.

We now return to our discussion on the jump measure \( N(dt, dz) \).

The jump measure \( N(dt, dz) \), \( 0 \leq t \leq T, z \in \mathbb{R}_0^d \), of a Lévy process is in fact a Poisson random measure with intensity \( dt \nu(dz) \), where \( \nu \) is the Lévy measure of the respective Lévy process. For this reason one sometimes refers to the jump measure of a Lévy process \( \eta = (\eta_t)_{0 \leq t \leq T} \) as the Poisson random measure associated with \( \eta \).

The theorem that follows gives some interesting properties of the jump measure.

**Theorem 3.3.** Let \( \eta = (\eta_t)_{0 \leq t \leq T} \) be a Lévy process with values in \( \mathbb{R}^d \) and Lévy measure \( \nu \). Let \( N(dt, dz) \), \( 0 \leq t \leq T, z \in \mathbb{R}_0^d \), be the jump measure of \( \eta \). We now have the following.

(i) For any \( A \in \mathcal{B}_0^d \), \( (N(t, A))_{0 \leq t \leq T} \) is a Poisson process with expected value

\[
\mathbb{E}[N(t, A)] = t \nu(A), \quad 0 \leq t \leq T.
\]

(ii) If \( (A_j)_{j=1}^n \subset \mathcal{B}_0^d \) is a sequence of disjoint sets and if \( (t_j)_{j=1}^n \subset [0, T] \) is a sequence of distinct times, then \( (N(t_j, A_j))_{j=1}^n \) is a sequence of independent stochastic variables.

**Proof.** See e.g. Theorem 2.3.5 in Applebaum [1].

Let \( \eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d \) be a Lévy process with Lévy measure \( \nu \). Let \( N(dt, dz) \), \( 0 \leq t \leq T, z \in \mathbb{R}_0^d \), be the jump measure of \( \eta \). Now in similar fashion to how we defined the compensated Poisson process in the previous chapter, we define the compensated jump measure (alternatively, compensated Poisson random measure) \( \tilde{N}(dt, dz) \), \( 0 \leq t \leq T, z \in \mathbb{R}_0^d \), as

\[
\tilde{N}(dt, dz) := N(dt, dz) - dt \nu(dz), \quad 0 \leq t \leq T, z \in \mathbb{R}_0^d.
\]

It is well known and easily verified that the above compensated jump measure is a martingale for any \( A \in \mathcal{B}_0^d \). To see this we compute the following conditional expectation, for all \( A \in \mathcal{B}_0^d \),

\[
\mathbb{E}[\tilde{N}(t, A)|\mathcal{F}_s] = \mathbb{E}[N(t, A) - t \nu(A)|\mathcal{F}_s]
\]

\[
= \mathbb{E}[N(t, A) - N(s, A)|\mathcal{F}_s] + N(s, A) - t \nu(A)
\]

\[
(\text{by \( \mathbb{E}[N(t - s, A)] + N(s, A) - t \nu(A) \))
\]

\[
= (t - s) \nu(A) + N(s, A) - t \nu(A)
\]

\[
= N(s, A) - s \nu(A), \quad 0 \leq s \leq t \leq T,
\]

where we have used:
The Poisson random measure will not be the only random measure we encounter. Sometimes we will consider more general integer valued random measures. More specifically, we will consider a \( N_0 \)-valued random measure \( M(dt, dz), 0 \leq t \leq T, z \in \mathbb{R}^d \), where for each \( A \in \mathcal{B}_0 \), \((M(t, A))_{0 \leq t \leq T}\) is a Cox process. Hence \( M(dt, dz) \) is not a Poisson random measure, but still a \( N_0 \)-valued random measure and behave quite similarly to the classical Poisson random measure.

### 3.1.3 Poisson integration

The theory in this subsection is mainly borrowed from Chapter 2 in Applebaum [1]. When trying to model discontinuous movement, one often applies integrals with respect to random measures and in particular Poisson random measures associated with Lévy processes. In this thesis we take advantage of such integrals extensively, to model jump behaviour.

Let, \( \eta = (\eta_t)_{0 \leq t \leq T} \) be a Lévy process with values in \( \mathbb{R}^d \). Let \( N(dt, dz), 0 \leq t \leq T, z \in \mathbb{R}^d \), be the jump measure of \( \eta \) and let \( A \in \mathcal{B}_0 \). Then for any function \( f : \mathbb{R}^d \to \mathbb{R}^d \) which is finite over \( A \), we define the Poisson integral

\[
\int_A f(z)N(t, dz) = \sum_{0 \leq s \leq t} f(\Delta \eta_s)1_A(\Delta \eta_s), \quad 0 \leq t \leq T. \tag{3.4}
\]

This integral (by the corollary on page 27 in Protter [18]) is a Lévy process.

**Theorem 3.4.** Let \( \eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d \) be a Lévy process with jump measure \( N(dt, dz), 0 \leq t \leq T, z \in \mathbb{R}^d \). Let \( \nu \) be the Lévy measure of \( \eta \) and let \( A \in \mathcal{B}_0 \). Now for any mapping \( f : \mathbb{R}^d \to \mathbb{R}^d \) which satisfies the integrability condition

\[
\int_A \|f(z)\| \nu(dz) < \infty,
\]

we have the expected value

\[
\mathbb{E} \left[ \int_A f(z)N(t, dz) \right] = t \int_A f(z) \nu(dz).
\]

**Proof.** See e.g. Theorem 2.3.7 in Applebaum [1].

### 3.1.4 Decomposition of Lévy processes

The following theorem shows that any Lévy process can be decomposed into a drift component, a Brownian motion, a compound Poisson process, and a jump process which is a martingale.

**Theorem 3.5** (Lévy-Itô decomposition). Let \( \eta = (\eta_t)_{0 \leq t \leq T} \) be a Lévy process with values in \( \mathbb{R}^d \), and Lévy measure \( \nu \). Then \( \eta \) may be decomposed into four independent Lévy processes, in the following way:

\[
\eta_t = \mu t + \sigma B_t + X_t + \lim_{\epsilon \searrow 0} \tilde{X}_t^\epsilon, \quad 0 \leq t \leq T, \tag{3.5}
\]
where $\mu \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d \times m}$ have constant entries; and $B$ is a $m$-dimensional Brownian motion. The processes $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ and $\tilde{X}^\varepsilon = (\tilde{X}_t^\varepsilon)_{0 \leq t \leq T} \subset \mathbb{R}^d$ are given by, for all $0 \leq t \leq T$,
\[
X_t := \int_{[0,t] \times \mathbb{R}^d} 1_{\{|z| \geq 1\}} z N(ds, dz) = \sum_{0 \leq s \leq t} \Delta \eta_s 1_{\{|\Delta \eta_s| \geq 1\}},
\]
and, for all $0 \leq t \leq T$,
\[
\tilde{X}_t^\varepsilon := \int_{[0,t] \times \mathbb{R}^d} 1_{\{|z| < 1\}} z \tilde{N}(ds, dz)
= \sum_{0 \leq s \leq t} \Delta \eta_s 1_{\{|z| < 1\}} z \nu(dz),
\]
where $N(dt, dz)$ is the jump measure of $\eta$ with intensity $dt \nu(dz)$; $\nu$ is the Lévy measure of $\eta$; and $\tilde{N}(dt, dz)$ is the compensated jump measure. Also, as $\varepsilon \searrow 0$, the convergence of $X_t^\varepsilon$ is almost sure and uniform in $t \leq T$.

Proof. See e.g. Chapter 4 in Sato [21]. For the proof of the part about the convergence of $X_t^\varepsilon$, the reader may alternatively consult the proof of Theorem 1.23 in Chapter 7 in Çınlar [5].

Remark 3. (i) $\tilde{X}^t = (\tilde{X}_t^\varepsilon)_{0 \leq t \leq T}$, in Theorem 3.5, is a Martingale (by Proposition 2.16 in Cont, Tankov [22]).

(ii) For the decomposition (3.5) in Theorem 3.5 we could have chosen to define
\[
X_t := \int_{[0,t] \times \mathbb{R}^d} 1_{\{|z| \geq K\}} z N(ds, dz), \quad 0 \leq t \leq T,
\]
and
\[
\tilde{X}_t^\varepsilon := \int_{[0,t] \times \mathbb{R}^d} 1_{\{|z| < K\}} z \tilde{N}(ds, dz), \quad 0 \leq t \leq T,
\]
for any $K > 0$. Thus the choice of $K$ is arbitrary, but one usually chooses $K = 1$ for simplicity.

The reason for not defining $\varepsilon \equiv 0$ directly in the decomposition (3.5), and including the compensated jump measure, is due to the problems that might occur in relation to the possibility of having an explosion of jumps close to zero.

The Lévy-Itô decomposition of a Lévy process $\eta = (\eta_t)_{0 \leq t \leq T}$ with values in $\mathbb{R}^d$ may be written in component form as, for all $i = 1, 2, \ldots, d$,
\[
\eta_t^{(i)} = \mu_t + \sum_{j=1}^m \sigma_{t,j} B_t^{(j)} + X_t^{(i)} + \lim_{\varepsilon \searrow 0} \tilde{X}_t^{\varepsilon,(i)}, \quad 0 \leq t \leq T,
\]
where, for all $i = 1, 2, \ldots, d$,
\[
X_t^{(i)} := \int_{[0,t] \times \mathbb{R}^d} 1_{\{|z| \geq 1\}} z_i N(ds, dz) = \sum_{0 \leq s \leq t} \Delta \eta_s^{(i)} 1_{\{|\Delta \eta_s| \geq 1\}}, \quad 0 \leq t \leq T,
\]
and, for all $i = 1, 2, \ldots, d$,

$$
\tilde{X}_t^{(i)} := \int \int_{[0,t] \times \mathbb{R}^d} 1_{\{\varepsilon \leq \|z\| < 1\}} z_i \tilde{N}(ds, dz)
$$

$$
= \sum_{0 \leq s \leq t} \Delta \eta_s^{(i)} 1_{\{\varepsilon \leq \|\Delta \eta_s\| < 1\}} - t \int_{\mathbb{R}^d} 1_{\{\varepsilon \leq \|z\| < 1\}} z_i \nu(dz), \quad 0 \leq t \leq T.
$$

For a Lévy process $\eta$ with the decomposition (3.5) we say that $\eta$ has the Lévy (or characteristic) triplet $(\mu, \sigma, \nu)$. By the Lévy-Itô decomposition, any Lévy process can be decomposed into a Brownian motion with drift, and two additional terms, where one of the terms represents the really small jumps and one term that represents the other jumps.

In the case of a Lévy process $\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$, where we have finite variation for the jumps of $\eta$, it follows from Proposition 3.2 that $\eta$ may be decomposed in the following way:

$$
\eta_t = \mu t + \sigma B_t + \int \int_{[0,t] \times \mathbb{R}^d} z N(ds, dz) = \mu t + \sigma B_t + \sum_{0 \leq s \leq t} \Delta \eta_s, \quad 0 \leq t \leq T,
$$

where $\mu \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d \times m}$ have constant entries; $B = (B_t)_{0 \leq t \leq T}$ is a $m$-dimensional Brownian motion; and $N(dt, dz)$ is the jump measure of $\eta$.

**Example 3.3.** Let $X = (X_t)_{0 \leq t \leq T}$ be a compound Poisson process with values in $\mathbb{R}$ and Lévy measure $\nu$. Let $N(dt, dz)$, $0 \leq t \leq T$, $z \in \mathbb{R}^d_0$, be the jump measure of $X$. Then

$$
X_t = \sum_{0 \leq s \leq t} \Delta X_s = \int \int_{[0,t] \times \mathbb{R}^d_0} z N(ds, dz), \quad 0 \leq t \leq T,
$$

such that $X$ is a Lévy process with Lévy triplet $(0, 0, \nu)$.

The above decompositions have dealt with Lévy processes in general, and Lévy processes with jumps that have finite variation. For continuous Lévy processes the following theorem gives a decomposition.

**Theorem 3.6** (Decomposition of continuous Lévy processes). Let $\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ be a continuous Lévy process. Then $\eta$ may be decomposed as,

$$
\eta_t = \mu t + \sigma B_t, \quad 0 \leq t \leq T,
$$

where $\mu \in \mathbb{R}^d$ and $\sigma \in \mathbb{R}^{d \times m}$ have constant entries; and $B = (B_t)_{0 \leq t \leq T}$ is a $m$-dimensional Brownian motion.

**Proof.** See e.g. Theorem 4.3 in Chapter 7 in Çınlar [5].

### 3.1.5 Distributional properties of Lévy processes

As mentioned in the previous chapter, there is a 1–1 relation between probability distributions and characteristic functions. When considering Lévy processes, it can be shown that their characteristic functions take on a specific form depending on their Lévy triplet. This makes the characteristic function a popular tool for analyzing the distributional properties of Lévy processes.

Now we introduce a definition which implies that the increments of Lévy processes can not take on any given distribution.
**Definition 3.5** (Infinite divisibility). Let $X$ be a stochastic variable with values in $\mathbb{R}^d$. We say that $X$ has an infinitely divisible distribution if for all integers $n \geq 2$, there exists a sequence $(\xi_i^{(n)})_{i=1}^n$ of i.i.d. stochastic variables such that

$$X \overset{d}{=} \sum_{i=1}^n \xi_i^{(n)}.$$ 

It is easily verified that any Lévy process has an infinitely divisible distribution. To see this we apply the properties of independent and stationary increments in the following way. Let $\eta = (\eta_t)_{0 \leq t \leq T}$ be a Lévy process with values in $\mathbb{R}^d$. Let $\Pi = (0 = t_0 < t_1 < \cdots < t_n = T)$ be the equidistant partitioning of $[0, T]$, $t \leq T$, where $t_j := jt/n$, $0 \leq j \leq n$, and $n \geq 2$ is an integer. Then

$$\eta_t = \eta_T - \eta_0 = \sum_{j=0}^{n-1} (\eta_{t_{j+1}} - \eta_{t_j}).$$

Let $X_{j+1}^{(n)} := \eta_{t_{j+1}} - \eta_{t_j}$, $0 \leq j \leq n - 1$. Then by the independent and stationary increments of Lévy processes, we find that $(X_{i}^{(n)})_{i=1}^n$ are i.i.d. and we have

$$\eta_t = \sum_{i=1}^n X_i^{(n)}.$$ 

It then follows that any Lévy process has an infinite divisible distribution.

Since all Lévy processes have an infinitely divisible distribution, the distributions that may be assigned to Lévy processes are limited as not all distributions are infinitely divisible.

A useful result which is argued on page 70 in Cont, Tankov [22], is that the characteristic function of a Lévy process $\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$, 

$$\varphi_t(u) = \mathbb{E}[\exp\{i(u, \eta_t)\}], \quad u \in \mathbb{R}^d, \quad 0 \leq t \leq T,$$

is an exponential function.

It is also argued on page 70 in Cont, Tankov [22], that the distribution of any Lévy process $\eta = (\eta_t)_{0 \leq t \leq T}$ is completely determined by the distribution of $\eta_1$. Hence, if one knows the distribution of $\eta_1$, one also knows the distribution of $\eta_t$ for all $0 \leq t \leq T$.

Below we give an alternative proof which mainly takes advantage of the independent and stationary increments of Lévy processes and their right continuity.

The proof of the proposition that follows is based on the idea on page 35 in Sato [21].

**Proposition 3.7.** Let $\eta = (\eta_t)_{0 \leq t \leq T}$ be a Lévy process with values in $\mathbb{R}^d$. Let $\varphi_t(\cdot)$ be the characteristic function of $\eta_t$, $0 \leq t \leq T$. Define the characteristic exponent of $\eta$ by

$$\psi_t(u) := \log\{\varphi_t(u)\}, \quad u \in \mathbb{R}^d, \quad 0 \leq t \leq T.$$ 

Then for all $0 \leq t \leq T$, $\psi_t(u) = t\psi_1(u)$, such that the distribution of $\eta_t$ is completely determined by the distribution of $\eta_1$.
Proof. For any pair \((K, J) \in \mathbb{N} \times \mathbb{N}\), we have (the calculations are shown in Appendix B) that \(\psi_K(u) = J \psi_{K/J}(u), u \in \mathbb{R}^d\), such that \(\psi_K(u) = K \psi_1(u), u \in \mathbb{R}^d\). Hence

\[
\psi_{K/J}(u) = \frac{K}{J} \psi_1(u), \quad u \in \mathbb{R}^d,
\]

such that for any rational \(t \leq T\),

\[
\log\{\varphi_t(u)\} = \psi_t(u) = t \psi_1(u) = t \log\{\varphi_1(u)\}, \quad u \in \mathbb{R}^d. \tag{3.6}
\]

For any non-rational \(t \leq T\), let \((t_n)_{n \in \mathbb{N}}\) be a sequence of strictly decreasing positive rational numbers, such that \(t_n \searrow t\), as \(n \to \infty\). By the right continuity of the paths \(t \mapsto \eta_t\), we have that \(\eta_{t_n} \to \eta_t\), as \(n \to \infty\). Now, note that for any pair \((x, y) \in \mathbb{R}^d \times \mathbb{R}^d\), we have \(|\exp\{i \langle x, y \rangle\}| = 1\). Then by the bounded convergence theorem (see Appendix B), the limit \(\lim_{n \to \infty} \exp\{i \langle u, \eta_{t_n} \rangle\}, u \in \mathbb{R}^d\), exists and

\[
\varphi_t(u) = E[\exp\{i \langle u, \eta_t \rangle\}] = E[\lim_{n \to \infty} \exp\{i \langle u, \eta_{t_n} \rangle\}] = \lim_{n \to \infty} E[\exp\{i \langle u, \eta_{t_n} \rangle\}] = \lim_{n \to \infty} \varphi_{t_n}(u), \quad u \in \mathbb{R}^d,
\]

where the second last equality follows from the bounded convergence theorem. Thus we have

\[
\psi_t(u) = \lim_{n \to \infty} \psi_{t_n}(u) = \lim_{n \to \infty} t_n \psi_1(u) = t \psi_1(u), \quad u \in \mathbb{R}^d.
\]

By the above argument, and the fact that there is a 1–1 relation between the characteristic functions and distributions, the distribution of \(\eta_t, 0 \leq t \leq T\), is completely determined by the distribution of \(\eta_1\). Thereby we have arrived at the desired result. ■

In relation to Proposition 3.7 it can be shown that the characteristic exponent of a Lévy process takes on a specific form, and we arrive at the following theorem.

**Theorem 3.8 (Lévy-Kinchin representation).** Let \(\eta = (\eta_t)_{0 \leq t \leq T}\) be a Lévy process with values in \(\mathbb{R}^d\) and Lévy triplet \((\mu, \sigma, \nu)\). Let \(\varphi_t(\cdot)\) be the characteristic function of \(\eta_t, 0 \leq t \leq T\). Then \(\varphi_t(\cdot)\) has the representation

\[
\varphi_t(u) = \exp\{t \psi(u)\}, \quad 0 \leq t \leq T, \quad u \in \mathbb{R}^d, \tag{3.7}
\]

where for all \(u \in \mathbb{R}^d\),

\[
\psi(u) = -\frac{1}{2} \langle u, \sigma u \rangle + i \langle \mu, u \rangle + \int_{\mathbb{R}^d} (\exp\{i \langle u, z \rangle\} - 1 - i \langle u, z \rangle \mathbf{1}_{\{||z|| \leq 1\}}) \nu(dz). \tag{3.8}
\]

Also, if \((\mu, \sigma, \nu)\) is a Lévy triplet, there exists a Lévy process with the above characteristic exponent.

**Proof.** See e.g. Chapter 2 in Sato [21]. ■
3.1.6 Lévy processes as martingales

When stochastic analysis is applied to practical problems, one often discuss the theory of martingales. We now give some propositions that showcase martingale related properties of Lévy processes in general. The following proposition can be found in Cont, Tankov [22]. Here we also add a proof of the proposition, by rather straightforward manipulation.

**Proposition 3.9.** Let \( \eta = (\eta_t)_{0 \leq t \leq T} \) be a Lévy process with values in \( \mathbb{R} \). Then we have the following:

(i) If for all \( 0 \leq t \leq T \), \( \mathbb{E}[\|\eta_t\|] < \infty \) then the process \( M = (M_t)_{0 \leq t \leq T} \subset \mathbb{R}^d \) given by \( M_t := \eta_t - \mathbb{E}[\eta_t], 0 \leq t \leq T \), is a martingale.

(ii) If for all \( 0 \leq t \leq T \), \( \text{var}[\eta_t] < \infty \), then with \( M \) as defined in (i), we have that the process \( (M_t^2 - \mathbb{E}[(M_t)^2])_{0 \leq t \leq T} \) is a martingale.

**Proof.** See Appendix B for proof.

The following proposition is borrowed from Cont, Tankov [22] (Proposition 3.18).

**Proposition 3.10.** Let \( \eta = (\eta_t)_{0 \leq t \leq T} \) be a Lévy process with values in \( \mathbb{R} \) and Lévy triplet \((\mu,\sigma,\nu)\). Then we have the following:

(i) \( \eta \) is a martingale if and only if
\[
\int_{\mathbb{R}_0} 1_{\{z \geq 1\}} z |\nu(dz)| < \infty \quad \text{and} \quad \sigma + \int_{\mathbb{R}_0} 1_{\{|z| \geq 1\}} z \nu(dz) = 0.
\]

(ii) The process \( (\exp\{\eta_t\})_{0 \leq t \leq T} \) is a martingale if and only if
\[
\int_{\mathbb{R}_0} \exp\{z\} 1_{\{|z| \geq 1\}} |\nu(dz)| < \infty,
\]
and
\[
\frac{1}{2} \sigma + \mu + \int_{\mathbb{R}_0} (\exp\{z\} - 1 - z 1_{\{|z| \leq 1\}}) \nu(dz) = 0.
\]

We finish this section with a proposition on Lévy processes as semimartingales.

**Proposition 3.11.** Every Lévy process is a semimartingale.

**Proof.** See e.g. Proposition 2.7.1 in Applebaum [1].

\[
\blacksquare
\]
Chapter 4

Nonlinear stochastic filtering

Bibliographical notes

The main references of this chapter are Bain, Crisan [2], Crisan [6], Duedahl [7], Mandrekar, Meyer-Brandis, Proske [16], Meyer-Brandis, Proske [17], Xiong [25] and Øksendal [26].

This chapter will give a short introduction to selected parts of the field of nonlinear filtering theory, where we have focused on filtering problems with a jump component in the observation process.

It is by nonlinear filtering techniques that the transition rates we ultimately seek, will be obtained.

4.1 An introduction to stochastic filtering

Consider a random system which evolves over time. In this random system we have some process that we are interested in, a process which we can not observe directly. Though we can not observe this process directly, we are able to collect partial observations of the process; that is, we observe a distorted and noisy version of the actual process itself. This is essentially the base of the stochastic filtering problem. A problem where we have an observed part of a random system, a part we are not interested in by itself, which we wish to use as a stepping stone to find another part of the random system, the part that we are really interested in.

The hidden process will be referred to as the signal process, and the distorted version of the signal will be referred to as the observation process.

We now give a simple example where the focus is on why it can be useful to consider filtering problems.

Example 4.1 (Inaccurate GPS). Consider a woman who has lost her dog and is looking for it. Luckily the dog has a GPS collar, such that the woman may track it with an application on her phone. Unluckily the GPS collar is of poor quality and is very inaccurate. The assumed positioning of the dog shown by the GPS is distorted and shifts around quite a lot. This is a typical filtering problem, where we have
distorted observations with some additional noise and we seek the original signal. Let $X = (X_t)_{t \geq 0}$ be the process which gives the exact coordinates of the dog; that is, $X$ takes values in $\mathbb{R}^2$. Let $Y$ be the process which gives the coordinates shown by the GPS on the phone; that is, the distorted and noisy observations. Now $Y$ also takes values in $\mathbb{R}^2$. Assume $Y$ to be given by some transformation $h(\cdot)$ applied to $X$ and a noisy part. In addition we will choose to model the noise by a 2-dimensional Brownian motion $W = (W_t)_{t \geq 0}$. Then the dynamics of $Y$ are given by

$$dY_t = h(X_t) \, dt + dW_t, \quad t \geq 0.$$ 

Now we may define a filtration $\mathcal{F}^Y_t := \sigma\{Y_s, 0 \leq s \leq t \cup \mathcal{N}\}$ which gives the information (coordinates) collected by the GPS over the time interval $[0, t]$, $t \geq 0$. It is based on this information, the information in $\mathcal{F}^Y_t$, that we wish to estimate the dogs exact coordinates at time $t \geq 0$.

An estimate for the sought after signal in Example 4.1 may be obtained by the means of stochastic filtering techniques.

4.2 Filtering problems

In what follows we consider a complete filtered probability space $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})$, where $T > 0$ is a finite time horizon.

We will in this section, as in the above introduction, consider a signal process $X = (X_t)_{0 \leq t \leq T}$ with values in $\mathbb{R}^d$ and an observation process $Y = (Y_t)_{0 \leq t \leq T}$ with values in $\mathbb{R}^m$. The signal process $X$ is a process which is hidden to us such that we are unable to observe it directly. On the other hand, $Y$ is a process which we observe directly and it provides us with information about the signal $X$. Hence, with $X$ hidden we must rely on $Y$ for information.

In the sequel let $\mathcal{F}^X := (\mathcal{F}^X_t)_{0 \leq t \leq T}$ and $\mathcal{F}^Y := (\mathcal{F}^Y_t)_{0 \leq t \leq T}$ be the natural filtrations of the processes $X$ and $Y$, respectively. In addition, let $\mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}$ be the filtration defined by $\mathcal{F}_t = \mathcal{F}^X_t \vee \mathcal{F}^Y_t$, $0 \leq t \leq T$.

When considering filtering problems we usually want to use the information provided to us by the observation process to do one of the following:

(A) Find the conditional probability that the process $X$ is contained in some set $A$ at a given time $t \leq T$, conditioned on $\mathcal{F}^Y_t$; that is, to find the regular conditional distribution* $\mathbb{P}[X_t \in A | \mathcal{F}^Y_t]$.

(B) Find the best estimate of $X$ at a given time $t \leq T$, given the information in $\mathcal{F}^Y_t$.

We tackle this problem by finding the stochastic variable which gives the best mean square estimate of $X_t$, among all $\mathcal{F}^Y_t$-adapted stochastic variables.

Below we will find that (A) and (B) are actually interconnected.

For (A) we wish to obtain information about the signals probabilistic behaviour; that is, to find a process $\pi = (\pi_t)_{0 \leq t \leq T}$ where $\pi_t$ is the regular conditional probability

*See Appendix B for definition of a regular conditional distribution.
measure for the signal $X_t$, given the information that the observation process provides us with up to time $t \leq T$. Thus we seek the process $\pi = (\pi_t)_{0 \leq t \leq T}$ given by

$$\pi_t(A, \omega) := \mathbb{P}[X_t \in A|\mathcal{F}_t^Y](\omega), \quad 0 \leq t \leq T, \; A \in \mathcal{B}(\mathbb{R}^d).$$

(4.1)

Now we have the following for the process $\pi$.

- For any fixed pair $(t, \omega) \in [0, T] \times \Omega$, the mapping $\mathcal{B}(\mathbb{R}^d) \ni A \mapsto \pi_t(A, \omega)$ is a probability measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.
- For any fixed pair $(t, A) \in [0, T] \times \mathcal{B}(\mathbb{R}^d)$, the mapping $\Omega \ni \omega \mapsto \pi_t(A, \omega)$ is a $\mathcal{F}_t^Y$-measurable stochastic variable.

In the case of (B) we wish for a given time $t \leq T$ to find the best possible estimate $\hat{X}_t$ of the signal $X_t$, based on $\mathcal{F}_t^Y$; that is, given the information that is provided to us by the observation process up to time $t$. This means in mathematical terms that $\hat{X}_t$ is to be $\mathcal{F}_t^Y$-measurable. Here "best estimate" means the best estimate in mean square. Hence we are interested in $(\hat{X}_t)_{0 \leq t \leq T}$ such that\footnote{We use the formulation from pages 86-87 in Øksendal [26].}

$$\mathbb{E}[\|X_t - \hat{X}_t\|^2] = \inf_{\xi \in \mathcal{U}_t} \left\{ \mathbb{E} \left[ \|X_t - \xi\|^2 \right] \right\}, \quad 0 \leq t \leq T,$$

where

$$\mathcal{U}_t := \{ \xi : \Omega \to \mathbb{R}^d \mid \xi \in L^2(\mathbb{P}), \xi \text{ is } \mathcal{F}_t^Y \text{-measurable} \}, \quad 0 \leq t \leq T. \quad (4.2)$$

As before, by $\xi \in L^2(\mathbb{P})$ we mean that

$$\|\xi\|_{L^2(\mathbb{P})} = \left( \mathbb{E} \left[ \|\xi\|^2 \right] \right)^{1/2} < \infty.$$

Now it can be shown that\footnote{See e.g. Theorem 6.1.2 in Øksendal [26].}

$$\mathbb{E} \left[ \|X_t - \mathbb{E}[X_t|\mathcal{F}_t^Y]\|^2 \right] = \inf_{\xi \in \mathcal{U}_t} \left\{ \mathbb{E} \left[ \|X_t - \xi\|^2 \right] \right\}, \quad 0 \leq t \leq T.$$

Hence, for any $t \leq T$, $\hat{X}_t = \mathbb{E}[X_t|\mathcal{F}_t^Y]$ is the best estimate of the signal $X_t$ based on $\mathcal{F}_t^Y$. This result will prove to be very useful as it will allow us to apply Bayes’ theorem (see Theorem 2.2) and then obtain $\hat{X}_t, 0 \leq t \leq T$, by Monte Carlo techniques.

Above we have only considered the case where one is interested in an estimate of $X = (X_t)_{0 \leq t \leq T}$. However, it will not always be the case that one is interested in estimating the pure signal $X$. More generally, one might be interested in a transformed version of $X$; that is, to estimate $f(X) = (f(X_t))_{0 \leq t \leq T}$ for a satisfactory rich enough class of functions $f(\cdot)$. It is then important to be aware of that $f(\mathbb{E}[X_t|\mathcal{F}_t^Y])$ is not the best estimate of $f(X_t), 0 \leq t \leq T$, even though $\mathbb{E}[X_t|\mathcal{F}_t^Y]$ is the best estimate for $X_t, 0 \leq t \leq T$. Thus we can not just obtain a best estimate $\hat{X}_t$ of $X_t$ at a time $t \leq T$ and then transform that estimate with $f(\cdot)$. However, the best estimate of $f(X_t)$ at a time $t \leq T$, based on $\mathcal{F}_t^Y$, is very similar to that of $X_t$. In fact we find that\footnote{See argument on page 84 in Xiong [25].}$\mathbb{E}[f(\hat{X}_t)|\mathcal{F}_t^Y]$ is the best possible estimate for $f(X_t), 0 \leq t \leq T$, based on the information $\mathcal{F}_t^Y$.

We will now see how the process of regular conditional probability measures $\pi$ and the best estimate $\hat{X}$ are interconnected, and arrive at the following lemma.
Lemma 4.1. Let \( \pi = (\pi_t)_{0 \leq t \leq T} \) be the conditional probability measure given by (4.1). Let \( f \in C_b(\mathbb{R}^d; \mathbb{R}) \). Then the optimal filter \( \langle \pi, f \rangle = (\langle \pi_t, f \rangle)_{0 \leq t \leq T} \) may be expressed in the following way:

\[
\langle \pi_t, f \rangle = \mathbb{E}[f(X_t)|\mathcal{F}_t], \quad 0 \leq t \leq T. \tag{4.3}
\]

By the above notation \( \langle \pi_t, f \rangle \) we mean the integral over \( f \) with respect to the measure \( \pi_t \); that is,

\[
\langle \pi_t, f \rangle(\omega) = \int_{\mathbb{R}^d} f(x) \pi_t(dx, \omega), \quad 0 \leq t \leq T.
\]

The above lemma will play a central role in obtaining a representation of the optimal filter \( \langle \pi, f \rangle = (\langle \pi_t, f \rangle)_{0 \leq t \leq T} \), which will in turn enable us to apply Monte Carlo techniques to obtain the optimal filter. This representation will be stated below in Theorem 4.3 and is called the Kallianpur-Striebel formula.

4.3 Formulation of the filtering problem

In most literature to date when a nonlinear filtering problem is considered, the observation process \( Y = (Y_t)_{0 \leq t \leq T} \subset \mathbb{R}^m \) is assumed to follow dynamics of the form

\[
dY_t = h(t, X_t) dt + dB_t^Y, \quad 0 \leq t \leq T,
\]

where \( B^Y = (B_t^Y)_{0 \leq t \leq T} \) is a \( m \)-dimensional Brownian motion and \( h: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^m \) is a continuous mapping.

Hence the distortion of the signal is explained through applying a transformation \( h(\cdot, \cdot) \), and the only type of noise which affects the observations is Gaussian. However, in this thesis we will consider a observation process with more flexible dynamics.

The type of nonlinear filtering problem considered in this thesis will have an observation process which includes a drift component, a Gaussian noise component, and a jump component. Also we will choose our jump component such that we may incorporate a compound Cox process, and hence a jump intensity which depends on \( \omega \in \Omega \). More specifically, we will choose the observation process \( Y \) to be given by what we will refer to as a "generalized Cox process"; that is, the dynamics of \( Y \) will be described by

\[
dY_t = h(t, X_t) dt + dB_t^Y + \int_{\mathbb{R}^m_0} z N_\lambda(t, dz), \quad 0 \leq t \leq T, \tag{4.4}
\]

where

- \( B^Y = (B_t^Y)_{0 \leq t \leq T} \) is a \( m \)-dimensional Brownian motion which is independent of \( N_\lambda(dt, dz), \ 0 \leq t \leq T, \ z \in \mathbb{R}^m_0 \).
- \( h: [0, T] \times \mathbb{R}^d \rightarrow \mathbb{R}^m \) is a continuous mapping.
- \( N_\lambda : [0, T] \otimes \mathcal{B}(\mathbb{R}^m_0) \times \Omega \rightarrow \mathbb{N}_0 \) is a random measure with the predictable compensator

\[
\hat{\mu}(dt, dz, \omega) := \lambda(t, X_t(\omega), z) dt \nu(dz)
\]

\[
= \lambda(t, X_t, z) dt \nu(dz), \quad 0 \leq t \leq T, \ z \in \mathbb{R}^m_0, \tag{4.5}
\]
where $X$ is the signal process with dynamics specified below; $\lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R}_0^m \rightarrow \mathbb{R}_{\geq 0}$ is a continuous mapping; and $\nu$ is a Lévy measure.

- The Lévy measure $\nu$ is assumed to satisfy the following integrability condition:
  \[
  \int_{\mathbb{R}^m_0} \|z\| \nu(dz) < \infty,
  \]
  which implies that the jumps of $Y$ are of finite variation.

For applicational purposes one should note that if $\lambda(t, x, z) = \tilde{\lambda}(t, x), \ 0 \leq t \leq T, \ x \in \mathbb{R}^d$; that is, $\lambda$ does not depend on $z$, then
  \[
  \left( \int_{\mathbb{R}^m_0} z N_\lambda(t, dz) \right)_{0 \leq t \leq T},
  \]
  is a compound Cox process with values in $\mathbb{R}^m$ and intensity $(\tilde{\mu}(t, \omega))_{0 \leq t \leq T}$ given by
  \[
  \tilde{\mu}(t, \omega) := \int_0^t \tilde{\lambda}(s, X_s) \, ds, \quad 0 \leq t \leq T.
  \]

We will get back to this representation of the jump component and provide a more in depth discussion in Chapter 6.

For our signal process we will choose one with dynamics as in the classical non-linear filtering problem; that is, we assume our signal process $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ to evolve according to the dynamics given by
  \[
  dX_t = b(X_t) \, dt + \sigma(X_t) \, dB^X_t, \quad 0 \leq t \leq T, \tag{4.6}
  \]
  where $B^X$ is a $d$--dimensional Brownian motion; $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ and $\sigma : \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$ are continuous mappings; and $X_0 := x^*$ is a stochastic variable independent of $Y$.

**Remark 4.** Later on, in chapter 6, when we state our model for the modeling of stochastic transition rates, we will point out that for some application one might be interested in choosing $b : \mathbb{R}^d \rightarrow \mathbb{R}^d$ as a discontinuous mapping. We do not work with such a discontinuous mapping in this thesis. By work with, we mean that we do not develop an implementation procedure for discontinuous $b(\cdot)$. However, we do discuss the possibilities a discontinuous $b(\cdot)$ would open for.

By describing our observation process $Y$ by the dynamics given in (4.4), we obtain a great deal of flexibility in our model. Since $X$ parametrizes the intensity of our compound Cox process, we may include effects such as regime switching or mean reversion, through the intensity of the compound Cox process; that is, we may modify the random intensity which the frequency of jumps follow. Also, we still include the Gaussian noise through the Brownian motion $B^Y$ as one typically does in the case of the classical nonlinear filtering problem. Hence we have kept the stochastic elements from the classical nonlinear filtering setting. However, we expand the flexibility by adding a jump component which could be a compound Cox process with intensity parametrized by the signal process.
Now it is clear that our signal process can be considered an unknown parametrization process, which parametrizes some of the terms in the dynamics of $Y$. Hence we will sometimes refer to $X$ as the unknown parametrization process.

The following theorem can be found in and is borrowed from, Mandrekar et al. [16], Meyer-Brandis, Proske [17] and Duedahl [7].

**Theorem 4.2** (Strong and unique solution). Let $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ and $Y = (Y_t)_{0 \leq t \leq T} \subset \mathbb{R}^m$ be two stochastic process. Assume that $X$ and $Y$ has dynamics described by the system of stochastic differential equations given by:

$$
\forall 0 \leq t \leq T, \quad \begin{cases}
    dX_t = b(X_t) \, dt + \sigma(X_t) \, dB_t, \\
    dY_t = h(t, X_t) \, dt + dB^Y_t + \int_{\mathbb{R}^m} z N_\lambda(dt, dz),
\end{cases}
$$

(*)

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$ and $h : [0, T] \times \mathbb{R}^d \to \mathbb{R}^m$ are continuous mappings; $B^X = (B^X_t)_{0 \leq t \leq T}$ and $B^Y = (B^Y_t)_{0 \leq t \leq T}$ are $m$- and $d$-dimensional Brownian motions, respectively; $N_\lambda(dt, dz)$, $0 \leq t \leq T$, $z \in \mathbb{R}^m_0$, is a $N_0$-valued random measure with predictable compensator

$$
\hat{\mu}(dt, dz, \omega) = \lambda(t, X_t, z) \, dt \, \nu(dz), \quad 0 \leq t \leq T, \ z \in \mathbb{R}^m_0,
$$

where $\lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}_0$ is a continuous mapping and $\nu$ is a Lévy measure. Now assume that there must exist a constant $C \in (0, \infty)$, such that the coefficients $b : \mathbb{R}^d \to \mathbb{R}^d$, $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$, $h : [0, T] \times \mathbb{R}^d \to \mathbb{R}^m$ and the intensity coefficient $\lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R}^m_0 \to \mathbb{R}$, satisfy the linear growth condition

$$
\forall (t, x, y) \in [0, T] \times \mathbb{R}^d \times \mathbb{R}^d, \quad \|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| + \|h(t, x) - h(t, y)\| \\
+ \int_{\mathbb{R}^m_0} |\lambda(t, x, z) - \lambda(t, y, z)| \nu(dz)
\leq C\|x - y\|,
$$

(4.7)

and the Lipschitz condition

$$
\forall (t, x) \in [0, T] \times \mathbb{R}^d, \quad \|b(x)\| + \|\sigma(x)\| + \|h(t, x)\| + \int_{\mathbb{R}^m_0} |\lambda(t, x, z)| \nu(dz)
\leq C(1 + \|x\|).
$$

(4.8)

Then we have a unique and strong solution to (*).

We have a unique solution for some SDE* if for any two solutions $(X^1_t)_{0 \leq t \leq T}$ and $(X^2_t)_{0 \leq t \leq T}$ to the SDE we have

$$
X^1_t = X^2_t, \quad 0 \leq t \leq T, \text{ a.s.}
$$

A SDE is said to have a strong solution if the solution is a measurable function of the driving noise. For more on this the reader may consult Øksendal [26].

---

* SDE — Abbreviation for stochastic differential equation.
CHAPTER 4. NONLINEAR STOCHASTIC FILTERING

4.4 The optimal filter

Here we will discuss the optimal filter \( \langle \pi, f \rangle = \langle \pi_t, f \rangle \) \( 0 \leq t \leq T \) given by (6.21) and how it can be obtained.

We start with formalizing what we require from the estimate \( \hat{X}_t \), \( 0 \leq t \leq T \). The following three properties (here we follow Crisan [6]) should be satisfied by the best estimate \( \hat{X}_t \), \( 0 \leq t \leq T \).

Casuality The estimate \( \hat{X}_t \) is to be based on \( \mathcal{F}_Y^t \), \( 0 \leq t \leq T \).

Optimality The estimate \( \hat{X}_t \) is to be the best mean square estimate of \( X_t \); that is,

\[
\mathbb{E}[(X_t - \hat{X}_t)^2] = \inf_{\xi \in \mathcal{U}_t} \mathbb{E}[(X_t - \xi)^2], \quad 0 \leq t \leq T, \tag{4.9}
\]

where \( \mathcal{U}_t \), \( 0 \leq t \leq T \) is defined by (4.2).

Online estimation The estimate \( \hat{X}_t \) is to be available at any time \( t \leq T \).

These three conditions should also hold for the optimal filter, \( \langle \pi, f \rangle = \langle \pi_t, f \rangle \) \( 0 \leq t \leq T \), \( f \in \mathcal{C}_b(\mathbb{R}^d; \mathbb{R}) \).

4.4.1 Change of measure

We now introduce our method for solving the nonlinear filtering problem; the change of measure method. Let \( \mathbb{Q} \) be the probability measure on the complete filtered probability space \( (\Omega, \mathcal{F}, \mathbb{F}) \), defined by

\[
\mathbb{Q}[A] = \mathbb{E}[1_A \Lambda_T], \quad A \in \mathcal{F}_T. \tag{4.10}
\]

Here \( \Lambda = (\Lambda_t)_{0 \leq t \leq T} \) is the Radon-Nikodym derivative on \( (\Omega, \mathcal{F}, \mathbb{F}) \) restricted to \( \mathbb{F} \), represented by the density process \( \Lambda_t \)

\[
\Lambda_t = \frac{d\mathbb{Q}}{d\mathbb{P}} \bigg|_{\mathcal{F}_t} = \exp \left\{ - \sum_{i=1}^{m} \int_0^t h_i(s, X_s) \, dB_s^{Y(i)} - \frac{1}{2} \int_0^t \| h(s, X_s) \|^2 \, ds 
- \int_{[0,t] \times \mathbb{R}^m} \log \{ \lambda(s, X_s, z) \} N_\lambda(ds, dz) 
+ \int_{[0,t] \times \mathbb{R}^m} (\lambda(s, X_s, z) - 1) \, ds \, \nu(dz) \right\}, \quad 0 \leq t \leq T. \tag{4.11}
\]

In addition we assume that

\[
\mathbb{E}[\Lambda_T] = 1. \tag{4.12}
\]

\( ^{\text{The density process is borrowed from Mandrekar et al. [16] (see also Meyer-Brandis, Proske [17] and Duedahl [7]).}} \)
Remark 5. The following three conditions are sufficient for (4.12) to hold (by Duedahl [7]):

\[ \sup_{0 \leq t \leq T} \mathbb{E} \left[ \exp \left\{ 6 \int_0^t \| h(s, X_s) \|^2 \, ds \right. \right. \\
+ 4 \int_{[0,t] \times \mathbb{R}^m_0} \{ 1 - \lambda(s, X_s, z)^{-1} \} \lambda(s, X_s, z) \, d\nu(dz) \\
- \int_{[0,t] \times \mathbb{R}^m_0} \{ 1 - \lambda(s, X_s, z)^{-4} \} \lambda(s, X_s, z) \, d\nu(dz) \right\} \right] \\
< \infty \]  
(4.13)

\[ \mathbb{E} \left[ \int_{[0,T] \times \mathbb{R}^m_0} \{ (\lambda(s, X_s, z)^{-4} - 1) \lambda(s, X_s, z) \} \, d\nu(dz) \right] \\
+ \mathbb{E} \left[ \int_0^T \left( \int_{\mathbb{R}^m_0} (\lambda(s, X_s, z) - 1) \, d\nu(dz) \right)^2 \, ds \right] \\
< \infty \]  
(4.14)

\[ \mathbb{E} \left[ \int_0^T \int_{\mathbb{R}^m_0} (\lambda(s, X_s, z) \log \{ \lambda(s, X_s, z) \}) \, d\nu(dz) \right] < \infty \]  
(4.15)

\( \Lambda \) is always a local martingale under \( Q \), and when condition (4.12) holds \( \Lambda \) is a martingale under \( Q \). In addition, \( Q \) and \( P \) are equivalent probability measures on \((\Omega, \mathcal{F}_T)\).

Now by the Girsanov theorem for random measures** we obtain a new representation for the observation process under \( Q \), in the following way: We define \( W = (W_t)_{0 \leq t \leq T} \) to be the process with values in \( \mathbb{R}^d \) and components

\[ W_t^{(i)} := B_t^{(i)} + \int_0^t h_i(s, X_s) \, ds, \quad 0 \leq t \leq T, \quad i = 1, \ldots, d. \]

Then \( W \) is a Brownian motion under the new probability measure \( Q \). In addition we may now under \( Q \) define a new Poisson random measure by

\[ N(dt, dz) := N_{\lambda}(dt, dz), \quad 0 \leq t \leq T, \quad z \in \mathbb{R}^m_0, \]  
(4.16)

where \( N(dt, dz) \) has the intensity measure \( dt \, d\nu(dz) \), \( 0 \leq t \leq T, \quad z \in \mathbb{R}^m_0 \). Then we have a new process \( \eta = (\eta_t)_{0 \leq t \leq T} \) defined by

\[ \eta_t := \int_{[0,t] \times \mathbb{R}^m_0} z N(ds, dz), \quad 0 \leq t \leq T, \]  
(4.17)

which is by our assumptions a finite variation Lévy process under \( Q \). Also, \( \eta \) is independent of the new Brownian motion \( W \) and the signal process \( X \). The Lévy

**See e.g. Jacod, Shiryaev [11].
process $\eta$ is completely described by the Lévy measure $\nu$, in the sense that it has the Lévy triplet $(0, 0, \nu)$.

We may now under $Q$ represent the observation process $Y$ in the following way:

$$ dY_t = dW_t + d\eta_t, \quad 0 \leq t \leq T, $$

where $Y$ is a Lévy process with Lévy triplet $(0, I, \nu)$, where $I \in \mathbb{R}^{m \times m}$ is the identity matrix. Now let $Z = (Z_t)_{0 \leq t \leq T}$ be the following Radon-Nikodym derivative restricted to $\mathcal{F}$:

$$ Z_t := \frac{dP}{dQ} \mid_{\mathcal{F}_t} = \Lambda_t^{-1}, \quad 0 \leq t \leq T, $$

where $\Lambda_t^{-1}$ is the inverse Radon-Nikodym derivative of $\Lambda_t$, $0 \leq t \leq T$ given by (4.18). Then for all $0 \leq t \leq T$, $Z_t$ is under $Q$ given by

$$ Z_t = \exp \left\{ \sum_{i=1}^{m} \int_0^t h_i(s, X_s) \, dB_s^{Y(i)} + \frac{1}{2} \int_0^t \| h(s, X_s) \|^2 \, ds \right\} 
+ \int_{[0,t] \times \mathbb{R}^m_0} \log \{ \lambda(s, X_s, z) \} N(ds, dz) 
+ \int_{[0,t] \times \mathbb{R}^m} \{ 1 - \lambda(s, X_s, z) \} ds \nu(dz) \right\}
= \exp \left\{ \sum_{i=1}^{m} \int_0^t h_i(s, X_s) \, dB_s^{Y(i)} + \int_0^t h_i(u, X_u) \, du - \int_0^t h_i(u, X_u) \, du \right\} 
+ \frac{1}{2} \int_0^t \| h(s, X_s) \|^2 \, ds + \int_{[0,t] \times \mathbb{R}^m_0} \log \{ \lambda(s, X_s, z) \} N(ds, dz) 
+ \int_{[0,t] \times \mathbb{R}^m} \{ 1 - \lambda(s, X_s, z) \} ds \nu(dz) \}
= \exp \left\{ \sum_{i=1}^{m} \int_0^t h_i(s, X_s) \, dW_s^{(i)} - \frac{1}{2} \int_0^t \| h(s, X_s) \|^2 \, ds 
+ \int_{[0,t] \times \mathbb{R}^m_0} \log \{ \lambda(s, X_s, z) \} N(ds, dz) 
+ \int_{[0,t] \times \mathbb{R}^m} \{ 1 - \lambda(s, X_s, z) \} ds \nu(dz) \right\}. \quad (4.18)$$

To summarize, our nonlinear filtering problem which consists of (4.4) and (4.6) takes the following form:

$$ \forall 0 \leq t \leq T, \quad \begin{cases} dX_t = b(X_t) \, dt + \sigma(X_t) \, dB_t^X, \\ dY_t = dW_t + d\eta_t, \end{cases} \quad (4.19) $$

under $Q$. Hence, when we consider our nonlinear filtering problem under $Q$, with a little help of the Girsanov theorem for random measures (see e.g. Jacod, Shiryaev [11]) we have arrived at a formulation where the observation process and the signal
This independence between $X$ and $Y$ is a crucial point when we later on prove Proposition 4.4 in the next section. In turn Proposition 4.4 will allow us to apply Monte Carlo techniques to obtain the optimal filter and provide us with a sort of error estimate.

### 4.4.2 The unnormalized filter and the optimal filter

Building on the above we here introduce our main tool for obtaining the optimal filter, the Kallianpur-Striebel formula. For any $f \in C_b(\mathbb{R}^d; \mathbb{R})$ we define the unnormalized filter $\langle \Psi, f \rangle = (\langle \Psi_t, f \rangle)_{0 \leq t \leq T}$ to be given by

$$
\langle \Psi_t, f \rangle := E^Q[Z_t f(X_t)|\mathcal{F}_t^Y], \quad 0 \leq t \leq T,
$$

(4.20)

where $Z = (Z_t)_{0 \leq t \leq T}$ is the Radon-Nikodym derivative given in (4.18). Why we define $\langle \Psi, \cdot \rangle$ this way, and why we call $\langle \Psi, \cdot \rangle$ the unnormalized filter, will be made clear by Theorem 4.3 which is stated below.

Now we will state a theorem, the Kallianpur-Striebel formula, which later on will prove to be of great importance to us in application.

**Theorem 4.3 (Kallianpur-Striebel formula).** For any function $f \in C_b(\mathbb{R}^d; \mathbb{R})$ the optimal filter $\langle \pi_t, f \rangle_{0 \leq t \leq T}$ may be expressed in the following way:

$$
\langle \pi_t, f \rangle = \frac{\langle \Psi_t, f \rangle}{\langle \Psi_t, 1 \rangle}, \quad 0 \leq t \leq T.
$$

(4.21)

By the division in the Kallianpur-Striebel formula it is reasonable for us to call $\langle \Psi, \cdot \rangle = (\langle \Psi_t, \cdot \rangle)_{0 \leq t \leq T}$ the unnomarilzed filter. Also, the Kallianpur-Striebel formula is really just Bayes’ theorem in a filtering setting.

**Remark 6.** The Kallianpur-Striebel formula holds for any mapping $f \in C(\mathbb{R}^d; \mathbb{R})$ for which the following condition is satisfied:

$$
E[|f(X_t)|] < \infty, \quad 0 \leq t \leq T.
$$

**Proof.** See e.g. page 139 in Bain, Crisan [2].

### 4.5 Numerics

When dealing with nonlinear filtering problems in this thesis, we will apply Monte Carlo techniques with respect to the Kallianpur-Striebel formula. The idea is to perform Monte Carlo simulations to obtain approximations for the two cases of the unnormalized filter,

$$
\langle \Psi_t, f \rangle = E^Q[Z_t f(X_t)|\mathcal{F}_t^Y] \quad \text{and} \quad \langle \Psi_t, 1 \rangle = E^Q[Z_t|\mathcal{F}_t^Y], \quad 0 \leq t \leq T,
$$

for a chosen $f \in C_b(\mathbb{R}^d; \mathbb{R})$. Then we will insert the approximations into the Kallianpur-Striebel formula and obtain an estimate for the optimal filter $\langle \pi_t, f \rangle$, $0 \leq t \leq T$. Here we are considering an infinite dimensional approximation problem since, for $f \in C_b(\mathbb{R}^d; \mathbb{R})$,

$$
\Omega \times [0, T] \ni (\omega, t) \mapsto \langle \pi_t, f \rangle(\omega) \in L^2(\Omega, C([0, T])).
$$
Where the above notation means that the paths $t \mapsto \langle \pi_t, f \rangle(\omega)$ are continuous over the interval $[0, T]$ and, for all $0 \leq t \leq T$, the random variable $\omega \mapsto \langle \pi_t, f \rangle(\omega)$ satisfies the condition

$$\|\langle \pi_t, f \rangle\|_{L^2(\mathbb{P})} = \left\{ \mathbb{E}[|\langle \pi_t, f \rangle|^2] \right\}^{1/2} < \infty.$$  

For us to justify the use of Monte Carlo techniques to obtain the unnormalized filter, we will need Proposition 4.4 given and proven below.

With the Kallianpur-Striebel formula (Theorem 4.3) in mind, the following proposition will allow us to obtain simulated paths of the optimal filter $(\langle \pi_t, f \rangle)_{0 \leq t \leq T}$ by Monte Carlo techniques, in addition to an error estimate.

**Proposition 4.4.** Assume that the functions $b(\cdot), \sigma(\cdot), h(\cdot)$ and $\lambda(\cdot, \cdot, \cdot)$ are all bounded and satisfy (4.13), (4.14), (4.15), (4.7) and (4.8). Let $X^n = (X^n_t)_{0 \leq t \leq T} \subset \mathbb{R}^d, n \in \mathbb{N}$, be a sequence of i.i.d. copies of the solution $X = (X_t)_{0 \leq t \leq T}$ to (4.6). For all $n \in \mathbb{N}$, let $Z^n = (Z^n_t)_{0 \leq t \leq T}$ be the density process (4.18) based on $X^n$. Let $f \in C_b(\mathbb{R}^d; \mathbb{R})$. Then

$$\varphi^n_t(f) := \frac{1}{\ell} \sum_{n=1}^{\ell} Z^n_t f(X^n_t) \quad \text{as} \quad \ell \to \infty$$  

where $(\Psi, \cdot) = (\langle \Psi_t, \cdot \rangle)_{0 \leq t \leq T}$ is the unnormalized filter given by (4.20). Also, there exists a constant $C \in (0, \infty)$ such that

$$\mathbb{E}^Q[|\varphi^n_t(f) - (\Psi_t, f)|^2] \leq \frac{1}{\ell} C \|f\|^2_{\infty, \mathbb{R}^d}, \quad 0 \leq t \leq T, \quad \ell \in \mathbb{N}. \quad (4.23)$$

**Proof.** By the independence of $X$ and $Y$ under $Q$, and since, for all $0 \leq t \leq T$, $Y_t$ is $\mathcal{F}_t^Y$-measurable, we can represent the unnormalized filter $(\Psi, f) = (\langle \Psi_t, f \rangle)_{0 \leq t \leq T}$ as

$$\langle \Psi_t, f \rangle = \mathbb{E}^Q[Z_t f(X_t)|\mathcal{F}_t^Y] = \mathbb{E}^Q_{X_t}[Z_t f(X_t)], \quad 0 \leq t \leq T.$$  

Here $\mathbb{E}^Q_{X_t}[\cdot]$ denotes the expectation with respect to $X_t$, $0 \leq t \leq T$, under the
probability measure $Q$. Now we obtain the following:

$$
\begin{align*}
\mathbb{E}^Q[\{\varphi^\ell_t(f) - \langle \Psi_t, f \rangle \}^2] &= \mathbb{E}^Q[\mathbb{E}^Q[\{\varphi^\ell_t(f) - \langle \Psi_t, f \rangle \}^2 | \mathcal{F}_t^Y]] \\
&= \mathbb{E}^Q_{X_t} \left[ \mathbb{E}^Q_{X_t} \left[ \left\{ \frac{1}{\ell} \sum_{i=1}^{\ell} Z_i^t f(X_i^t) - \frac{1}{\ell} \langle \Psi_t^1, f \rangle \right\}^2 \right] \right] \\
&= \frac{1}{\ell^2} \mathbb{E}^Q_{X_t} \left[ \mathbb{E}^Q_{X_t} \left[ \left( \sum_{i=1}^{\ell} Z_i^t f(X_i^t) - \ell \langle \Psi_t^1, f \rangle \right)^2 \right] \right] \\
&= \frac{1}{\ell^2} \mathbb{E}^Q_{X_t} \left[ \text{var}_{X_t} \left[ \sum_{i=1}^{\ell} Z_i^t f(X_i^t) \right] \right] \\
&= \frac{1}{\ell^2} \mathbb{E}^Q_{X_t} \left[ \mathbb{E}^Q_{X_t} \left[ (\{ Z_i^t f(X_i^t) \})^2 \right] - \left( \mathbb{E}^Q_{X_t} [M_i^1 f(X_i^1)] \right)^2 \right] \\
&\leq \frac{1}{\ell} \mathbb{E}^Q_{X_t} \left[ \mathbb{E}^Q_{X_t} \left[ \{ Z_i^t f(X_i^t) \}^2 \right] \right] \\
&\leq \frac{1}{\ell} \| f \|_{L^2, \mathbb{R}^d} \mathbb{E}^Q[\{ Z_i^t \}^2], \quad 0 \leq t \leq T.
\end{align*}
$$

Since we have assumed for conditions (4.13), (4.14) and (4.15) to hold, it follows that there exists a constant $C \in (0, \infty)$, such that

$$
\mathbb{E}^Q[\{ Z_i^t \}^2] \leq C,
$$

where the size of $C$ will depend on the choice of $h(\cdot, \cdot)$ and $\lambda(\cdot, \cdot, \cdot)$. This proves relation (4.23).

We now proceed to prove relation (4.22). Since $(X^n)_{n \in \mathbb{N}}$ are i.i.d. copies of the solution $X$, and for all $n \in \mathbb{N}$, $Z_i^n$ is the density process (6.15) based on $X^n_t$ it follows from the SLLN that, for all $0 \leq t \leq T$,

$$
\varphi^\ell_t(f) := \frac{1}{\ell} \sum_{n=1}^{\ell} Z_i^n f(X_i^t) \xrightarrow{\ell \to \infty} \mathbb{E}^Q_{X_t} [Z_t f(X_t)] = \mathbb{E}^Q[Z_t f(X_t) | \mathcal{F}_t^Y] =: \langle \Psi_t, f \rangle, \text{ a.s.}
$$

Which is what we wanted to show. 

The $\ell$ in (4.23) gives an error estimate with respect to $L^2(Q)$. An error estimate which tells us how exact our optimal filter really is, and how much we have to gain from increasing the number of paths $\ell \in \mathbb{N}$, to be simulated. In Chapter 6 Proposition 4.4 will be used for obtaining the optimal filter.
Chapter 5

Life insurance

Bibliographical notes

The main reference of this chapter is Koller [12].

In what follows we will give an introduction to concepts of life insurance mathematics; that is, mathematical concepts regarding disability-, life- and pension insurance. This chapter provides an intuition into why transition rates of insured individuals are of great importance for the risk management of life insurance companies. Hence, the following will serve as a motivation for why it is important to have good methods for the modeling of transition rates, before we introduce methods for modeling in the next chapter.

The state space, which contains the states that the insured might find him- or herself in, will often be given by $S = \{*, \dagger\}$, where $*$ and $\dagger$ are the states where the insured is alive and dead, respectively.

Now we give the definition of a life annuity, which is useful to keep in mind when considering risk in the setting of life and pension insurance. Thereafter we give a couple of examples of life insurance contracts, one with a life annuity included.

**Definition 5.1** (Life Annuity). A life annuity is a series of payments, referred to as annuities, payed by an insurer to an insured, from a chosen time of maturity, while the insured is alive.

Typically, the insured either pays a lump sum when the life annuity is issued, or a series of premiums up until the time of maturity. Time of maturity will usually be the date when the policy holder retires.

A life annuity is a good way for an individual to be secured after retirement. It removes the fear of outliving ones assets, as the payments from the life annuity are guaranteed for as long as the insured is alive. The fear of outliving ones assets, certainly becomes ever more realistic as time goes on, considering that the life expectancy of people in general is expected to increase (see e.g. Willets et al. [24] for a thorough discussion on future life expectancy).

Life annuities are good contracts to consider to gain an intuition about risks the insurer must consider, and why it is important with good models for the transition
Two examples of insurance contracts that the insured might get from an insurer, are:

**Example 5.1** (Pension). The insured pays premiums up to a time of maturity, to the insurer. Thereafter from the time when the insured reaches the age of maturity of the pension contract, the insurer has to pay annuities to the insured.

**Example 5.2** (Permanent life insurance). The insured pays premiums to the insurer until the insured’s time of death. When the insured dies the insurer has to pay according to the policy agreement.

There are many different types of contracts to consider in life and pension insurance, varying greatly in complexity. However, Example 5.1 and Example 5.2 along with Definition 5.1, are sufficient for one to understand why the concepts and tools developed in this chapter are of importance to us in life insurance.

Before we continue to discuss risk and transition rates, we establish an important definition when considering risks related to the life span of insured individuals.

**Definition 5.2** (Cohort). In statistics, a cohort is a group of subjects, that experience a shared event over a given period in time.

A cohort could, for example, be the group of people born in London during the years 1999 and 2000. Though, here in this chapter cohorts will be people of who are born in the same year, unless vi specify otherwise. Thus, the cohort of 1995 will be the group of individuals born in 1995.

### 5.1 Risk in life and pension insurance

When considering life insurance contracts or pensions, there are mainly two types of risks, given that we do not consider stochastic interest rates, that the insurance company have to consider:

- Mortality risk: Exposure to unexpected increase in mortality.
- Longevity risk: Exposure to unexpected increase in mortality improvement.

When we write "mortality improvement", we mean that the mortality rate has improved, in the sense that it has been reduced; that is, if we have an improvement in mortality from one year to the next, we have a lower mortality rate in the next year, than we did in the year before.

The two risks are discussed below.

#### 5.1.1 Mortality risk

More deaths than one expects might occur when we have a natural disaster like the 2004 Indian Ocean earthquake and tsunami, which struck countries such as India, Indonesia and Thailand. Another possible cause of dramatic increase in mortality rates, for a specific interval in time, could be an epidemic like the 1918 influenza pandemic.
The two scenarios mentioned above are of course highly unlikely, especially in countries such as the Scandinavian ones, but it is of importance to consider them as they could have catastrophical consequences for the insurer. If an insurance company experiences a lot more deaths among the policy holders, who have contracts as the one in Example 5.2, the insurance company would have to give up a lot more capital, than what they originally expected. Such unexpectedly large mortality rates, among the policy holders, would naturally be quite a burden for the insurer. Depending on the capital requirements that the insurer is subject to, the insurer might not have enough reserves set aside to cover the claims that arise. If in addition to not having set aside enough reserves, the insurer has most of its investments tied up in illiquid assets, the insurer might not even have enough liquid assets to liquidate to cover the expenses. Such a scenario could, in a worst case scenario, even lead to the insurer going insolvent.

5.1.2 Longevity risk

The life span for people constantly evolve, and is of great concern when considering longevity risk. Life expectancy for people has grown substantially, globally, due to matters such as better diets, better hygiene and medical advances, over the last century. How much the life expectancy will increase with in the coming decades, if it will increase at all, is almost impossible to say. We can not be certain about things such as how large the medical advances will be over a future time interval. If large advances in medicine did occur over some future interval in time, to which extent could these advances possibly affect the length of people’s life span? Also, how long will it take before there is made a monumental breakthrough in anti-ageing research, if any is made. Issues such as these, and their effect on life expectancy in the UK, are discussed in greater depths in Willets et al. [24], where the authors discuss the outlook on life expectancy for the 21st century.

Though we might predict substantial increases in mortality improvement due to medical advances, or other similar factors, the life expectancy might not actually increase so much over, say, the next fifty years. Perhaps the government, in the country where the insurer operates, make cuts in funding of social or medical care, such that the quality of care for pensioners deteriorate enough to slow down the mortality improvement in the respective country.

Longevity risk is of great importance to any company making longterm guarantees. If an insurance company underestimates the average life time of the holders of life insurance and pension contracts, they might end up with premiums that are relatively low, compared to the cost incurred by the contracts issued by the insurance company. One could easily see why such mispricing of contracts could have catastrophical consequences, by considering the classical life annuity given in Definition 5.1.

In addition, it is important to note that the decline or increase in mortality improvement, might be different for different age groups. We might have a much higher mortality improvement among individuals aged between 75 and 100, than for those aged between 0 and 25. Hence, perhaps the medical advances only affect either longevity risk or mortality risk, and not both. Also, how important different risks
are to the insurer, and how important these risks are relative to each other, would depend on the country which the insurer operates in.

With all this in mind, it is clear that it is of great importance to life insurance companies to have an accurate estimate of the transition rates of the policy holders and pensioners. If the insurance company does not estimate the transition rates accurately, and end up with having a lot more contract holders transitioning from alive to dead in a given year. There might be quite a bit of hurdles in the way when modeling transition rates, where one of the big ones is data. An insurance company might be able to obtain life tables for the general population of the country it operates in, but this might not reflect the life expectancy of the insured population accurately. Perhaps in the country where the insurance company operates, life- or pension insurance might be considered a luxury good. Then perhaps the mortality rates of the general population is quite different than that of the insured population.

5.1.3 Overestimation versus underestimation of risk

Above, we have only considered underestimation of risk which is certainly the main concern, but it is also disadvantageous to overestimate risk. Clearly the insurer does not want to overestimate the risk and as a consequence ties up to much capital in reserves, capital that could have been invested and earned return for the insurer. Hence the insurer has to consider both the problems tied to underestimation and overestimation of risk. None of the scenarios of overestimation or underestimation are good, but obviously the case of underestimation is the most dangerous of the two, as this one could lead to insolvency, rather than just lower returns which is a possible consequence of overestimation of risk.

5.1.4 Costs

As mentioned earlier, the transition rates are of great importance in connection with pricing of insurance; that is, to calculate the premiums which are to be paid by the insured to the insurer, at some rate. The transition rates are deeply integrated in the pricing formulas typically used to price the insurance contracts. For such pricing formulas, see for example Koller [12]. Also, the transition rates are systematically used to compute the value of an insurance contract; that is, the reserve, the value that the insurance company must set aside for the respective insurance contract. The importance of transition rates are obvious: If one expect 10 claims of 100 000$ and instead aggregates 1000 claims of 100 000$, the insurance company, most likely, have not set aside the correct amount of capital, to cover the losses. The previous is just a simple example to illustrate a point, and typically the capital requirements are very strict. Hence, unless some extreme catastrophe occurs, the insurance company will typically have set aside enough reserves to cover potential losses. However, as discussed earlier, such an underestimation of risk could in a worst case scenario lead to insolvency. For pension contracts, the following illustrates the consequence of underestimation of longevity risk: If the insurance company expects none of its holders of a particular pension contract, to live past 100 years of age, and they end up with half of the contract holders living until the age of 130, they will incur unexpected losses over a time period of 30 years.
The above illustrates the main problems with having estimated transition rates incorrectly. An insurer could end up issuing contracts over a long period of time, without realising the faulty estimation that has been made. Say we have a group of individuals aged around thirty, which an insurance company issues pension contracts to. If the insurance company have underestimated future mortality improvement, they might realise it has performed a faulty estimation within the next twenty years or so, as they observe the mortality improvement year by year. Though they might not realize their mistake before half a century has passed since they issued the first contract which relied on the wrongly estimated transition rates. During which time they could end up issuing many more pension contracts, based on the transition rates which are based on an unrealistic view of future mortality improvement, before they realise that they have made a mistake. In which case, the insurance company would have made guarantees which could generate great losses. Though the insurer will not see the full ramification of the faulty estimation until the insured group has lived out their lives, which might be seventy years from now.

In life and pension insurance, one often assumes the state of the insured to follow a regular Markov chain. Concepts and properties of Markov chains are introduced below.

5.2 Markov Chains

The material developed in this section is based on, and the given definitions are borrowed from, Koller [12].

Definition 5.3 (Markov chain). Let $X = (X_t)_{0 \leq t \leq T}$ be a stochastic process with values in $S$. Let $\Pi = (0 = t_0 < t_1 < \cdots < t_n = T)$ be a partitioning of $[0, T]$. We say that $X$ is a Markov chain if we have, for all states $s_0, s_1, \ldots, s_n \in S$,

$$P[X_{t_0} = s_0, X_{t_1} = s_1, \ldots, X_{t_n} = s_n] > 0,$$

and

$$P[X_{t_n} = s_n | X_{t_0} = s_0, X_{t_1} = s_1, \ldots, X_{t_{n-1}} = s_{n-1}] = P[X_{t_{n+1}} = s_n | X_{t_{n-1}} = s_{n-1}].$$

Hence, in the setting of life insurance, the conditional probability of which state the insured will be in, in the next time step, only depends on the state the insured currently is in. When considering the case of a life insurance contract this is obviously reasonable, as one only considers the states alive and dead.

However, this assumption might not be so reasonable in nonpermanent disability insurance; that is, insurance contracts where we consider the case where the insured could transition back and forth between the state of disabled and active. Now the assumption that the evolution of the state of the insured, follow a Markov chain, is slightly less reasonable. As an example, consider the following: We have two individuals with nonpermanent disability contracts, who are both 60 years old, where one of the individuals has been sick for thirty days and the other has been sick for thirty years; that is, declared too sick to work. It is intuitively much more likely that the individual who has been sick for thirty days will return to work, rather than the
insured who has sick for thirty years. Hence the assumption of the insured following
a Markov chain, in the case where we consider contracts such as disability contracts
where the disability might not be permanent, might be less reasonable than in case
of life insurance contracts as the one considered in Example 5.2.

In the sequel, to make notation less tedious, we shall make use of the following
notation:

\[ p_{ij}(s, t) := \mathbb{P}[X_t = j | X_s = i], \quad 0 \leq s \leq t \leq T, \ i, j \in S. \]

The conditional probability above, denoted by \( p_{ij}(\cdot, \cdot) \), \( i, j \in S \), is what we will refer
to as a transition probability.

**Definition 5.4 (Transition rates).** Let \( X = (X_t)_{0 \leq t \leq T} \) be a Markov chain. We say
that \( X \) is regular, if the following functions are continuous in time and well defined:

\[
\mu_i(t) = \lim_{\Delta t \downarrow 0} \frac{1 - p_{ii}(t, t + \Delta t)}{\Delta t}, \quad i \in S, \tag{5.1}
\]

\[
\mu_{ij}(t) = \lim_{\Delta t \downarrow 0} \frac{p_{ij}(t, t + \Delta t)}{\Delta t}, \quad i \neq j, i, j \in S. \tag{5.2}
\]

In addition we define

\[
\mu_{ii}(t) = -\mu_i(t), \quad i \in S. \tag{5.3}
\]

The functions \( \mu_i(\cdot) \) and \( \mu_{ij}(\cdot) \), are called the transition rates of \( X \).

Let \( X = (X_t)_{0 \leq t \leq T} \) be a regular Markov chain with state space \( S \). A possible
way to interpret the transition rates, is to interpret them as the derivatives of the
transition probabilities. For example, by (5.2), we have for all \( i \neq j, i, j \in S \),

\[
\mu_{ij}(t) \overset{(*)}{=} \lim_{\Delta t \downarrow 0} \frac{p_{ij}(t, t + \Delta t) - p_{ij}(t, t)}{\Delta t} = \left. \frac{d}{dr} p_{ij}(t, r) \right|_{r=t} \tag{5.4}
\]

Where we have used the following:

\( (*) \) Follows from: For all \( i \neq j, i, j \in S \),

\[
p_{ij}(t, t) = \mathbb{P}[X_t = j | X_t = i] = 0, \quad 0 \leq t \leq T.
\]

Another result which proves particularly useful when performing calculations related
to transition rates, is the following:

**Proposition 5.1.** Let \( X = (X_t)_{0 \leq t \leq T} \) be a regular Markov chain with state space \( S \).
Then, for all \( i \in S \), we have

\[
\mu_i(t) = \sum_{\substack{j \in S \\text{ } \\text{ } \text{ } \text{ } \text{ } \text{ } \text{ } j \neq i}} \mu_{ij}(t), \quad 0 \leq t \leq T.
\]
Proof. For all $i \in S$ we have
\[ \mu_i(t) = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} (1 - \mathbb{P}[X_{t+\Delta t} = i|X_t = i]) \]
\[ = \lim_{\Delta t \downarrow 0} \sum_{\substack{j \in S \\setminus \{i\}}} \frac{1}{\Delta t} \mathbb{P}[X_{t+\Delta t} = j|X_t = i] \]
\[ = \sum_{\substack{j \in S \\setminus \{i\}}} \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} p_{ij}(t, t + \Delta t) \]
\[ = \sum_{\substack{j \in S \\setminus \{i\}}} \mu_{ij}(t), \quad 0 \leq t \leq T. \]

A set of equations which ties the transition probabilities together with the transition rates, are the Kolmogorov equations, given by the following result.

**Theorem 5.2 (Kolmogorov equations).** Let $X = (X_t)_{0 \leq t \leq T}$ be a regular Markov chain with values in $S$. Then the following hold:

(i) The backward Kolmogorov equations:
\[
\frac{d}{ds} p_{ij}(s, t) - \mu_i(s)p_{ij}(s, t) = -\sum_{k \neq i} \mu_{ik}(s)p_{kj}(s, t), \quad i, j \in S.
\]

(ii) The forward Kolmogorov equations:
\[
\frac{d}{dt} p_{ij}(s, t) + p_{ij}(s, t)\mu_j(t) = \sum_{k \neq i} p_{ik}(s, t)\mu_{kj}(t), \quad i, j \in S.
\]

Theorem 5.2 shows that if we assume the state of the insured to follow a regular Markov chain, we are able to compute the transition probabilities after having estimated the transition rates. Transition rates can be estimated in different ways, but typically it will be done based on information from a given data set.

The next chapter will give a more thorough discussion on the subject of estimating transition rates, and we will even introduce a new model for such estimation.

The following gives an example of how one can obtain the transition probabilities from the transition rates, in the case of contracts where we only concern ourselves with if the insured is alive or dead.

**Example 5.3 (Transition probability $p_{*\dagger}(\cdot, \cdot)$).** Let $X = (X_t)_{0 \leq t \leq T}$ be a regular Markov chain, with state space $S := \{*, \dagger\}$, where $X_t$ gives the state of the insured at time $0 \leq t \leq T$. And assume that the rate at which the insured transitions from $*$ to $\dagger$, at age $x \geq 0$, is given by
\[ \mu_{*\dagger}(x) = A + B \exp\{\lambda x\}, \]
where $A, B \geq 0$ and $\lambda \in \mathbb{R}$ are constants. Then by the forward Kolmogorov equations in Theorem 5.2, we have the differential equation:

$$
\frac{d}{dt} p^*(s,t) + p^*(s,t) \mu_*(t) = p^†(s,t) \mu^†(t), \quad 0 \leq s \leq t \leq T,
$$

where $\mu^†(t) \equiv 0$, $0 \leq t \leq T$, and by Proposition 5.1 $\mu_*(t) = \mu^†(t)$, $0 \leq t \leq T$. Now we find the following for the above differential equation:

$$
\frac{d}{dt} p^*(s,t) \exp \left\{ \int_s^t \mu_*(u) \, du \right\} = 0 \quad \Leftrightarrow \quad p^*(s,r) \exp \left\{ \int_s^r \mu^†(u) \, du \right\} = 1,
$$

such that we obtain the transition probability

$$
p^*(s,r) = \exp \left\{ - \int_s^r (A + B \exp \{\lambda u\}) \, du \right\} = \exp \{-A(r-s) - B\lambda^{-1}(\exp\{\lambda r\} - \exp\{\lambda s\})\}, \quad 0 \leq s \leq r \leq T.
$$

Here we have used:

$(\ast)$: Multiply both sides of the equation by $\exp \left\{ \int_s^t \mu^†(u) \, du \right\}$.

$(\ast\ast)$: $\int_s^r \cdot \, dt$, $0 \leq s \leq r \leq T$ — on both sides of the equation.

Now we may obtain the transition probability $p^†(\cdot,\cdot)$ by

$$
p^†(s,t) = 1 - p^*(s,t), \quad 0 \leq s \leq t \leq T.
$$

5.3 Transition rates in life insurance

In this section we consider transition rates in disability-, life- and pension insurance. However, we will for the most part consider the types of insurance contracts were we are only concerned with if the insured is alive or dead. This transition rate, the rate at which the insured transitions from alive to dead, will be referred to as the mortality rate. We will sometimes make use of the notation $j \rightarrow k$, which denotes the transition from state $j$ to state $k$.

When we model transition rates between the state where insured is alive and the state where the insured is dead, an important variable is $m_{t,x}$, $t \in \mathbb{T}$, $x \in \mathbb{X}$. Here $\mathbb{T}$ denotes a set of years and $\mathbb{X}$ denotes a set of ages. $m_{t,x}$ is called the central mortality rate in year $t \in \mathbb{T}$ for age $x \in \mathbb{X}$. The central mortality rate $m_{t,x}$ is defined by

$$
m_{t,x} := \frac{d_{t,x}}{L_{t,x}}, \quad x \in \mathbb{X}, \ t \in \mathbb{T},
$$

where,
• \( L_{t,x} \) denotes the number of years lived in the age-interval \([x, x + 1)\), in year \([t, t + 1)\);

• \( d_{t,x} \) denotes the number of persons who dies aged in the interval \([x, x + 1)\), in year \([t, t + 1)\).

By the above definition of \( m_{t,x} \), the central mortality rates must take values in the unit interval \([0, 1]\). When we model our transition rates from data, we will consider the central death rates, as it is obvious from Theorem 5.4 that we can not observe the exact transition rates directly. Hence, central mortality rate is a variable which will prove crucial to us when we attempt to model mortality rates in next chapter.

For the modeling of transition rates, a model must first be chosen for the transition rates, and then the parameters for the respective model must either be chosen or estimated.

The following gives an example of a type of disability insurance, and a diagram where the transitions relevant to the given insurance contract are illustrated.

**Example 5.4 (Disability income insurance).** A disability income insurance, provides benefits to the insured during periods where the insured is unable to work due to illness. Usually the premiums will be paid while the insured is healthy, and annuities will be paid by the insurer while the insured is ill. In such a case we will let the state space be given by \( S = \{\ast, \diamondsuit, \dagger\} \), where \( \ast \), \( \diamondsuit \) and \( \dagger \) are the states where the insured is active, sick and dead, respectively. The following diagram illustrates the transition rates:

\[
\begin{align*}
\ast & \quad \mu_{\ast\diamondsuit}(t, x) \\
\diamondsuit & \quad \mu_{\diamondsuit\ast}(t, x) \\
\dagger & \quad \mu_{\dagger\ast}(t, x) \\
\ast & \quad \mu_{\ast\dagger}(t, x)
\end{align*}
\]

5.3.1 Gompertz-Makeham model

Perhaps the most famous model for modeling transition rates in life insurance, is the Gompertz-Makeham model, or Gompertz-Makeham law of mortality. In the Gompertz-Makeham model one considers the case where the insured is either alive or dead; that is, the state space of the insured is given by \( S = \{\ast, \dagger\} \), such that we are interested in the mortality rate. Gompertz-Makeham, assumes that the transition rate from alive to dead, for an individual aged \( x \), is given by

\[
\mu_{\ast\dagger}(x) = \lambda + \alpha \exp\{\beta x\}, \quad x \geq 0,
\]

where \( \lambda, \alpha, \) and \( \beta \) are reasonable constants, in the sense that we avoid negative transition rates, and transition rates greater than 1.
More generally we introduce the following. We say that the transition rate from state $i$ to $j$, $i,j \in S$, for an individual aged $x$, follows a generalized Gompertz-Makeham model, if the transition rate from state $i$ to $j$, at time $t$, is given by

$$
\mu_{ij}(t,x) = f_{ij}^{1,\ell_1}(t,x) + \exp\{f_{ij}^{2,\ell_2}(t,x)\}, \quad x \geq 0, \quad 0 \leq t \leq T. \quad (5.6)
$$

Above, $f_{ij}^{1,\ell_1}(\cdot,\cdot)$ and $f_{ij}^{2,\ell_2}(\cdot,\cdot)$ are polynomials of degree $\ell_1$ and $\ell_2$, respectively. This generalization of the Gompertz-Makeham model gives rise to a wide variety of possible modifications. One could even let the polynomials have both stochastic coefficients and nonstochastic coefficients. However, perhaps the most important improvement which the generalized model provides, is the inclusion of a dependency on time, and not just age. Hence we could obtain different mortality rates for an individual aged $x$ at time $t < T$ and an individual aged $x$ at time $s < t$. Also the generalization of the Gompertz-Makeham model is not only restricted to mortality rates, but could be modified to be flexible enough for modeling of other transition rates as well.

We give an example of a possible type of generalized Gompertz-Makeham model, where we will not vary the transition rates over time. The case where the transition rates vary over time will be treated thoroughly in the next chapter.

We consider modeling of mortality rates. Let $Y$ be a uniformly distributed stochastic variable, and choose $f_{s*}^{1,\ell_1}(\cdot,\cdot)$, $f_{s*}^{2,\ell_2}(\cdot,\cdot)$, with $\ell_1 = 0$ and $\ell_2 = 1$ in (5.6), to be given by, for all $(t,x) \in [0,T] \times \mathbb{R}_{\geq 0}$,

$$
\begin{align*}
\left\{ 
\begin{array}{ll}
  f_{s*}^{1,\ell_1}(t,x) &= \alpha, \quad \alpha \geq 0, \\
  f_{s*}^{2,\ell_2}(t,x) &= \beta_1 + \beta_2 x + \beta_3 Y, \quad \beta_1,\beta_2,\beta_3 \in \mathbb{R}, \quad Y \sim U(-\epsilon,\epsilon),
\end{array}
\right.
\end{align*}
$$

for a chosen $\epsilon > 0$, where $\alpha, \beta_1, \beta_2$ and $\beta_3$ are constants that one could obtain from data. Now the mortality rate is given by

$$
\mu_{s*}(t,x) = \alpha + \exp\{\beta_1 + \beta_2 x + \beta_3 Y\}, \quad x \geq 0, \quad 0 \leq t \leq T.
$$

To obtain the coefficients $\beta_1, \beta_2$ and $\beta_3$ one could take central death rates from given data, and use software to perform curve fitting to fit $\mathbb{R}_{\geq 0} \ni x \mapsto \mu_{s*}(t,x)$ to the central death rates.

As mentioned above, in the generalized Gompertz-Makeham model we have a model for transition rates that depends both on time and age. This enables us to include stochastic processes that evolve over time to model future uncertainty. Since there is so much uncertainty about the future transition rates for $\star \sim \dagger$, as discussed in Section 5.1, the ability to include stochastic processes as coefficients in the generalized Gompertz-Makeham model is a huge advantage.

In Figure 5.1 a standard Gompertz-Makeham model given by (5.5) is compared to a generalized Gompertz-Makeham model, without stochastic elements, given by

$$
g(x) = \beta_0 + \beta_1 x + \exp\{\beta_2 + \beta_3 x + \beta_4 x^2\}, \quad x \geq 0,
$$

where $\beta_j \in \mathbb{R}$, $j = 0,\ldots,4$, are reasonable constants, such that we avoid negative transition rates and transition rates larger than one. These two models have been fitted by obtaining the coefficients with the curve-fitting tool in MATLAB. The data*

*The data has been obtained from the human mortality database [9].
Figure 5.1: A generalized Gompertz-Makeham given by \( g(x) = \beta_0 + \beta_1 x + \exp(\beta_2 + \beta_3 x + \beta_4 x^2) \), \( \beta_i \in \mathbb{R}, i = 0, \ldots, 4 \), symbolized by "--", and a standard Gompertz-Makeham given by \( f(x) = \lambda + \alpha \exp(\gamma x) \), \( \lambda, \alpha, \gamma \in \mathbb{R} \), symbolized by "—". The models have been compared to central mortality rates of females in Austria, symbolized by "★ ☆", for the years 1947, 1981 and 2007 from left to right, respectively.

that have been used to perform the curve-fitting, are central mortality rates for females in Austria for the years 1947, 1981 and 2007. The results are plotted in Figure 5.1.

In Figure 5.1 we have an almost perfect fit for the generalized Gompertz-Makeham model. For the standard Gompertz-Makeham model the fit is not bad, however it is not as good as the fit we obtained with the generalized Gompertz-Makeham model. This is rather obvious, that we obtain a better fit with higher degree polynomials, but one gets a good view of how big the difference actually is, by considering Figure 5.1. Especially, the standard Gompertz-Makeham model tends to clearly overestimate the mortality rate for individuals aged between, roughly, 60 and 85.

Here in this chapter we have mainly focused on transition rates that do not vary over time, and without stochastic elements. In the chapter that follows we will introduce time-dependent stochastic models for transition rates, and introduce our new model which models the transition rates by means of nonlinear filtering theory.
In addition our model will be continuous in time.
Chapter 6

Stochastic transition rates

In this chapter we start with a discussion on the very popular model presented by Lee and Carter in [14] (henceforth Lee-Carter model), and a suggested extension to the model presented by Wang et al. in [23]. This will strictly be a discussion on what effects the elements included in the two models may capture, and not an in-depth technical discussion on simulation procedures. For a more thorough treatment of the structure of these models the reader may consult Lee, Carter [14] and Wang et al. [23]. However, the models are introduced and we do give a quick overview as to how to implement the model presented in Lee and Carter [14]. We do so mostly to give an understanding of why the models are attractive for practical purposes, and how their construction compare to our model.

Our model is a stochastic model which is continuous in time and non-Gaussian; more specifically, our model will be based on generalized Cox processes. Hence our model includes both drift, Gaussian noise and jumps. The techniques applied when attempting to obtain the transition rates with our model, will be Monte Carlo techniques in relation to nonlinear filtering theory.

By having been introduced to two already widely used models, hopefully one will gain insight into why our model has some very attractive futures, especially in the sense of flexibility, compared to already popular and well established models.

We introduce our model as a whole, discuss the elements incorporated, and discuss which effects can be captured by each component. Much is explained by examples with specific suggestions for the functions that one is to decide on when wanting to capture a specific effect. However, when we proceed to simulate the model, we will mainly focus on the jump component and mean reversion effects through the intensity of the jump component.

As in previous chapters, in the sequel we will consider a complete probability space \((\Omega, \mathcal{F}, \mathbb{P})\) and a finite time horizon \(T\).

In what follows we will denote the state space of the insured by \(S\), where \(S\) consists of all possible states for the insured. We will consider two state spaces. One with respect to permanent disability insurance and one with respect to "pure" life insurance. However, all simulation procedures introduced are simulated with regard to the case of pure life insurance; that is, we consider the state space \(S = \{*, \dagger\}\) for the insured when we simulate our new model. When we discuss the case of permanent disability insurance, we consider the state space \(S = \{*, \diamondsuit, \dagger\}\) for the state of the insured.
6.1 POPULAR MODELS

The transition rate from state \( i \in S \) to state \( j \in S \) for an individual aged \( x > 0 \) at time \( t \leq T \), will be denoted by \( \mu_{ij}(t,x) \). When we model our transition rates we will be using empirical data from the human mortality database [9].

To make notation less cumbersome we introduce the following:

- \( T \) —The set containing the years we have data for, where \( T \) will denote the last year of data.
- \( X \) —The set of all ages in the data, where \( x \) will denote the oldest age, or age group, in the data.

Typically \( x \) will be e.g. \( 110^+ \); that is, the age group of ages \( 110 \) and older, such that \( X := \{0,1,2,\ldots,108,109,110^+\} \).

6.1 Popular models

Here we first give a short introduction to the perhaps most popular model in the modeling of mortality rates, the Lee-Carter model. We give a bit of insight into the properties of the model, its weaknesses and strenghts, as well as a procedure for implementation.

After we have discussed the Lee-Carter model we will give an introduction to, and discuss, the model introduced by Wang et al. in [23].

6.1.1 The Lee-Carter model

This subsection is essentially a summary of Lee, Carter [14], with accompanying discussion. We also give a short walkthrough of a implementation procedure for the model, based on Giacometti et al. [8].

In [14], Lee and Carter introduce an extrapolative method for forecasting of mortality, where they focus on the U.S. population. The following is a summary of this model and includes a discussion on the elements in the model. The method suggested by in Lee, Carter [14] has been discussed in many papers to date*, where weaknesses of the model has been pointed out and possible improvements have been suggested.

Let \( m_{tx}, t = 1,2,\ldots,T, x \in X \), be the central mortality rate as defined in the previous chapter. In the Lee-Carter model the following model is chosen for the central death rates:

\[
\log\{m_{tx}\} = a_x + b_xk_t + \varepsilon_{tx}, \quad x \in X, \ 0 \leq t \leq T, \tag{6.1}
\]

where \( a_x, b_x, x \in X \), and \( k_t, t = 1,2,\ldots,T \), are the parameters of the model, and \( \varepsilon_{tx}, t = 1,2,\ldots,T, x \in X \), are Gaussian stochastic variables with mean zero and variance \( \sigma^2 \).

The parameters have the following interpretation.

- \( a_x \) describes the general mortality for a specific age \( x \in X \).
- \( k_t, t = 1,2,\ldots,T \), gives the trend which the overall mortality follows, relative to time.

*See e.g. Giacometti et al. [8] or Renshaw, Haberman [20]
• $b_x$ describes change in mortality for a specific age $x \in \mathbb{X}$ as a consequence of change in the mortality trend $k_t$; that is, the larger $b_x$ is, for some $x \in \mathbb{X}$, the larger the change in the mortality rate is when there is a change in the trend $k_t$ at time $t \leq T$.

Remark 7. The model (6.1) is invariant. To see this, let $(a, b, k)$ be a solution of (6.1). Then let $c > 0$ be a constant, and put $\tilde{b} = bc$ and $\tilde{k} = k/c$. Now $(a, \tilde{b}, \tilde{k})$ also form a solution of (6.1). Thus we have an infinite amount of maxima in the likelihood function of (6.1).

To make ensure that the model is not unaltered by transformation, Lee and Carter ([14]) impose the following constraints

$$\sum_{x \in \mathbb{X}} b_x = 1 \quad \text{and} \quad \sum_{t \in \mathbb{T}} k_t = 0. \quad (6.2)$$

Fitting the Lee-Carter model

By summing over all years in the data set, we have by the constraint on $(k_t)_{t \in \mathbb{T}}$ in (6.2) that we may, for all $x \in \mathbb{X}$, obtain an estimate $\hat{a}_x$ for $a_x$ in the following way:

$$\sum_{t \in \mathbb{T}} \log \{ m_{t,x} \} = a_x T + b_x \sum_{t \in \mathbb{T}} k_t \implies \hat{a}_x = \frac{1}{T} \sum_{t \in \mathbb{T}} \log \{ m_{t,x} \}, \quad x \in \mathbb{X}. \quad (6.3)$$

In Lee, Carter [14] the authors point out that a good and simple method for obtaining estimates $\hat{b}_x$, $x \in \mathbb{X}$, and $\hat{k}_t$, $t \in \mathbb{T}$, is by singular value decomposition of a matrix consisting of the differences $\log \{ m_{t,x} \} - \hat{a}_x$, $t \in \mathbb{T}$, $x \in \mathbb{X}$.

Singular value decompositions can easily be found by use of software such as e.g. MATLAB. We do not give any explanation of singular value decompositions, for this the reader may consult Lay [13].

By Giacometti et al. [8], the following gives estimates for $b_x$ and $k_t$. We begin with sequences $(m_{t,x})_{x \in \mathbb{X}, t \in \mathbb{T}}$, from given data, consisting of central death rates, and compute $\hat{a}_x$ for $x \in \mathbb{X}$ by (6.3). Then we construct a matrix $M = (\log \{ m_{t,x} \} - \hat{a}_x)_{(t,x) \in \mathbb{T} \times \mathbb{X}}$. Following the steps in Giacometti et al. [8], we first compute the following:

• $u_1 = (u_{1,j})_{j \in \mathbb{T}} \in \mathbb{R}^\mathbb{T}$ — The normalized eigenvector of $MM'$, with respect to the largest eigenvalue of $MM'$.

• $v_1 = (v_{1,j})_{j \in \mathbb{X}} \in \mathbb{R}^\mathbb{X}$ — The normalized eigenvector of $M'M$, with respect to the largest eigenvalue of $M'M$.

Now let $\lambda_1^a$ be the largest eigenvalue of $MM'$. Then we obtain the optimal estimates that satisfy the constraints in (6.2), to be given by

$$\left( \hat{b}_0, \ldots, \hat{b}_x \right)' = \frac{1}{\sum_{j \in \mathbb{X}} v_{1,j}} v_1 \quad \text{and} \quad \left( \hat{k}_1, \ldots, \hat{k}_T \right)' = \lambda_1^a \left( \sum_{j \in \mathbb{X}} v_{1,j} \right) u_1.$$

Hence, when we have obtained estimates for $(a_x)_{x \in \mathbb{X}}$, $(b_x)_{x \in \mathbb{X}}$ and $(k_t)_{t \in \mathbb{T}}$ from data, all we need to find to obtain $(m_{t,x})_{x \in \mathbb{X}}$ for times $t > \mathbb{T}$, is $k_t$. In the Lee-Carter
model one assumes that \( k_t, t = 1, \ldots, T \), follows a random walk with drift; that is, more specifically,

\[
k_t = \theta + k_{t-1} + \xi_t, \quad t = 2, \ldots, T.
\]

Here \( (\xi_t)_{t=2}^T \) is a sequence of stochastic variables that are Gaussian in law with mean zero and finite variance \( \sigma^2_\xi \), and \( \theta \) is a constant. It then follows that \( k_t \sim \mathcal{N}(k_{t-1} + \theta, \sigma^2_\xi), \ t = 2, \ldots, T \). Here we find\(^1\) that the maximum likelihood estimates for \( \theta \) and \( \sigma^2_\xi \) are given by

\[
\hat{\theta} = \frac{k_T - k_1}{T - 1} \quad \text{and} \quad \hat{\sigma}^2 = \frac{1}{T - 1} \sum_{j=1}^{T-1} (k_{j+1} - k_j - \hat{\theta})^2.
\]

Lee-Carter model — concluding remarks

The Lee-Carter model is easy to implement and is widely recognized as one of the best models to date, for the modeling of mortality rates. However, the only way for the model to capture future uncertainty, is by a Gaussian noise terms with mean zero. Hence the Lee-Carter model has no way of capturing cohort effects or events that have short term catastrophic consequences. Suggestions have been made for extensions to the Lee-Carter model, for the capturing of more extreme events and cohort effects. In Giacometti et al. \([8]\) the authors show evidence for the residuals of the Lee-Carter model being non-Gaussian, and obtain better results by fitting the model with a Normal Inverse Gaussian distribution imposed on the residuals.

In the next subsection, Section 6.1.2, we give a brief introduction to the model introduced in Wang et al. \([23]\), where the authors apply heavy-tailed distributions to the error term and also offer a way of capturing cohort effects.

### 6.1.2 A model with non-Gaussian error

This subsection provides a short summary of selected parts of Wang et al. \([23]\), with accompanying discussion.

In Wang et al. \([23]\) the authors introduce a model for mortality rates with non-Gaussian error terms. The following is a summary of this model, and includes a discussion on the elements in the model. The authors consider a Lee-Carter type model, proposed by Renshaw and Haberman in \([20]\), which is intended to capture effects by age, period in time and cohort. Wang et al. extend the model suggested by Renshaw and Haberman by imposing heavy-tailed distributions on the error term. In \([20]\) the authors suggest the following model for the central mortality rates in calendar year \( t \):

\[
\log\{m_{t,x}\} = a_x + b_x k_t + \eta_x \xi_{t-x} + \varepsilon_{t,x}, \quad x \geq 0,
\]

where \( m_{t,x} \) is the central mortality rate as defined before and \( a_x, b_x \) and \( k_t \) have the same interpretation as in the classical Lee-Carter model. To be clear, the calendar year \( t \in \mathbb{N} \) is defined to cover the whole interval \([t, t + 1)\). The term \( k_t \) still follows

\(^1\)The calculations are shown in Appendix B.
the same random walk as in the Lee-Carter model above; that is, for a calendar year \( t \),

\[
k_t = k_{t-1} + \theta + \varepsilon^k_t.
\]  

(6.5)

Since \( t \) is the calendar year and \( x \) is the age of an individual, we may define the year of birth \( c = t - x \). Now the factor \( \xi_c \), which explains cohort effects for the cohort born in year \( c \), is assumed to be given by

\[
\Delta \xi_c = \theta \xi + a \xi (\Delta \xi_{c-1} - \theta \xi) + \sigma \xi \varepsilon^\xi_c, \quad c \in \mathbb{N}.
\]  

(6.6)

Here \((\varepsilon^\xi_c)_{c \in \mathbb{N}_0}\) is a sequence of i.i.d. standard Gaussian stochastic variables. Also, the notation \( \Delta \xi_c \) means \( \xi_{c+1} - \xi_c \). Now the three random terms in the model suggested by Renshaw Haberman are \( \varepsilon_{t,x} \), \( \varepsilon^k_t \) and \( \varepsilon^\xi_c \).

In Wang et al. [23] the authors impose heavy tailed distributions on the error terms, \( \varepsilon_{t,x} \), \( \varepsilon^k_t \) and \( \varepsilon^\xi_c \), in the Renshaw-Haberman model. These heavy tailed distributions are given by a jump diffusion, variance gamma and normal inverse Gaussian, and are imposed in the same way on all three error terms. For more information about the heavy tailed distributions, and a more thorough discussion of the model, the reader may consult Wang et al. [23].

**Lee-Carter extension by Wang et al. — concluding remarks**

Wang et al. suggest an extension to the Lee-Carter model, which offers a way of predicting mortality rates under less "usual" circumstances. The extension models the mortality rates with heavier tails applied to the error terms, than in the classical Lee-Carter model. Hence, the extension offers a way of capturing more extreme events, with possible short term catastrophic consequences. In addition, the model offers a way of capturing cohort effects. These are some of the main futures considered to be weaknesses of the original Lee-Carter model. Hence, in conclusion, the model gives a significant improvement to the Lee-Carter model, in the sense of flexibility, and in the sense of modeling mortality rates under less usual circumstances.

### 6.2 Stochastic transition rates based on generalized Cox processes

In this section we will first give a general introduction to our model with rather general functions, before we in the next section choose specific functions for the purpose of simulations. The structure of the general model will essentially be the same as the general filtering problem suggested in Duedahl [7]. We give a general introduction, before we introduce a new model ready for implementation, to give the reader insight into how flexible the model is and how it can be modified. Then as we go through a procedure for simulations for a specific case of the model, the reader might get an idea of which functions might be relevant to use.

Our model opens for the possibility of capturing both regime switching effects and mean reversion, through an unknown "parametrization process". We structure our

\[\text{‡See also Mandrekar et al. [16].}\]
6.2. STOCHASTIC TRANSITION RATES BASED ON GENERALIZED COX PROCESSES

model as a nonlinear filtering problem, where the observation process will be chosen as a generalized Cox process. Hence, our filtering problem will take on the same form as the one considered in Chapter 4, and we will repeat some of the specifications that were given there.

The strength behind modeling the transition rates stochastically, is that we are taking future uncertainty into account, and since we incorporate jumps into the model we will be able to model big and sudden changes.

In what follows let $T > 0$ be a finite time horizon. Let the process $X = (X_t)_{0 \leq t \leq T}$, with values in $\mathbb{R}^d$, be the hidden signal process and the process $Y = (Y_t)_{0 \leq t \leq T}$, with values in $\mathbb{R}^m$, be the observation process which corresponds to $X$. Also, in the rest of this chapter, let $\mathbb{F}^X := (\mathcal{F}_t^X)_{0 \leq t \leq T}$ and $\mathbb{F}^Y := (\mathcal{F}_t^Y)_{0 \leq t \leq T}$ be the natural filtrations of $X$ and $Y$, respectively, and let $\mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}$ be the filtration given by $\mathcal{F}_t = \mathcal{F}_t^X \vee \mathcal{F}_t^Y$, $0 \leq t \leq T$.

In the sequel we let $\mathcal{S}$ be the state space which consists of the different states of an insured individual. For our model we will assume the stochastic transition rates $(\mu_{ij}(t, x))_{0 \leq t \leq T}$, $i, j \in \mathcal{S}$, $x \geq 0$, to be given by a generalized Gompertz Makeham model

$$\mu_{ij}(t, x) = \alpha_{ij} + h_{ij}^{1,\ell_1}(t, x) + \beta_{ij} \exp \left\{ h_{ij}^{2,\ell_2}(t, x) \right\}, \quad (6.7)$$

where $h_{ij}^{1,\ell_1}(t, x)$ and $h_{ij}^{2,\ell_2}(t, x)$, $0 \leq t \leq T$, $x \geq 0$, are polynomials with stochastic coefficients of degree $\ell_1 \in \mathbb{N}$ and $\ell_2 \in \mathbb{N}$, respectively. Here we choose the polynomials $h_{ij}^{1,\ell_1}(:, :)$ and $h_{ij}^{2,\ell_2}(:, :)$ to be given by

$$h_{ij}^{1,\ell_1}(t, x) = \sum_{k=0}^{\ell_1} Y_t^{(k+1)} x^k, \quad i, j \in \mathcal{S}, \quad (6.8)$$

$$h_{ij}^{2,\ell_2}(t, x) = \sum_{k=0}^{\ell_2} Y_t^{(k+\ell_1+2)} x^k, \quad i, j \in \mathcal{S}. \quad (6.9)$$

Hence the value of our observation process $Y$ is given by the stochastic vector

$$Y_t = (Y_t^{(1)}, \ldots, Y_t^{(\ell_1)}, \ldots, Y_t^{(\ell_1+\ell_2+2)})', \quad 0 \leq t \leq T,$$

such that $m = \ell_1 + \ell_2 + 2$. In addition, for all states $i, j \in \mathcal{S}$, we will require $h_{ij}^{1,\ell_1}(:, :)$ to satisfy the following condition:

$$h_{ij}^{1,\ell_1}(t, x) \geq 0, \quad 0 \leq t \leq T, \quad x \geq 0.$$

This requirement is imposed for us to be sure to avoid negative transition rates. The reason for excluding the possibility of negative transition rates is the following. For a given time interval we can not have a negative amount of individuals transitioning from one state to another, which is what negative transition rates would imply.

Perhaps the simplest and most intuitive way of ensuring non-negative transition rates, is to impose the conditions:

$$\alpha_{ij}, \beta_{ij} \geq 0, \quad i, j \in \mathcal{S}, \quad \text{and} \quad Y_t^{(k)} \geq 0, \quad k = 1, \ldots, \ell_1 + 1, \quad 0 \leq t \leq T.$$

This is also the way we will choose to ensure non-negative transition rates later in this chapter, when we simulate a specific model.
To describe the dynamics of the observation process $Y$ we choose a generalized Cox process; that is, $Y$ is described by the dynamics
\[
dY_t = h(t, X_t) dt + dB_t^Y + \int_{\mathbb{R}_0^m} z N_\lambda(dt, dz), \quad Y_0 = 0, \quad 0 \leq t \leq T,
\] (6.10)
where
- $B^Y = (B^Y_t)_{0 \leq t \leq T}$ is a $m$-dimensional Brownian motion which is independent of the signal process.
- The drift coefficient $h : [0, T] \times \mathbb{R}^d \to \mathbb{R}^m$ is a continuous mapping.
- $N_\lambda(dt, dz)$, $0 \leq t \leq T$, $z \in \mathbb{R}_0^m$ is a $\mathbb{N}_0$-valued random measure, with the following predictable compensator:
  \[
  \hat{\mu}(dt, dz, \omega) := \lambda(t, X_t, z) dt \nu(dz), \quad 0 \leq t \leq T, \quad z \in \mathbb{R}_0^m,
  \]
  where $\nu$ is a Lévy measure and $\lambda : [0, T] \times \mathbb{R}^d \times \mathbb{R}_0^m \to \mathbb{R}_{\geq 0}$ is some function.
- The Lévy measure $\nu$ satisfies the following integrability condition:
  \[
  \int_{\mathbb{R}_0^m} \|z\| \nu(dz) < \infty.
  \]

It is important to note the dependency on the signal process in the predictable compensator. Since the signal process parametrizes the intensity of the jumps of $Y$, we are able to include different effects, such as e.g. regime switching, through the integral with respect to the integer valued random measure $N_\lambda(dt, dz)$.

In our simulation procedure the following will be of great importance: If $\lambda$ does not depend on $z$; that is, $\lambda(t, x, z) = c\check{\lambda}(t, x)$ where $c > 0$ is a constant, and we define
\[
\check{\mu}(t, \omega) := \int_0^t c\check{\lambda}(s, X_s) ds, \quad 0 \leq t \leq T,
\]
then we may write the jump component, in the observation process (6.10), in the following way:
\[
\int_{[0, t] \times \mathbb{R}_0^m} z N_\lambda(ds, dz) = \sum_{j=1}^{N_1(\check{\mu}(t, \omega))} \zeta_j, \quad 0 \leq t \leq T,
\] (6.11)
where, $N_1 = (N_1(t))_{0 \leq t \leq T}$ is a Poisson process with intensity equal to 1 which is independent of $(\zeta_j)_{j \in \mathbb{N}}$ and $\check{\mu}(t, \omega)$, $0 \leq t \leq T$; and $(\zeta_j)_{j \in \mathbb{N}} \subset \mathbb{R}^m$ is a sequence of independent and identically distributed stochastic variables. Also, for all $j \in \mathbb{N}$, the components of $\zeta_j$ are independent. Here the distribution of $\zeta_1$ is derived from the Lévy measure $\nu$; that is, we define the Lévy measure in the following way: For all $A = (A_1, \ldots, A_m) \in \mathcal{B}(\mathbb{R}^m)$,
\[
\nu(A) = \mathbb{P}[\zeta_1 \in A] = \mathbb{P}[\zeta_1^{(1)} \in A_1, \ldots, \zeta_1^{(m)} \in A_m] = \mathbb{P}[\zeta_1^{(1)} \in A_1] \times \cdots \times \mathbb{P}[\zeta_1^{(m)} \in A_m].
\]
6.2. STOCHASTIC TRANSITION RATES BASED ON GENERALIZED COX PROCESSES

Later we will perform a change of measure as in Chapter 4. This change of measure will give us a Poisson random measure $N(dt, dz)$, $0 \leq t \leq T$, $z \in \mathbb{R}^m_0$, for which it will be useful to remember the following representation, from the section on Poisson integration in Chapter 3:

$$
\int_{[0,t] \times \mathbb{R}^m_0} zN(ds, dz) = \sum_{0 \leq s \leq t} \Delta Y_s \mathbf{1}_{\{\Delta Y_s \neq 0\}}, \quad 0 \leq t \leq T.
$$

In the case of our nonlinear filtering problem the jumps will be known from data, so this representation will be utilized when we obtain the optimal filter.

By including $N_\lambda(dt, dz)$ in our model we gain great flexibility. We are now able to model large and sudden changes in the transition rates. Such sudden changes cannot be modeled, without fixing other parameters unrealistically high, without the jump component.

Regarding the choice of functions in the dynamics of the observation process: We will give an example for a possible choice of the function $h(\cdot, \cdot)$. However, we will do this after we have explained the different effects that can be included in the dynamics of the signal process, as this will provide greater understanding for the consequences of different choices of $h(\cdot, \cdot)$. Choices for the other functions; that is, $\lambda(\cdot, \cdot, \cdot)$ and the Lévy measure $\nu$, in the dynamics (6.10), will be discussed in greater detail later in this section.

We will assume for the signal process $X$ to be described by the dynamics

$$
dX_t = b(X_t) dt + \sigma(X_t) dB^X_t, \quad X_0 := x^*, \quad 0 \leq t \leq T,
$$

(6.12)

where $B^X$ is a $d$-dimensional Brownian motion which is independent of $B^Y$; the drift coefficient $b : \mathbb{R}^d \to \mathbb{R}^d$ and the diffusion coefficient $\sigma : \mathbb{R}^d \to \mathbb{R}^{d \times d}$, are both continuous mappings; and the initial value $x^*$ will be a stochastic variable which is independent of the observation process. In addition, the coefficients $b(\cdot)$ and $\sigma(\cdot)$ should be such that we have strong solution to (6.12).

Here there are many possible choices for the coefficients $b(\cdot)$ and $\sigma(\cdot)$. The choice of these coefficients must be made on the grounds of which phenomena one wishes to capture with the model. Perhaps the most interesting choice of coefficient, is the choice of function for $b(\cdot)$. One could for example choose $b(\cdot)$ as a mean reversion coefficient or as a discontinuous mapping to include the possibility of regime switching, or even both. In the case of regime switching, a general form of $b(\cdot)$ is

$$
b(x) = \sum_{i=1}^\ell c_i(x) \mathbf{1}_{\Pi_i}(x), \quad x \in \mathbb{R}^d,
$$

(6.13)

where $\Pi = (\Pi_i)_{i=1}^\ell$ is some partitioning of $\mathbb{R}^d$ and $(c_i(\cdot))_{i=1}^\ell$ are $C(\mathbb{R}^d; \mathbb{R}^d)$ mappings. Here $(c_i(\cdot))_{i=1}^\ell$ represent the effects of the different regimes we might find ourselves in. As for $b(\cdot)$ as a mean reversion coefficient, one could choose

$$
b(x) = \alpha(\beta - x), \quad x \in \mathbb{R}^d,
$$

where $\alpha, \beta \in \mathbb{R}_{\geq 0}^d$ have constant entries and $\beta$ is the stabilization level. Mean reversion will be explained in greater detail below.
These two effects, mean reversion and regime switching, are not mutually exclusive and can be combined. In the case of combining mean reversion with regime switching, one could choose \( b(\cdot) \) as the representation (6.13), with \((c(\cdot))_{i=1}^\ell \) as

\[
c_i(x) = \alpha_i(\beta_i - x), \quad i = 1, \ldots, \ell, \quad x \in \mathbb{R}^d,
\]

where \( \alpha_i, \beta_i \in \mathbb{R}^d_{\geq 0}, \ i = 1, \ldots, \ell \), have constant entries, and \( \beta_i, \ i = 1, \ldots, \ell \), are the stabilization levels in the different regimes.

When we implement our model later on in this chapter, we will mainly concentrate on mean-reversion by use of the Vasicek model, most famous for its application in the modeling of short rates. In the vasicek model, one considers a stochastic differential equation which explains the short rates \( r_t = (r_t)_{0 \leq t \leq T} \subset \mathbb{R} \) by the dynamics

\[
dr_t = \alpha(\beta - r_t) \, dt + \sigma \, dB_t, \quad 0 \leq t \leq T, \tag{6.14}
\]

where \( \alpha, \beta > 0 \) and \( \sigma \) are constants, and \( B \) is a 1-dimensional Brownian motion. A rather simple interpretation of the Vasicek model given by (6.14), is as follows: If \( r_t > \beta \), then the drift is negative at time \( t \leq T \), such that \( r \) will decrease towards \( \beta \) while being affected by noise from the Brownian motion. However, if \( r_t < \beta \), then the drift is positive at time \( t \leq T \), which will lead to \( r \) increasing towards \( \beta \) while being affected by noise from the Brownian motion. The parameter \( \alpha \) gives the speed at which the paths approach the long term expected stabilization level \( \beta \), and \( \sigma \) explains how "volatile" the movement of \( r \) is; that is, how much noise \( r \) is affected by.

The Vasicek model is mean reverting in the sense that, over time \( r \) is expected to stabilize around \( \beta \); that is,

\[
\mathbb{E}[r_t] \longrightarrow \beta \quad \text{as} \quad t \rightarrow \infty, \quad \text{if} \quad \beta \geq r_0,
\]

\[
\mathbb{E}[r_t] \nrightarrow \beta \quad \text{as} \quad t \rightarrow \infty, \quad \text{if} \quad \beta < r_0.
\]

A solution to (6.14) as well as a proof of the limit result above, is given in Appendix B.

Now that we have gone through some of the effects we could include through the signal process \( X \), we return to the choice of \( h(\cdot, \cdot) \) in the dynamics of the observation process.

The choice of the function \( h(\cdot, \cdot) \) will depend very much on the chosen dimension for both the signal process and the observation process. Say, for example, that we choose a signal process with values in \( \mathbb{R}^2 \), and choose \( \ell_1 = 0 \) and \( \ell_2 = 1 \) in (6.8) and (6.9), respectively. Then the observation process takes values in \( \mathbb{R}^3 \) and we could possibly choose \( h : [0, T] \times \mathbb{R}^2 \rightarrow \mathbb{R}^3 \) to be given by

\[
h(t, z) = (0, z_1, z_2)', \quad z = (z_1, z_2) \in \mathbb{R}^2.
\]

By choosing \( h(\cdot, \cdot) \) in such a way, we are able to include effects such as e.g. regime switching from the components of \( X \). Now, effects such as regime-switching or mean-reversion are included much more "directly", than when we included these effects strictly through the intensity of the jump component. The effects will now have a direct effect on the coefficients of the stochastic polynomials in the expression for the transition rate.
6.2. STOCHASTIC TRANSITION RATES BASED ON GENERALIZED COX PROCESSES

The reasoning behind the suggestion given for \( h(\cdot, \cdot) \), above, is that we wish to avoid unreasonable transition rates; that is, to avoid negative transition rates.

When we obtain the optimal filter we will rely heavily on the change of measure method. We remind the reader that (by Mandrekar et al. [16])\(^\text{§}\) the density process \( Z = (Z_t)_{0 \leq t \leq T} \), which will be used in the Kallianpur-Striebel formula, is the Radon-Nikodym derivative restricted to \( \mathcal{F} \),

\[
Z_t = \frac{d\mathbb{P}}{d\mathbb{Q}} \bigg|_{\mathcal{F}_t}, \quad 0 \leq t \leq T,
\]

where \( \mathbb{Q} \) is a probability measure equivalent to \( \mathbb{P} \) on \( (\Omega, \mathcal{F}_T) \). Here, the density process \( Z \) has the following representation:

\[
Z_t = \exp \left\{ \sum_{i=1}^{m} \int_0^t h_i(s, X_s) \, dW_s^{(i)} - \frac{1}{2} \int_0^t \|h(s, X_s)\|^2 \, ds + \int_{[0,t] \times \mathbb{R}_0^m} \log \{\lambda(s, X_s, z)\} N(ds, dz) + \int_{[0,t] \times \mathbb{R}_0^m} (1 - \lambda(s, X_s, z)) \, ds \nu(dz) \right\}, \quad 0 \leq t \leq T, \quad (6.15)
\]

where under \( \mathbb{Q} \),

- \( W = (W_t)_{0 \leq t \leq T} \) is a \( m \)-dimensional Brownian motion;
- \( N(dt, dz), 0 \leq t \leq T, z \in \mathbb{R}_0^m, \) is a Poisson random measure with intensity measure

\[
\mathbb{E}[N(dt, dz)] = dt \nu(dz), \quad 0 \leq t \leq T, z \in \mathbb{R}_0^m.
\]

For the integral with respect to the Poisson random measure \( N(dt, dz) \), inside the exponential of (6.15), we have the following:

\[
\int_{[0,t] \times \mathbb{R}_0^m} \log \{\lambda(s, X_s, z)\} N(ds, dz) = \sum_{0 \leq s \leq t} \log \{\lambda(s, X_s, \Delta Y_s)\} 1_{\{\Delta Y_s \neq 0\}}, \quad 0 \leq t \leq T. \quad (6.16)
\]

In practice we will have obtained the observation process from data. In other words we will have observations \( (Y_{t_j})_{t_j \in \mathbb{T}} \) over the time interval \([0, t^*]\), where \( \mathbb{T} \) is the set of points in time, in the interval \([0, t^*]\), which we have data for. Hence, with respect to the time interval \([0, t^*]\), the sum in (6.16) will be approximated by

\[
\sum_{0 \leq s \leq t^*} \log \{\lambda(s, X_s, \Delta Y_s)\} 1_{\{\Delta Y_s \neq 0\}} \approx \sum_{j:(j,j+1) \in \mathbb{T} \times \mathbb{T}} \log \{\lambda(s, X_j, \Delta Y_j)\} 1_{\{\Delta Y_j \neq 0\}},
\]

\(^\text{§}\) See also, Meyer-Brandis, Proske [17] or Duedahl [7]
where $\Delta Y_j := Y_{j+1} - Y_j$. When handling practical problems one must decide when $\Delta Y_j$ is zero, which might not be as straightforward as one might think. The following must be decided on.

**How much movement is required for us to say that a jump has occurred?**

Here, in this thesis, we will choose to define when jumps occur in the following way: First we will find the average distance between two instances of the observation process, for two following years. Then we say, that if the distance between the two points is a certain amount bigger than the average distance, a jump has occurred. We thereby define the following:

$$\Delta Y := \frac{1}{T} \sum_{j=0}^{T-1} \|\Delta Y_j\|,$$

where $\Delta Y_j := Y_{j+1} - Y_j$, $j = 0, 1, \ldots, T - 1$. Hence $\Delta Y$ is the average difference for $Y$ when having moved one step in time. As jumps are to be sudden large movements, it is reasonable to say that if $\|\Delta Y_j\| < \Delta Y$, a jump did not occur from time $j$ to time $j + 1$, $j = 0, 1, \ldots, T - 1$. One way of choosing a definition that determines when a jump occurs, and the way we have done it in this thesis, is to choose a constant $\varepsilon > 0$, and say that a jump occurs from time $j$ to time $j + 1$, $j = 0, 1, \ldots, T - 1$, if

$$\|\Delta Y_j\| > \Delta Y + \varepsilon \Delta.$$

(6.17)

Now by obtaining the observation process $Y$ from data, we have the approximation for $t^* \leq T$, with $\tilde{T}$ as the set of points in time we have data for up to time $t^*$,

$$\int_{[0,t^*]} \times \mathbb{R}^m \log \{\lambda(s, X_s, z)\} N(ds, dz) \approx \sum_{j: (j, j+1) \in \tilde{T} \times \tilde{T}} \log \{\lambda(k, X_j, \Delta Y_j)\} 1\{\|\Delta Y_j\| > \Delta Y + \varepsilon \Delta\}.$$

(6.19)

To compute the above approximation, we must both decide on the choice of $\varepsilon \Delta$ and also compute $(X_j)_{j \in \mathbb{T}}$. The latter will be obtained by simulating paths $[0, \tilde{T}] \ni t \mapsto X_t$, for a chosen partitioning of $[0, \tilde{T}]$. However, for the partitioning which one has chosen, one might not find a value for $X$ at all time points $j \in \mathbb{T}$, for that partitioning. More specifically, for some time point $j \in \mathbb{T}$, $j$ might fall in between two members of the chosen partitioning. This obstacle will now be discussed in greater detail, where we suggest a possible solution to the problem.

Considering an interval in time $[0, T^*]$, where $T^*$ is a finite time horizon. When simulating paths of a process $X = (X_t)_{0 \leq t \leq T^*} \subset \mathbb{R}^d$ we will use the following: For a chosen number of instances $n \in \mathbb{N}$, let $\Delta t := T^*/n$. In all simulation procedures conducted in this thesis, we use an equidistant partitioning $\Pi = (0 = t_0 < t_1 < \cdots < t_{n-1} < t_n = T^*)$ of the interval we are considering. Here $t_j := j \Delta t$, $0 \leq j \leq n$, such that the distance of the subintervals in $\Pi$ is given by $\|\Pi\| = \Delta t$. Now, say we

\*Notation: $|\Pi| = \sup_{0 \leq j \leq n-1} |t_{j+1} - t_j|$
have simulated a path of $X$, $[0, T^*] \ni t \mapsto X_t$, with the above partitioning, and have obtained $(X_{t_i})_{i=0}^n \subset \mathbb{R}^d$. Then consider the case where we want to obtain the value $X_t$ for a time $t \leq T^*$ that falls in between two of the members of our partitioning $\Pi$. In such a situation we must first identify $t_j$ and $t_{j+1}$, $0 \leq j \leq n - 1$, such that $t_j < t < t_{j+1}$. We do so in the following way: Let $\xi_1$ and $\xi_2$ be the closest whole numbers rounded down to and up to, respectively, of $t/\Delta t$; that is $\xi_1 = \lfloor t/\Delta t \rfloor$ and $\xi_2 = 1 + \lfloor t/\Delta t \rfloor$. Now $\xi_1$ and $\xi_2$ are the numbers $t_j$ and $t_{j+1}$, respectively, such that $t_j < t < t_{j+1}$. Next, in connection with the result that follows, we will use some linear algebra to obtain the value of $X_t$. There exists a scalar $s \in \mathbb{R}$, such that

$$
\begin{pmatrix}
\xi_1 \\
X_x^{(1)} \\
\vdots \\
X_x^{(d)} \\
\xi_1
\end{pmatrix}
+ s 
\begin{pmatrix}
\xi_2 - \xi_1 \\
X_x^{(1)} - X_x^{(1)} \\
\vdots \\
X_x^{(d)} - X_x^{(d)} \\
\xi_1
\end{pmatrix}
= 
\begin{pmatrix}
t \\
X_x^{(1)} \\
\vdots \\
X_x^{(d)}
\end{pmatrix}.
$$

Then by now solving the system of linear equations, we have the value $X_t$ which we were looking for.

The above interpolation procedure will be particularly important when trying to simulate the jump component of the density process $Z$, where the jump component is given by (6.16). This interpolation procedure is also given in Algorithm 6.2.

**Algorithm 6.1 Linear interpolation to obtain $X_t$**

**Input:** Matrix $(X_{t_i}^{(i)})_{0 \leq i \leq d, 0 \leq j \leq n}$; time $t < T$, $t \neq t_j$, $0 \leq j \leq n$

1: $\xi_1 \leftarrow \lfloor \frac{t}{\Delta t} \rfloor$
2: $\xi_2 \leftarrow 1 + \lfloor \frac{t}{\Delta t} \rfloor$
3: $s \leftarrow \frac{t - \xi_1}{\xi_2 - \xi_1}$
4: for $i = 1, \ldots, d$ do
5: $X_t^{(i)} \leftarrow s (X_{\xi_2}^{(i)} - X_{\xi_1}^{(i)})$
6: end for
7: return $(X_t^{(i)})_{i=1}^d$

In the next subsection we will discuss our model in the setting of permanent disability insurance, and how the model can be specified for the purpose of capturing different effects.

### 6.2.1 Transition rates for permanent disability

When considering a permanent disability insurance contract, the state space of the insured is given by $\mathcal{S} = \{*, \diamondsuit, \dag\}$. Then, since the disability would be permanent, we would be considering three transitions; that is, more specifically, the transitions:

$$
* \rightsquigarrow \diamondsuit, \quad * \rightsquigarrow \dag \quad \text{and} \quad \diamondsuit \rightsquigarrow \dag.
$$

Hence we would like to model the transition rates $\mu_{*\diamondsuit}(t, x)$, $\mu_{*\dag}(t, x)$ and $\mu_{\diamondsuit\dag}(t, x)$, $0 \leq t \leq T$, $x \geq 0$.

\(^{\text{Notation: } \lfloor x \rfloor \text{ denotes the integer part of } x \in \mathbb{R}.}\)
Typically the transition rates behave differently in between different states, such that we might have to structure the transition rates differently for each transition. Perhaps one of the transition rates require more flexibility in its structure; that is, higher degrees of polynomials in (6.8) and (6.9), in the case when one chooses generalized Gompertz-Makeham models for the transition rates. In the case of transition rates structured as generalized Gompertz-Makeham models, we could e.g. choose, for all $(t, x) \in [0, T] \times \mathbb{R}_{\geq 0}$,

$$
\mu_{\|}(t, x) = \mu_{\mathcal{O}1}(t, x) = Y_t^{(1)} + A \exp \left\{ Y_t^{(2)} + Y_t^{(3)} x \right\}, \\
\mu_{\mathcal{O}}(t, x) = Y_t^{(4)} + B \exp \left\{ Y_t^{(5)} + Y_t^{(6)} x + Y_t^{(7)} x^2 \right\},
$$

where $Y = (Y_t)_{0 \leq t \leq T} = ((Y_t^{(1)}) \ldots, Y_t^{(7)})_{0 \leq t \leq T}$, is the observation process with values in $\mathbb{R}^7$ and dynamics described by the generalized Cox process (6.10).

For the choice of functions in the filtering problems, one would have to decide which elements one wants to capture for each of the transition rates.

Perhaps in the case of the transition rate $\mu_{\mathcal{O}}(\cdot, \cdot)$ we would be most interested in capturing the effects of regulatory changes in the insurance market. In the model this could be incorporated through the jump component. One way of capturing the effects of regulatory changes, through the jump component, could be to include regime switching effects through the choice of the coefficient $b(\cdot)$ in the signal process.

For the transition rates $\mu_{\|}(\cdot, \cdot)$ and $\mu_{\mathcal{O}1}(\cdot, \cdot)$, perhaps one expects some kind of stabilization in the long run. A possible choice could be a mean reversion model for $X$; that is, to e.g. let $X$ have dynamics described by the Vasicek model given by (6.14).

More precisely we could e.g. structure the dynamics of the signal process $X = (X_t)_{0 \leq t \leq T}$ as, for all $0 \leq t \leq T$,

$$
\begin{align*}
\begin{cases}
   dX_t^{(1)} &= \left\{ 1_{\{X_t^{(1)} < \theta\}} \tilde{b}_1 + 1_{\{X_t^{(1)} \geq \theta\}} \tilde{b}_2 \right\} X_t^{(1)} dt + d\bar{B}_t^{(1)}, & X_0^{(1)} := x_1^* \in \mathbb{R}, \\
   dX_t^{(2)} &= 0, & X_0^{(2)} := \tilde{b}_1 \in \mathbb{R}, \\
   dX_t^{(3)} &= 0, & X_0^{(3)} := \tilde{b}_2 \in \mathbb{R}, \\
   dX_t^{(4)} &= 0, & X_0^{(4)} := \theta > 0, \\
   dX_t^{(5)} &= a_1 (a_2 - X_t^{(4)}) dt + d\bar{B}_t^{(2)}, & X_0^{(5)} := x_5^* \geq 0, \\
   dX_t^{(6)} &= 0, & X_0^{(6)} := a_1 \geq 0, \\
   dX_t^{(7)} &= 0, & X_0^{(7)} := a_2 \geq 0,
\end{cases}
\end{align*}
$$

where $\bar{B} = ((\bar{B}_t^{(1)}, \bar{B}_t^{(2)}))_{0 \leq t \leq T}$ is a 2-dimensional Brownian motion. Here $\theta$ is a threshold, and $\tilde{b}_1$ and $\tilde{b}_2$ represent the effects of different regimes. With these dynamics we could use the dynamics of $X^{(5)}$ to capture the expected long term stabilization for $\mu_{\|}(\cdot, \cdot)$ and $\mu_{\mathcal{O}1}(\cdot, \cdot)$. To capture regime switching effects for $\mu_{\mathcal{O}}(\cdot, \cdot)$, we could use the dynamics of $X^{(1)}$.

Here, in the above dynamics, $\tilde{b}_1$ and $\tilde{b}_2$ could also be $C(\mathbb{R}; \mathbb{R})$ functions, such that we could also capture mean reversion in the dynamics of $X^{(1)}$. We will later in this section give an example of this, where one of the components of the signal process captures regime switching, and mean reversion in the different regimes.
6.2. STOCHASTIC TRANSITION RATES BASED ON GENERALIZED COX PROCESSES

The implementation of a model with mean reversion in the dynamics of the signal process will be discussed in greater detail in the next section, where we will perform simulations for such a model.

6.2.2 Simulating the general model

For simplicity, in the sequel we will only consider the case of a state space $S = \{*, \dagger\}$. Hence we confine ourselves to obtaining the transition rate $\mu_{*\dagger}(\cdot, \cdot)$. However, the general simulation procedure discussed in this section can be applied in similar fashion to obtain the transition rates $\mu_{*\diamond}(\cdot, \cdot)$ and $\mu_{\diamond\dagger}(\cdot, \cdot)$.

Here we will also consider another finite time horizon $\tilde{T}$. Also, we remind the reader that $T$ denotes the last year we have data for. Hence we are considering two time intervals:

- $[0, T]$ — The time interval we have data for.
- $[0, \tilde{T}]$ — The time interval we wish to obtain transition rates for.

Here it is important to note that the time intervals are separate, in the sense that time 0 in one interval is not the same as time 0 in the other interval. We will consider the time interval $[0, \tilde{T}]$ to be some future time interval.

The following diagram illustrates the main steps in the simulation procedure that we will implement to simulate paths $[0, \tilde{T}] \ni t \mapsto \mu_{*\dagger}(t, x)$, for ages $x \geq 0$, of the transition rates.

\[
\begin{array}{c}
\mathcal{F}_T^T \\
\mathcal{X}_T \\
(Y_t)_{0 \leq t \leq \tilde{T}} \\
(\mu_{*\dagger}(t, x))_{0 \leq t \leq \tilde{T}}
\end{array}
\]

All information extracted from the data over the time interval $[0, T]$ is contained in $\mathcal{F}_T^T$, and it is based on this information we obtain the optimal filter $\hat{X}_T$. However, we do not stop there. In our case the ultimate goal is not to obtain the optimal filter, rather this is just a stepping stone to simulate future transition rates. Hence, after we have obtained the best estimate for the signal process at time $T$, we simulate future paths $[0, \tilde{T}] \ni t \mapsto X_t$ of the signal process. Then we may simulate paths $[0, \tilde{T}] \ni t \mapsto Y_t$ of the stochastic coefficients in (6.8) and (6.9). Having simulated paths $t \mapsto Y_t$ for a future time interval, we may in turn obtain paths $t \mapsto \mu_{ij}(t, x)$, $x \geq 0$, of the stochastic transition rates given by the generalized Gompertz-Makeham model (6.7).

Here in this section we will mainly focus on procedures for implementing Monte Carlo techniques to obtain the optimal filter; that is, $\langle \pi_t, f_i \rangle$, $0 \leq t \leq \tilde{T}$, $i = 1, \ldots, d$. Where we will choose to work with the functions

\[ \mathbb{R}^d \ni (x_1, \ldots, x_d) = x \mapsto f_i(x) = x_i, \quad i = 1, \ldots, d, \]

throughout this chapter.

Then we wish to compute the following by Monte Carlo techniques:

\[ \hat{X}_t^{(i)} = \langle \pi_t, f_i \rangle = \frac{\langle \Psi_t, f_i \rangle}{\langle \Psi_t, 1 \rangle}, \quad 0 \leq t \leq T, \ i = 1, \ldots, d. \quad (6.21) \]
Hence, by Proposition 4.4, we wish to obtain the approximations, for a sufficiently large \( \ell \in \mathbb{N} \),
\[
\forall 0 \leq t \leq T, \quad \begin{cases}
(\Psi_t, f_i) \approx \frac{1}{\ell} \sum_{j=1}^{\ell} Z_{jt}^{i} f_i(X_t^{j}), & i = 1, \ldots, d, \\
(\Psi_t, 1) \approx \frac{1}{\ell} \sum_{j=1}^{\ell} Z_{jt}^{d},
\end{cases}
\]  
(6.22)
where \((X_t^{j})_{j=1}^{\ell}, 0 \leq t \leq T\), are \( \ell \) simulated paths \([0, T] \ni t \mapsto X_t\), and \((Z_{jt}^{i})_{j=1}^{\ell} \) are \( \ell \) simulated paths of the density process based on the simulated paths of \( X \).

Before we go into the implementation the following technicalities must be decided on.

(i) Specific functions for the coefficients \( b(\cdot), h(\cdot, \cdot) \) and \( \sigma(\cdot) \).

(ii) Choice of Lévy measure \( \nu \).

(iii) Choice of a priori distributions \((\Theta_i)_{i=1}^{d}\). These are to be generated from to initialize simulations of the paths of the signal process \( X \).

(iv) Choice of the function \( \lambda(\cdot, \cdot, \cdot) \) in the predictable compensator of \( N_{\lambda}(\cdot, \cdot) \).

(v) Definition of a jump; that is, if we for example choose to define a jump by \((6.17)\), we must choose an \( \varepsilon X > 0 \).

The choice of the coefficient \( b(\cdot) \) will rely on what kind of effects one wishes to capture through the signal process; that is, if one wants to capture effects such as e.g. regime switching or mean reversion, or both.

For \( \sigma(\cdot) \) we do not suggest any particular representation to capture specific effects. Rather we suggest to choose a matrix with constant real-valued entries, for \( \sigma(\cdot) \), such that we include the Brownian motion for general noise. Hence a reasonable choice could be a matrix with elements equal to 0 or 1, depending on in which components we want to include the Brownian motion for noise. This is also the way we will define \( \sigma(\cdot) \) when we introduce a specific model and simulate future mortality rates. Hence the main use of \( \sigma(\cdot) \) will be to scale the noise provided by the Brownian motion, and not capture any specific effects.

As mentioned earlier, the function \( h(\cdot, \cdot) \) can be utilized if one wishes to include effects by other means than through the jump component of the observation process.

The Lévy measure \( \nu \) will take the form \( \nu(dz) := \varphi(z)dz \), with \( \varphi(\cdot) \) as the density function of a chosen probability distribution. Here, the density \( \varphi(\cdot) \) will give the distribution of the jump sizes. For \( \varphi(\cdot) \) it could be wise to choose the density function by means of the uniform distribution
\[
U([\delta_1^L, \delta_1^U] \times [\delta_2^L, \delta_2^U] \times \cdots \times [\delta_m^L, \delta_m^U]),
\]
where \((\delta_1^L, \delta_1^U), (\delta_2^L, \delta_2^U), \ldots, (\delta_m^L, \delta_m^U) \in \mathbb{R}^2\), and \( \delta_i^L \geq 0, i = 1, \ldots, \ell_1 + 1 \), with \( \ell_1 \) as in \( h_{s_1^L}^{1, \ell_1}(\cdot, \cdot) \) given by \((6.8)\), to avoid negative transition rates. The reason for choosing a uniform distribution is that it is easier to stay in control of the jump sizes, such that we avoid unrealistically large transition rates and negative transition rates.

For the choice of the a priori distributions \((\Theta_i)_{i=1}^{d}\), one could, for example, choose some type of Gaussian distribution or uniform distribution. Though the choice of \((\Theta_i)_{i=1}^{d}\) will depend on the dynamics chosen for the signal process. However, often
6.2. STOCHASTIC TRANSITION RATES BASED ON GENERALIZED COX PROCESSES

A good place to start with would be some uniform distribution. For some choices of, for example, \( b(\cdot) \), we could end up with problems for certain values of the initial values. With different types of uniform distributions we are able to stay in complete control of the interval where the initial values end up, without any worry of more extreme and possibly problematic values. This will be discussed in further depth, with an accompanying example, in the next section.

When choosing \( \lambda(\cdot,\cdot,\cdot) \) we recommend a "multiplicative" form

\[
\lambda(t,x,z) := c\tilde{\lambda}(t,x), \quad 0 \leq t \leq T, \ x \in \mathbb{R}^d, \ z \in \mathbb{R}_0^m,
\]

with \( c > 0 \) as a constant, and \( [0,T] \times \mathbb{R}^d \ni (t,x) \mapsto \tilde{\lambda}(s,x) \in \mathbb{R}_{\geq 0} \) as surjectiv. We want \( \tilde{\lambda}(\cdot,\cdot) \) to be surjective for us to be able to capture more extreme events. The reason for choosing such a multiplicative form is to remove the computational difficulties of including a dependency on \( z \) in \( \lambda(\cdot,\cdot,\cdot) \). If we have a dependency on \( z \) in the function \( \lambda(\cdot,\cdot,\cdot) \), then we will encounter problems during the simulation phase, when we are attempting to simulate paths of the observation process over future time intervals. For this reason, we will henceforth only consider the representation \( \lambda(s,x,z) := c\tilde{\lambda}(s,x), \ 0 \leq s \leq T, \ x \in \mathbb{R}^d \), where \( c > 0 \) is some real valued constant.

As for defining what a jump is, there is no one right answer, one has to decide that for oneself when implementing the model. In our case we will choose to define jumps as specified in (6.17). Hence, in our case it would probably be wise to consult with experts about what years one could consider to be years where jumps occurred, and then go from there to pick \( \varepsilon_\mathcal{X} \) in (6.17).

When having decided on the above technicalities (i)—(v), we proceed to simulations.

We begin with a simulation procedure for obtaining the optimal filters

\[
\hat{X}^{(i)}_T = \langle \pi_T, f_i \rangle, \quad i = 1, \ldots, d.
\]

**General simulation procedure**

For us to be able to perform simulations to obtain the mortality rates \( \mu_{\ast\dagger}(\cdot,\cdot) \), we need a data set with central mortality rates \( m_{x,t} \) for ages \( x \in X \) and years \( t \in T \).

The simulation procedure that follows is rather general one for obtaining optimal filters. However, we do work with the functions \( f_i(\cdot), \ i = 1, \ldots, d \), as chosen above, in the Kallianpur-Striebel formula. Hence the below simulation procedure helps obtain to the best estimates \( \hat{X}^{(i)}_T \) \( f_j = 1 \), given a matrix \( (m_{t,x})_{t \in T, x \in X} \).

**Step 1. Extract data.** From given data we extract sequences \( (m_{t,x})_{x \in X} \) for each year \( t \in T \).

**Step 2. Obtain the observation process.** With the sequences \( (m_{t,x})_{x \in X}, \ t \in T \), we perform curve fitting. When performing the curve fitting we fit the function \( \mathbb{R}_{\geq 0} \ni x \mapsto \mu_{\ast\dagger}(t,x) \), where \( \mu_{\ast\dagger}(\cdot,\cdot) \) is given by (6.7), to the sequences \( (m_{t,x})_{x \in X}, \) for each \( t \in T \). When the curve fitting has been performed, the coefficients in year \( t \in T \) gives \( Y_t \) for that year. Hence, in this step we have obtained the observation process \( (Y_t)_{t \in T} \).
Step 3. **Generate the initial values.** Generate the initial values \( (x^*_{i,j})_{j=1}^\ell \subset \mathbb{R}^d \) from the a priori distributions \( (\Theta_j)_{j=1}^\ell \); that is, generate
\[
x^*_{i,j} \sim \Theta_i, \quad i = 1, \ldots, d, \ j = 1, \ldots, \ell.
\]

Step 4. **Simulate paths of the signal process.** Simulate \( \ell \in \mathbb{N} \) paths \( [0, T] \ni t \mapsto X^j_t, \ j = 1, \ldots, \ell, \) of the signal process by the Euler scheme given in Algorithm 6.2. Here we initialize the Euler scheme by
\[
X^j_{0,t} := x^*_{i,j}, \quad i = 1, \ldots, d, \ j = 1, \ldots, \ell.
\]

Step 5. **Simulate the density process.** We use the obtained observation process \((Y_t)_{t \in \mathbb{T}}\), and the simulated paths of the signal process, to simulate sequences \((Z^j_T)_{j=1}^\ell\) of the density process \(Z\) given by (6.15). Here we might need to perform linear interpolation to obtain \((X^j_t)_{t \in \mathbb{T}}\), \( j = 1, \ldots, \ell \), to be able to simulate the jump component in the density process. This can be done by Algorithm 6.2.

Step 6. **Obtain the unnormalized filter.** We compute \( \langle \Psi^T, f^j \rangle \) and \( \langle \Psi^T, 1 \rangle \) by the Monte Carlo approximations in (6.22), where the functions \((f^j(\cdot))_{j=1}^d\) are given by
\[
\forall x = (x_1, \ldots, x_d) \in \mathbb{R}^d, \quad f^j(x) = x_j, \quad j = 1, \ldots, d.
\]

Step 7. **Obtain optimal filter.** Then we obtain the value of the best estimate \( \hat{X}^T = (\hat{X}^j_T)_{j=1}^d \) by the Kallianpur-Striebel formula given by (6.21).

---

**Algorithm 6.2** Euler scheme for path of \( X \), with dynamics (6.12).

**Input:** Time horizon \( T \); \( n \) for partitioning; initial values \((x^*_i)^\ell_{i=1} \); coefficients \( b(\cdot), \sigma(\cdot) \).

1. \( \Delta t \leftarrow T/n \)
2. \( X^i_{t_0} \leftarrow x^*_i, \quad i = 1, \ldots, d \)
3. generate \( \xi_i \sim \mathcal{N}(0,1), \quad i = 1, \ldots, n \)
4. for \( i = 1, \ldots, d \) do
5. for \( j = 0, 1, \ldots, n - 1 \) do
6. \( X^i_{t_{j+1}} \leftarrow X^i_{t_j} + b(X^i_{t_j}) \Delta t + \sum_{k=1}^d \sigma_{ik}(X^i_{t_j}) \xi_k \sqrt{\Delta t} \)
7. end for
8. end for
9. return \((X^i_{t_j})_{0 \leq j \leq n, 1 \leq i \leq d}\)

When we wish to simulate future transition rates, typically we will be interested in obtaining the best estimate of some specific components of the signal process. Why we would be particularly interested in obtaining the optimal filter for only some of the components of the signal process, we now explain through an example.
We consider the case of a signal process with dynamics that experience mean reversion and regime switching, jointly. More specifically, we consider a signal process $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^6$ with the dynamics

$$
\forall 0 \leq t \leq T,
\begin{aligned}
\begin{cases}
\frac{dX_t^{(1)}}{dt} = \left\{1_{\{||X_t|| < \theta\}} \tilde{b}_1(X_t) + 1_{\{||X_t|| \geq \theta\}} \tilde{b}_2(X_t) \right\} dt + \tilde{\sigma} dB_t,
\end{cases}
\end{aligned}
$$

where $X_0 := x$ is a stochastic variable that is independent of the observation process; $B$ is a Brownian motion; $X_0^{(6)} := \tilde{\sigma}$ is a constant; $\theta > 0$ is a, constant, chosen threshold; and $\tilde{b}_1 : \mathbb{R}^6 \to \mathbb{R}$ and $\tilde{b}_2 : \mathbb{R}^6 \to \mathbb{R}$ are the mappings

$$
\forall x = (x_1, \ldots, x_6) \in \mathbb{R}^6, \quad \tilde{b}_i(x) = \begin{cases}
 x_3(x_2 - x_1), & \text{if } i = 1, \\
 x_5(x_4 - x_1), & \text{if } i = 2.
\end{cases}
$$

Now we would be interested in obtaining the optimal filters $(\tilde{X}_j^T)^{j=2}_{j=2}$. The reason why we are interested in these particular filters is the following: The filters provide the best estimates for the parameters needed to simulate the paths $t \mapsto X_t^{(1)}$, over the future time interval $[0, \tilde{T}]$ which we wish to obtain transition rates for. Where we initialize the numerical scheme for the paths $[0, \tilde{T}] \ni t \mapsto X_t^{(1)}$, by $X_0^{(1)} \sim \Theta_1$ and $(\tilde{X}_0^j)^{j=2}_{j=2} = (\tilde{X}_j^T)^{j=2}_{j=2}$. Having performed the above simulation procedure, we may simulate future paths $[0, \tilde{T}] \ni t \mapsto X_t$. After we have simulated future paths of $X$, we are able to obtain paths $[0, \tilde{T}] \ni t \mapsto \tilde{\mu}(t, \omega)$ based on the paths of $X$. Then we arrive at the final step for obtaining the future transition rates that we seek. We continue to simulate paths of the observation process $Y$ over a future time interval $[0, \tilde{T}]$. This will in turn give us paths of the transition rates that we ultimately seek. Here $Y$ is given by

$$
Y_t = \int_0^t h(s, X_s) \, ds + B_t^Y + \sum_{j=1}^{N_1(\tilde{\mu}(t, \omega))} \zeta_j, \quad 0 \leq t \leq \tilde{T}.
$$

(6.23)

Since $h(\cdot, \cdot)$ is a continuous mapping and we have already simulated paths $[0, \tilde{T}] \ni t \mapsto X_t$, the integral of $h(\cdot, \cdot)$ in the expression for $Y$, is a deterministic integral. Hence we can for example use the trapezoidal rule to compute this integral. Paths of the Brownian motion can be simulated as before by Algorithm 2.1. For the compound Cox process we have the following for an equidistant partitioning $\Pi = (0 = t_0 < t_1 < \cdots < t_n = \tilde{T})$:

$$
N_1(\tilde{\mu}(t_{j+1}, \omega)) = N_1(\tilde{\mu}(t_j, \omega)) + N_1(\tilde{\mu}(t_{j+1}, \omega)) - N_1(\tilde{\mu}(t_j, \omega)), \quad j = 0, 1, \ldots, n - 1,
$$

where by the stationary increments of the Poisson process,

$$
\{N_1(\tilde{\mu}(t_{j+1}, \omega)) - N_1(\tilde{\mu}(t_j, \omega))\}_{0 \leq t \leq t_{j+1}} = \{N_1(\tilde{\mu}(t_{j+1}, \omega) - \tilde{\mu}(t_j, \omega))\}_{0 \leq t \leq t_{j+1}}, \quad j = 0, 1, \ldots, n - 1,
$$

(6.23)
and by the incremental distribution of \( N_1 \),

\[
N_1(\tilde{\mu}(t_{j+1}, \omega) - \tilde{\mu}(t_j, \omega)) | (X_t)_{t \leq t_{j+1}} \sim \text{Poisson} \left( \int_{t_j}^{t_{j+1}} c\tilde{\lambda}(s, X_s) \, ds \right), \quad j = 0, 1, \ldots, n - 1.
\]

Having already simulated paths \([0, \tilde{T}] \mapsto X_t\), the above integral can easily be computed by e.g. the trapezoidal rule. By the above reasoning, we may now simulate the following sequences:

\[
(\zeta^j_i : 1 \leq i \leq N_1 \left( \int_{t_j}^{t_{j+1}} c\tilde{\lambda}(s, X_s) \, ds \right), \quad j = 0, 1, \ldots, n - 1,
\]

of i.i.d. stochastic variables, where the distribution of \( \zeta_1 \) is derived from the Lévy measure.

Now all is in place for us to obtain paths of the observation process, over future time intervals. A technical walkthrough of how one may simulate these paths, will be given by example in the next subsection, Section 6.3.

After having simulated paths \([0, \tilde{T}] \ni t \mapsto Y_t\) we obtain the mortality rates, based on the future paths of the observation process, given by

\[
[0, \tilde{T}] \ni t \mapsto \mu^*\uparrow(t, x) = \alpha^*\uparrow + \sum_{i=0}^{\ell_1} Y_t^{(i+1)} x^i + \beta^*\uparrow \exp \left( \sum_{i=0}^{\ell_2} Y_t^{(i+\ell_1+2)} x^i \right),
\]

for \( x \geq 0 \). Now we have arrived at the desired transition rates.

We go into greater depths of the simulation procedure in the next section. There we will provide algorithms for simulations of specific cases of the both the compound Cox process in the dynamics of \( Y \), and the density process \( Z \).

### 6.3 Simulations of the model

In this section we introduce and simulate a new model for mortality rates, where we choose an observation process which evolves with the dynamics of a non-Gaussian jump process. Hence, as above we will be considering a state space given by \( S = \{*, \uparrow\}\).

To perform our simulations we will collect our data from the human mortality database [9]. We have chosen to use Norwegian life tables for both genders. The data contains central death rates for the years 1846—2009 for ages 0, 1, 2, \ldots, 110, where the age 110 contains individuals aged 110 and older. For the time interval \([0, \tilde{T}]\) we will let 1846 be the first year and let 2009 be the last year.

Since the state space we are considering only contains the states where the insured is alive or dead, we will only be interested in modeling the mortality rate, since it is impossible to go through the transition \( \uparrow \mapsto * \). Hence we confine ourselves to simulations of the transition rate \( \mu^*\uparrow(\cdot, \cdot) \).

---

**As before, * and \( \uparrow \) are the states where the insured is alive and dead, respectively.**
We will choose to structure our transition rate \( \mu_{\ast}(t, x) \) as the generalized Gompertz-Makeham model given by, for all \((t, x) \in [0, T] \times \mathbb{R}_{\geq 0}, \)

\[
\mu_{\ast}(t, x) = Y_t^{(1)} + Y_t^{(2)} x + \beta \exp \left\{ Y_t^{(3)} + Y_t^{(4)} x + Y_t^{(5)} x^2 \right\}, \tag{6.24}
\]

Hence we are considering a observation process with values in \( \mathbb{R}^5 \). In addition we will impose the condition \( Y_t^{(1)}, Y_t^{(2)} \geq 0, 0 \leq t \leq T \), to ensure that we avoid negative transition rates.

For \( \beta \) we will choose \( \beta = \exp\{-25\} \). The choice of \( \beta \) has been made on the basis of performing curvefitting of \( \mathbb{R}_{\geq 0} \ni x \mapsto \mu_{\ast}(t, x) \) to the data described above, and then looking at the coefficients for the last years. In particular we have looked at the coefficients obtained for the last years of data. Since the coefficients are so sensitive, and with only jump observations, it is very useful to include a constant to have greater control over the behaviour of the future paths \([0, \tilde{T}] \ni t \mapsto Y_t \). Without the non-stochastic constant term, it would be incredibly difficult to obtain realistic transition rates for all ages \( x \geq 0 \), especially with our choice of observation process and the structure of the transition rates. The reason for this is the sensitivity of the transition rates, with respect to the coefficients in (6.8) and (6.9).

Also, we have chosen a second degree polynomial in the exponential in (6.24), due to the comparison in Figure 5.1.

### 6.3.1 The observation process

For the dynamics of the observation process \( Y = (Y_t)_{0 \leq t \leq T} \) with values in \( \mathbb{R}^5 \), we choose to exclude the Brownian motion \( B^\mu \), and we choose \( h(t, x) = 0, 0 \leq t \leq T, x \in \mathbb{R}^d \), such that \( Y \) is given by

\[
dY_t = \int_{\mathbb{R}_0^m} zN_\lambda(dt, dz), \quad 0 \leq t \leq T. \tag{6.25}
\]

Here, for the predictable compensator of \( N_\lambda(dt, dz) \) we will choose

\[
\tilde{\mu}(dt, dz, \omega) := \tilde{\lambda}(t, X_t) dt \nu(dz), \quad 0 \leq t \leq T, z \in \mathbb{R}_0^m,
\]

with

\[
\tilde{\lambda}(t, X_t) := \varepsilon^\lambda \|X_t\|_1, \quad 0 \leq t \leq T, \tag{6.26}
\]

where \( \varepsilon^\lambda > 0 \) is a chosen scalar and \( \| \cdot \|_1 \) is the Manhattan norm; that is, for a vector \( u = (u_1, \ldots, u_d) \in \mathbb{R}^d \),

\[
\|u\|_1 := \sum_{i=1}^d |u_i|.
\]

For practical purposes we will choose a "multiplicative" form of our Lévy measure; that is, \( \nu(dz) = \varphi(z) dz \). More specifically, we choose \( \varphi(\cdot) \) to be a density function by means of a uniform probability distribution

\[
\text{U}([\delta^L_1, \delta^U_1] \times [\delta^L_2, \delta^U_2] \times \cdots \times [\delta^L_5, \delta^U_5]), \tag{6.27}
\]
where $(\delta^L_1, \delta^U_1), (\delta^L_2, \delta^U_2), \ldots, (\delta^L_5, \delta^U_5) \in \mathbb{R}^2$, and $\delta^L_j < \delta^U_j, j = 1, \ldots, 5$. Since $\varphi(\cdot)$ is the density of a probability distribution, we have

$$\int_{\mathbb{R}^m} \varphi(z) dz = 1.$$ 

Now our observation process has the following representation:

$$Y_t = \int \int_{[0,t] \times \mathbb{R}^m} z N_\lambda(ds, dz) = \sum_{i=1}^{N_1(\tilde{\mu}(t, \omega))} \zeta_i,$$

where

- $N_1 = (N_1(t))_{0 \leq t \leq T}$ is a Poisson process, independent of the signal process $X$, with intensity equal to 1.
- The intensity $\tilde{\mu}(t, \omega)$ is given by
  $$\tilde{\mu}(t, \omega) := \int_0^t \tilde{\lambda}(s, X_s) ds, \quad 0 \leq t \leq T.$$
- $(\zeta_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. stochastic variables, where the distribution of $\zeta_1$ is derived from the Lévy measure; that is,
  $$\zeta_1 \sim \text{U}([\delta^L_1, \delta^U_1] \times [\delta^L_2, \delta^U_2] \times \cdots \times [\delta^L_5, \delta^U_5]).$$

Here we see the reason for choosing the density (6.27) as a form of uniform. Now we are able to stay in control of the intervals where our jump sizes occur, and avoid jumps so extreme that we may end up with unreasonable transition rates. As for the choice of $(\delta^L_i, \delta^U_i)_{i=1}^5$, naturally we choose $\delta^L_1, \delta^L_2 \geq 0$ to avoid negative transition rates. It could be good to choose the lower limits $\delta^L_1$ and $\delta^L_2$, to be non-zero to avoid too small transition rates at the lower ages; that is, to choose $\delta^L_1, \delta^L_2 > 0$. For the upper bounds, we will choose $\delta^U_1$ and $\delta^U_2$ quite small to avoid unrealistically large transition rates for the lower ages. For the other limits $(\delta^L_i, \delta^U_i)_{i=3}^5$, one suggestion could be to look at the coefficients obtained from the curve-fitting procedure, and then go from there.

### 6.3.2 The density process

Now the density process $Z = (Z_t)_{0 \leq t \leq T}$ given by (6.15), may in our case be represented in the following way:

$$Z_t = \exp \left\{ \int \int_{[0,t] \times \mathbb{R}^m} \log(\tilde{\lambda}(s, X_s)) N(ds, dz) \ight.$$  

$$\left. + \int \int_{[0,t] \times \mathbb{R}^m} (1 - \tilde{\lambda}(s, X_s)) ds \nu(dz) \right\}, \quad 0 \leq t \leq T. \quad (6.30)$$
Here, by our choice of $\tilde{\lambda}(\cdot, \cdot)$, the integral with respect to the Poisson random measure in the expression (6.30) is given by

$$\int_0^t \int_{\mathbb{R}^n_0} \log\{\tilde{\lambda}(s, X_s)\} N(ds, dz) = \sum_{0 \leq s \leq t} \log\{\varepsilon^\lambda ||X_s||_1\} 1_{\{\Delta Y_s \neq 0\}}, \quad 0 \leq t \leq T. \quad (6.31)$$

We also find that by our choice of Lévy measure, the other integral in (6.30) takes the following form:

$$\int_0^t \int_{\mathbb{R}^n_0} (1 - \tilde{\lambda}(s, X_s)) ds \nu(dz) = \int_0^t (1 - \varepsilon^\lambda ||X_s||_1) ds dz = \int_0^t (1 - \varepsilon^\lambda ||X_s||_1) ds, \quad 0 \leq t \leq T. \quad (6.32)$$

Now with the expressions (6.31) and (6.32) for the integrals inside (6.30), we obtain the following expression for the density process in our model:

$$Z_t = \exp\left\{ \sum_{0 \leq s \leq t} \log\{\varepsilon^\lambda ||X_s||_1\} 1_{\{\Delta Y_s \neq 0\}} + \int_0^t (1 - \varepsilon^\lambda ||X_s||_1) ds, \quad 0 \leq t \leq T. \right\} \quad (6.33)$$

When we perform simulations we will use the following approximation for the integral with respect to the Poisson random measure:

$$\int_0^t \int_{\mathbb{R}^n_0} \log\{\tilde{\lambda}(s, X_s)\} N(ds, dz) \approx \sum_{j:(j, j+1) \in T \times T} \log\{\tilde{\lambda}(s, X_j)\} 1_{\{||\Delta Y_j|| > \Delta X + \varepsilon X\}}.$$

### 6.3.3 The signal process

We choose a signal process $X = (X_t)_{0 \leq t \leq T}$ with values in $\mathbb{R}^3$. For the dynamics of $X$ we will choose to work with the Vasicek model with the functions

$$\forall x \in \mathbb{R}^{3 \times 3}, \quad \sigma_{i,j}(x) = \begin{cases} 1, & \text{if } i = j = 1, \\ 0, & \text{else}, \end{cases} \quad (6.34)$$

$$\forall x \in \mathbb{R}^3, \quad b_i(x) = \begin{cases} x_3(x_2 - x_1), & \text{if } i = 1, \\ 0, & \text{else}. \end{cases} \quad (6.35)$$

Hence the dynamics of the components of $X$ are given by

$$\forall 0 \leq t \leq T, \quad \begin{cases} dX_t^{(1)} = X_t^{(3)}(X_t^{(2)} - X_t^{(1)}) dt + dB_t^{(1)}, \quad X_0^{(1)} := x_1^* \sim \Theta_1, \\ dX_t^{(2)} = 0, \quad X_0^{(2)} := x_2^* \sim \Theta_2, \\ dX_t^{(3)} = 0, \quad X_0^{(3)} := x_3^* \sim \Theta_3. \end{cases} \quad (6.36)$$
Figure 6.1: Two paths of $X^{(1)}$ in (6.36) with an equidistant partitioning $\Pi$ of $[0,20]$ with $|\Pi| = \Delta t = 20/100$. The initial values for the two plots are $(X_0^{1,(1)}, X_0^{1,(2)}, X_0^{1,(3)}) = (-1, -2, -3)$ and $(X_0^{2,(1)}, X_0^{2,(2)}, X_0^{2,(3)}) = (-1, -1, -1)$, from left to right, respectively.

where $(\Theta_i)_{i=1,2,3}$ are a priori distributions that the initial values $(x_i^*)_{i=1,2,3}$ are to be generated from, respectively. We will here choose a uniform distribution for the a priori distributions $(\Theta_i)_{i=1,2,3}$. More specifically, we generate

$$x_1^* \sim U(0, 0.1), \quad x_2^* \sim U(0, 0.1) \quad \text{and} \quad x_3^* \sim U(0, 0.5). \quad (6.37)$$

The reason for choosing positive uniform distributions is that problems arise if we were to choose, for example, a Gaussian distribution for the a priori distributions of the Vasicek parameters. Say, for example, that we choose a priori distributions, such that we generate the following initial values, for two paths of (6.36):

$$X_0^1 := \{x_1^{*,1}, x_2^{*,1}, x_3^{*,1}\} = \{-1, -2, -3\}$$

$$X_0^2 := \{x_1^{*,2}, x_2^{*,2}, x_3^{*,2}\} = \{-1, -1, -1\}.$$

Then we could end up with paths of $X^{(1)}$ illustrated in Figure 6.1. However, as long as the initial values are positive in (6.36) we should obtain a mean-reversion effect. Hence the chosen distributions in (6.37) are unproblematic. Another suggestion could for example be to use gamma distributions for $(\Theta_i)_{i=1,2,3}$. In Figure 6.2, two paths of $X^{(1)}$ are plotted with different a priori distributions for the initial values. The first path $X^{1,(1)}$ has initial values generated as in (6.37), and the second path $X^{2,(2)}$ has initial values generated by

$$X_0^{2,(1)} := x_1^{*,2} \sim \text{Gamma}(1, 2),$$

$$X_0^{2,(2)} := x_2^{*,2} \sim \text{Gamma}(1, 2),$$

$$X_0^{2,(3)} := x_3^{*,2} \sim \text{Gamma}(1, 2). \quad (6.38)$$

Simulation procedure for pure jump observations

Before we simulate the actual transition rates, we must go through a simulation procedure to obtain the Vasicek parameters $\hat{X}^{(2)}_T$ and $\hat{X}^{(3)}_T$. To obtain these estimates
we will go through a simulation procedure as described in the previous section. The procedure is given below, and is essentially the same as in the previous section only more specific.

Step 1. Extract data. We obtain central mortality rates for years $\mathbb{T} := \{0, 1, \ldots, 163\}$ and ages $\mathbb{X} := \{0, 1, \ldots, 110\}$, from the data described earlier in this section.

Step 2. Obtain the observation process. Same as in previous section.

Step 3. Generate initial values. Generate the initial values $(x^{*j})_{j=1}^\ell \subset \mathbb{R}^3$ by

$$
x^{*j}_1 \sim \text{U}(0, 0.1), \quad x^{*j}_2 \sim \text{U}(0, 0.1) \quad \text{and} \quad x^{*j}_3 \sim \text{U}(0, 0.05), \ j = 1, \ldots, \ell.
$$

Step 4. Simulate paths of the Vasicek model. Simulate $\ell$ paths of $[0, T] \ni t \mapsto X^{(1)}_{t, j}, \ j = 1, \ldots, \ell$, of the process $X^{(1)}$ with dynamics given in (6.36). Here we initialize the numerical scheme with

$$
X_0^{i,(j)} := x^{*j}_i, \ i = 1, 2, 3, \ j = 1, \ldots, \ell,
$$

and the paths may be simulated by the Euler scheme in Algorithm 6.2.

Step 5. Simulation of the density process. Simulate the density process at time $T$ by Algorithm 6.3.

Step 6. Compute the unnormalized filter. Compute $(\langle \Psi_T, f_i \rangle)_{i=2,3}$ and $(\langle \Psi_T, 1 \rangle)$ by the Monte Carlo approximations in (6.22). Here the functions $(f_i(\cdot))_{i=2,3}$ are given by

$$
\forall x = (x_1, x_2, x_3), \quad f_i(x) = x_i, \ i = 2, 3.
$$

Step 7. Estimate the Vasicek parameters. Obtain estimates for the Vasicek parameters; that is, the optimal filters $(\hat{X}_{T}^{(i)})_{i=2,3}$ by the Kallianpur-Striebel formula (6.21).
Algorithm 6.3 Compute $Z_T$ given by (6.30).

**Input:** Time horizon $\bar{T}$; $n$ for choice of partitioning; observation process $(Y_t)_{t=0}^{\bar{T}}$; path of signal process $(X_t)_{t=0}^{n}$; function $\tilde{\lambda}(t, x)$; jump threshold $\varepsilon$.  

1. $\Delta t \leftarrow \bar{T}/n$
2. $\Delta Y_j \leftarrow Y_{j+1} - Y_j, \quad j = 0, \ldots, n - 1$
3. $\Delta Y \leftarrow \frac{1}{n} \sum_{i=0}^{n-1} \| \Delta Y_i \|$  
4. for $j = 0, \ldots, \bar{T} - 1$ do  
5. \quad if $\| \Delta Y_j \| > \Delta Y + \varepsilon X$ then  
6. \qquad $I_{j,1} \leftarrow \log\{\tilde{\lambda}(j, X_j)\}$  
7. \quad end if  
8. end for  
9. $I_1 \leftarrow \sum_{j=0}^{\bar{T}-1} I_{j,1}$  
10. $I_2 \leftarrow \Delta t \left\{ \frac{1}{2} \left( 2 - \tilde{\lambda}(0, X_0) - \tilde{\lambda}(t, X_t) \right) + \sum_{j=1}^{n-1} (1 - \tilde{\lambda}(t_j, X_{t_j})) \right\}$ \quad \text{Trapezoidal rule}  
11. $Z_T \leftarrow \exp\{I_1 + I_2\}$  
12. return $Z_T$

After we have performed the simulation procedure by going through steps 1—7, we will have obtained the Vasicek parameters $\hat{X}_T^{(2)}$ and $\hat{X}_T^{(3)}$. Then, when the parameters have been obtained, we can simulate paths of $[0, \bar{T}] \ni t \mapsto X^{(1)}_t$ by the Vasicek dynamics,

$$dX^{(1)}_t = \hat{X}_T^{(3)} (\hat{X}_T^{(2)} - X^{(1)}_t) \, dt + dB^{X,(1)}_t, \quad X^{(1)}_0 := x^*_1, \quad 0 \leq t \leq \bar{T}. \quad (6.39)$$

Now that we are able to simulate the unknown parametrization process $X$ at a future time $t \leq \bar{T}$, we are able to simulate the model as a whole. With simulated paths of the signal process we can simulate the paths $[0, \bar{T}] \ni t \mapsto Y_t$ of the observation process, and then obtain the stochastic coefficients in the generalized Gompertz-Makeham model (6.24). However, before we can simulate the actual observation process, we must first compute $\tilde{\mu}(t, \omega)$ over the increments in the equidistant partitioning $\Pi = (0 = t_0 < t_1 < \cdots < t_n = \bar{T})$; that is, for all $k = 0, 1, \ldots, n - 1$, we must compute

$$\tilde{\mu}(t_{k+1}, \omega) - \tilde{\mu}(t_k, \omega) = \int_{t_k}^{t_{k+1}} \tilde{\lambda}(s, X_s) \, ds. \quad (6.40)$$

Then we simulate $\ell$ paths $[0, \bar{T}] \ni t \mapsto Y^{(j)}_t, \quad j = 1, \ldots, \ell$, of the observation process $Y$, recursively by

$$Y^{(j)}_{t_{k+1}} = Y^{(j)}_{t_k} + \sum_{i=1}^{N_i(\tilde{\mu}(t_{k+1}, \omega) - \tilde{\mu}(t_k, \omega))} \zeta^{(j,k)}_i, \quad k = 0, 1, \ldots, n - 1, \quad j = 1, 2, \ldots, \ell,$$

where

$$(\zeta^{(j,k)}_i)_{i \in \mathbb{N}} \subset \mathbb{R}^n, \quad k = 0, 1, \ldots, n - 1, \quad j = 1, 2, \ldots, \ell,$$
are sequences of i.i.d. stochastic variables. Here, as specified earlier, by the choice of Lévy measure we have the following:

\[
\zeta_{j,k}^1 \sim U([\delta_{L1}, \delta_{U1}] \times \cdots \times [\delta_{L5}, \delta_{U5}]), \quad k = 0, 1, \ldots, n - 1, \ j = 1, 2, \ldots, \ell.
\]

An algorithm for simulating paths of the observation process \(Y\) is given in Algorithm 6.4.

Since we have chosen a 5-dimensional observation process, with a second degree polynomial in the exponent of (6.24), the coefficients are extremely sensitive. And since we have chosen an observation process which only move by jumps, it is extremely difficult to calibrate the model such that we consistently obtain reasonable transition rates. This is the main reason for including the constant term \(\beta := \exp\{-25\}\) in (6.24).

Having estimated paths of the observation process for a future time interval, we finally obtain paths of future mortality rates given by, for \(j = 1, 2, \ldots, \ell, k = 0, 1, \ldots, n\),

\[
\mu_{\ast}^j(t_k, x) = Y_{t_k}^{j,(1)} + Y_{t_k}^{j,(2)} x + \exp\{-25 + Y_{t_k}^{j,(3)} + Y_{t_k}^{j,(4)} x + Y_{t_k}^{j,(5)} x^2\}, \ x \geq 0.
\]

(6.41)

In the next section we give the results and our choices for the unspecified technicalities.

---

**Algorithm 6.4 Simulation of the path \(t \mapsto Y_t\).**

**Input:** Time horizon \(T\); \(n\) for partitioning; path of signal process \((X_t)_{t=0}^n\); function \(\tilde{\lambda}(t, X_t)\); jump size distributions \((\Theta_j)_{j=1}^m\).

1: \(\Delta t \leftarrow T/n\)
2: \(\tilde{\mu}_0 \leftarrow 0\)
3: for \(j = 0, \ldots, n - 1\) do
4: \(\tilde{\mu}_{t_{j+1}} = \tilde{\mu}_{t_j} + \frac{\Delta t}{2} \left\{ \tilde{\lambda}(t_j, X_{t_j}) + \tilde{\lambda}(t_{j+1}, X_{t_{j+1}}) \right\} \quad \triangleright \text{Trapezoidal rule}\)
5: generate \(N \sim \text{Poisson}\left(\tilde{\mu}_{t_{j+1}} - \tilde{\mu}_{t_j}\right)\)
6: if \(N > 0\) then
7: for \(i = 1, \ldots, m\) do
8: generate \(\zeta_{k}^{j,(i)} \sim \Theta_k, \ k = 1, \ldots, N\)
9: \(Y_{t_{j+1}}^{(i)} \leftarrow Y_{t_j}^{(i)} + \sum_{k=1}^N \zeta_{k}^{j,(i)}\)
10: end for
11: else
12: \(Y_{t_{j+1}}^{(i)} \leftarrow Y_{t_j}^{(i)}, \ i = 1, \ldots, m\)
13: end if
14: end for
15: return \((Y_{t_j}^{(i)})_{0 \leq j \leq n, 1 \leq i \leq m}\)

---

### 6.3.4 Results from simulations

In this subsection we go through the results obtained from the simulation procedure described above. We have used data, described earlier in the section, from the human
mortality database [9]. Also, the complete program we have used to perform all simulations, can be found in Appendix C. For the remaining unspecified technicalities we have chosen:

- \( \tilde{T} = 26 \) — We simulate paths of transition rates for a future time interval \([0, 26]\).
- Length of subintervals in the equidistant partitioning for simulation of all paths: \( \Delta t = 26/n, n = 1000 \).
- Number of paths: \( \ell = 1000 \).
- Number of times \( \ell \) paths have been simulated to obtain the optimal filter: \( K = 100 \).
- Constants:
  - \( \varepsilon^\pi = 0.15 \) — This has been chosen rather arbitrarily and gives 44 jumps in the data.
  - \( \varepsilon^\lambda = 0.00023 \) — This has been chosen to be able to simulate the density process \( Z \); that is, to avoid too extreme values for \( Z \) which could cause computational difficulties.
- Jump size distributions:
  \[
  \begin{align*}
  \zeta^{(1)}_1 &\sim \text{U}(0, 0.5 \times 10^{-4}), \\
  \zeta^{(2)}_1 &\sim \text{U}(10^{-5}, 3 \times 10^{-5}), \\
  \zeta^{(3)}_1 &\sim \text{U}(-1, 1), \\
  \zeta^{(4)}_1 &\sim \text{U}(43055 \times 10^{-6}, 55815 \times 10^{-6}), \\
  \zeta^{(5)}_1 &\sim \text{U}(-1405 \times 10^{-7}, -955 \times 10^{-7}).
  \end{align*}
  \]

These jump sizes are based on the coefficients obtained in the observation process for the last years of data.

When we estimated the Vasicek parameters as the optimal filters \( \hat{X}^{(i)}_T = (\hat{\pi}_T, f_i), i = 2, 3 \), by Monte Carlo techniques applied in relation to the Kallianpur-Striebel formula, the results in Table 6.1 were obtained.

<table>
<thead>
<tr>
<th>Path</th>
<th>( X^{(2)}_T )</th>
<th>( X^{(3)}_T )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( j = 1 )</td>
<td>0.0470</td>
<td>0.2277</td>
</tr>
<tr>
<td>( j = 2 )</td>
<td>0.0485</td>
<td>0.2291</td>
</tr>
<tr>
<td>( j = 3 )</td>
<td>0.0494</td>
<td>0.2370</td>
</tr>
<tr>
<td>( j = 4 )</td>
<td>0.0493</td>
<td>0.2502</td>
</tr>
<tr>
<td>( j = 5 )</td>
<td>0.0506</td>
<td>0.2190</td>
</tr>
<tr>
<td>( j = 6 )</td>
<td>0.0484</td>
<td>0.2169</td>
</tr>
<tr>
<td>( j = 7 )</td>
<td>0.0499</td>
<td>0.2249</td>
</tr>
</tbody>
</table>

We chose to simulate 100 instances, \( (\hat{X}^{(2)}_T)_{j=1}^{100} \) and \( (\hat{X}^{(3)}_T)_{j=1}^{100} \), of the optimal filters at time \( \tilde{T} \). Here, there is no reason to pick one of the simulated instances over
the other amongst the 100. Thereby we choose to average the instances and use

$$\hat{X}_T^{(2)} = \frac{1}{100} \sum_{j=1}^{100} \hat{X}_T^{(2)}, \quad \text{and} \quad \hat{X}_T^{(3)} = \frac{1}{100} \sum_{j=1}^{100} \hat{X}_T^{(3)},$$

when we simulate the paths $[0, \tilde{T}] \ni t \mapsto X_t^{(1)}$. Plots of these future paths of $X^{(1)}$, obtained by the simulation procedure in this section, are given in Figure 6.3. The main thing to notice about the plot is that the paths do not extend to extreme values, and stay rather contained. This makes the calibration of the model a bit easier, since as a consequence of this, the paths of the intensity of the Cox process will not take on too extreme values. Hence it will be easier to obtain transition rates within reasonable bounds. Here, plots of the future paths of $\tilde{\mu}(t, \omega)$ are given in Figure 6.4. Finally, plots for future transition rates, 13 and 26 years into the future, are given in Figure 6.5.

Clearly, by Figure 6.5, there is a chance of obtaining unreasonable transition rates; that is, unreasonable in the sense that their value exceed 1. Eventhough we do obtain some unreasonable transition rates, in the first plot there are only four out of a houndred paths that give unreasonable transition rates, so the results are not terrible. However, the way we have structured the transition rates $\mu_\ast(t, x)$, $0 \leq t \leq T, x \geq 0$, and by our choice of jump sizes, the transition rates increase in size as we simulate paths further into the future. Hence it is inevitable to obtain unreasonable transition rates at some point with this structure. Possible extensions that could help fix this are discussed in the next chapter.

In the next section, Section 6.4, we will give a short summary of the thesis and some concluding remarks.
Figure 6.3: Path $[0, \tilde{T}] \mapsto X_1^{(1)}$, $\tilde{T} = 26$, of (6.39), with an equidistant partitioning $\Pi$, where $|\Pi| = \Delta t := 26/1000$.

Figure 6.4: Path $[0, \tilde{T}] \mapsto \tilde{\mu}(t, \omega)$, $\tilde{T} = 26$, with increments (6.40). Here we have used an equidistant partitioning $\Pi$ with $|\Pi| = \Delta t := 26/1000$. 
82  6.3. SIMULATIONS OF THE MODEL

Figure 6.5: Mortality rates from the last year of data plotted as a thick black line. Future mortality rates obtained from paths given by (6.41), 13 and 26 years into the future, from top to bottom, respectively, are plotted in dark grey. There are 1000 simulated paths with an equidistant partitioning $\Pi$, where $|\Pi| = \Delta t := 26/1000$. 
6.4 Concluding remarks

In this thesis we have built a solid framework based on probability theory, stochastic analysis, Lévy processes, and nonlinear filtering theory with observations that capture jumps. Also, in the chapter on nonlinear filtering theory, a bound for the error estimate when performing Monte Carlo techniques with respect to the Kallianpur-Striebel formula was given, and proven. After the mathematical foundation had been built, Chapter 5 offered a motivation for why transition rates are of such great importance in risk management when dealing with life insurance products.

After all this, we have here in the final chapter first given a brief discussion of already popular models for stochastic transition rates in life insurance. Then finally we introduced a new model for modeling stochastic transition rates in life insurance, where hopefully, having been introduced to already popular models in the field, the reader has gained an insight into why our model has some very attractive futures. In addition, we have given a rather general simulation procedure, as well as a detailed account on how to calibrate the model.

After the general simulation procedure had been introduced, we simulated a special case where the focus was on the jump component with mean reversion in the unknown parametrization process. Here we made specific choices for all the different elements of the model, and gave specific algorithms regarding simulation of each of the relevant processes. Also, the results obtained from performing simulations of our model were given.

In Figure 6.5 the final transition rates were given, where some of the obtained transition rates were unreasonable; that is, unreasonable in the sense that their values exceeded 1.

Even though our model is difficult to calibrate, needs many years of data, and is quite complex compared to other already popular models, it does offer a great deal of flexibility. Each user can decide on which effects the user would like for the model to capture, and how these effects are to be capture. Desired effects such as regime switching and mean reversion can even be captured simultaneously through choice of functions in the unknown parametrization process. These effects can even be captured through the intensity of a jump component, as we did in our specific model which we introduced in the previous section. Generally, by including the unknown parametrization process we are able to capture effects that lead to different types of jump behaviour, such as e.g. cohort effects or natural disasters. In addition the model can be scaled through many of the functions to obtain as good of balance as one wants, between the elements of the model.

In conclusion: The model is quite complex, could be challenging to implement, but offers a great deal of flexibility. Where perhaps the most interesting part of the model is the opportunity for modeling change in the jump behaviour, by use of the unknown parametrization process. However, as of now there is still work to be done to calibrate the model such that reasonable transition rates are obtained on a consistent basis. The next chapter gives possible suggestions for extensions and future work, where the structuring of the transition rates is discussed.
Chapter 7

Extensions and future work

In the sequel we consider an insured individual with a state space $S := \{\ast, \diamondsuit, \dagger\}$. 

7.1 Structure of transition rates

It would be interesting to test out more ways of structuring the transition rates. As of now, we have chosen the transition rates $\mu_{ij}(\cdot, \cdot), i, j \in S$, to be given by

$$
\mu_{ij}(t, x) = \alpha_{ij} + h_{ij}^{1,\ell_1}(t, x) + \beta_{ij} \exp\left\{h_{ij}^{1,\ell_2}(t, x)\right\}, \ 0 \leq t \leq T, \ x \geq 0,
$$

where $\alpha_{ij}, \beta_{ij} \geq 0$ are constants and $h_{ij}^{1,\ell_1}(\cdot, \cdot)$ and $h_{ij}^{2,\ell_2}(\cdot, \cdot)$ are specified by (6.8) and (6.9), respectively. One suggested modification to this, could be to exchange $h_{ij}^{1,\ell_1}(\cdot, \cdot)$ and $h_{ij}^{2,\ell_2}(\cdot, \cdot)$ with $\tilde{h}_{ij}^{1,\ell_1}(\cdot, \cdot)$ and $\tilde{h}_{ij}^{2,\ell_2}(\cdot, \cdot)$, respectively, where the latter two are given by

$$
\forall(i, j, t, x) \in S \times S \times [0, T] \times \mathbb{R}_{\geq 0}, \begin{cases}
\tilde{h}_{ij}^{1,\ell_1}(t, x) = \sum_{n=0}^{\ell_1} \binom{n}{\ell_1} t^{n+1} + \kappa_{n+1} x^n, \\
\tilde{h}_{ij}^{1,\ell_1}(t, x) = \sum_{n=0}^{\ell_2} \binom{n}{\ell_1+2} t^{n+\ell_1+2} + \kappa_{n+\ell_1+2} x^n,
\end{cases}
$$

where $(\kappa_n)_{n=1}^{\ell_2+\ell_1+2} \subset \mathbb{R}$ could be independent random variables, or just constants. As always we will need to impose some constraints on the coefficients in the polynomial $\tilde{h}_{ij}^{1,\ell_1}(\cdot, \cdot)$ to avoid unreasonable transition rates. Adding constants to the coefficients of the polynomials inside the functions for the transition rates, is thought to make it significantly easier to obtain reasonable transition rates. Perhaps one could set the constants to be coefficients obtained from curvefitting for the last year of data, and then let the transition rates evolve from there. This might also help with the problem of having transition rates that grow over time.

7.2 Extend the dynamics of $(Y_t)$

In this thesis we have implemented a model with stochastic coefficients in the transition rates, that evolve with dynamics given by a pure jump process. It is quite difficult to calibrate the model when it is only carried by a pure jump process. We would like to attempt to implement the model with an observation process with
7.3 The function \( \lambda(\cdot,\cdot,\cdot) \)

As of today, we have only developed a simulation procedure for the case where \( \lambda(s,x,z) := c\tilde{\lambda}(s,x) \in \mathbb{R}_{\geq 0}, 0 \leq s \leq T, x \in \mathbb{R}^d \), and \( c \) is some real-valued, non-negative constant. It would be interesting to further research the possibility of a dependency on \( z \) in \( \lambda(s,x,z) \in \mathbb{R}_{\geq 0}, 0 \leq s \leq T, x \in \mathbb{R}^d \), \( z \in \mathbb{R}_0^m \).

7.4 Dynamics of \( (X_t) \)

Here we discuss different functions for the coefficients in the dynamics of \( X \), with regard to future work.

7.4.1 The coefficient \( b(\cdot) \)

It would be of interest to consider discontinuous functions for \( b(\cdot) \) in the dynamics of \( X = (X_t)_{0 \leq t \leq T} \). As of today, the first function we would like to try for \( b(\cdot) \) is

\[
b(x) = a_1(\tilde{b}_1 - x)1_{\{\|x\| < \theta\}} + a_2(\tilde{b}_2 - x)1_{\{\|x\| \geq \theta\}}, \quad x \in \mathbb{R}^d,
\]
where $\theta \in \mathbb{R}$ is some, constant, chosen threshold, and $a_1, a_2, \tilde{b}_1, \tilde{b}_2 > 0$ are constant parameters to be estimated. With such a choice for $b(\cdot)$ one could assume for $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^5$ to have dynamics given by, for all $0 \leq t \leq T$,

\[
\begin{cases}
  dX_t^{(1)} = \left\{ a_1(\tilde{b}_1 - X_t^{(1)})1_{\{\|x\| < \theta\}} + a_2(\tilde{b}_2 - X_t^{(1)})1_{\{\|x\| \geq \theta\}} \right\} dt + d\tilde{B}_t, \\
  dX_t^{(2)} = 0, \quad X_0^{(2)} = a_1, \\
  dX_t^{(3)} = 0, \quad X_0^{(3)} = \tilde{b}_1, \\
  dX_t^{(4)} = 0, \quad X_0^{(4)} = a_2, \\
  dX_t^{(5)} = 0, \quad X_0^{(5)} = \tilde{b}_2, \\
  dX_t^{(6)} = 0, \quad X_0^{(6)} = \theta,
\end{cases}
\]

where $a_1, a_2, \tilde{b}_1, \tilde{b}_2, \theta \in \mathbb{R}_{\geq 0}$ are constants and $\tilde{B} = (\tilde{B}_t)_{0 \leq t \leq T}$ is a Brownian motion.

Now we would find the best estimates of $a_1, a_2, \tilde{b}_1, \tilde{b}_2$ and $\theta$ by applying Monte Carlo techniques to obtain the optimal filters.

This representation has been discussed in the previous chapter, however it has not been implemented. The attractive part of such a representation is, amongst other things, the opportunity to capture the effects of regulatory changes in the insurance market, or political changes, with respect to disability insurance.

Before we can implement our model with such functions for $b(\cdot)$, we might have to find another way of simulating paths of the signal process. The reason for this being that the Euler scheme we have used in Chapter 6, might not work with discontinuous functions for $b(\cdot)$.

### 7.4.2 The diffusion coefficient $\sigma(\cdot)$

As of now, we have only attempted for $\sigma(\cdot)$ to be a matrix with constant entries. However, it would be interesting to attempt a higher dimension of $X$ than we tried to implement with the Vasicek dynamics. Perhaps a suggestion could be to try the following dynamics for $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^4$:

\[
\forall 0 \leq t \leq T, \quad \begin{cases}
  dX_t^{(1)} = X_t^{(3)}(X_t^{(2)} - X_t^{(1)}) dt + X_t^{(4)} d\tilde{B}_t, \\
  dX_t^{(2)} = 0, \\
  dX_t^{(3)} = 0, \\
  dX_t^{(4)} = 0,
\end{cases}
\]

with $X_0 := x^*$ as a stochastic variable, with values in $\mathbb{R}^4$, that is independent of the observation process, and $\tilde{B} = (\tilde{B}_t)_{0 \leq t \leq T}$ is a Brownian motion.
7.4. DYNAMICS OF \((X_T)\)
Appendix A

Distributions

A.1 Distributions for \( \mathbb{R} \)-valued stochastic variables

A.1.1 The gamma distribution

Let \( X \) be a stochastic variable with values in \( \mathbb{R} \). We say that \( X \) is gamma distributed with shape parameter \( \alpha > 0 \) and scale parameter \( \beta > 0 \), if the density \( \varphi(\cdot) \) of \( X \) is given by

\[
    \varphi(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} \exp \left\{ -\frac{x}{\beta} \right\}, \quad x \in \mathbb{R},
\]

and we write \( X \sim \text{Gamma}(\alpha, \beta) \). Here \( \Gamma(\cdot) \) is the gamma function.

A.1.2 The Gaussian distribution

Let \( X \) be a stochastic variable with values in \( \mathbb{R} \). We say that \( X \) is Gaussian with expectation \( \mu \) and variance \( \sigma^2 \), if the density \( \varphi(\cdot) \) of \( X \) is given by

\[
    \varphi(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma^2} (x - \mu)^2 \right\}, \quad x \in \mathbb{R},
\]

and we write \( X \sim \text{N}(\mu, \sigma^2) \). If \( \mu = 0 \) and \( \sigma^2 = 1 \), we say that \( X \) is standard Gaussian.

A.1.3 The uniform distribution

Let \( X \) be a stochastic variable with values in \( \mathbb{R} \). We say that \( X \) is uniformly distributed, over \( [a, b] \), \( a < b \), \( a, b \in \mathbb{R} \), if the density \( \varphi(\cdot) \) of \( X \) is given by

\[
    \varphi(x) = \begin{cases} 
    \frac{1}{b-a} & \text{if } x \in [a, b], \\
    0 & \text{otherwise},
    \end{cases}
\]

and we write \( X \sim \text{U}(a, b) \).
A.2 Distributions for \( \mathbb{N}_0 \)-valued stochastic variables

A.2.1 The Poisson distribution

Let \( X \) be a random variable with values in \( \mathbb{N}_0 \). We say that \( X \) is Poisson distributed with intensity \( \lambda > 0 \), if

\[
P[X = k] = e^{-\lambda} \frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0,
\]

and we write \( X \sim \text{Poisson}(\lambda) \).
Appendix B

Calculations

B.1 Chapter 3

B.1.1 Calculations in proof of Proposition 1.6

The following theorem is borrowed from page 59 in Çınlar [5].

**Theorem B.1** (Bounded convergence theorem). Let \((X_n)_{n \in \mathbb{N}} \subset \mathbb{R}^d\) be a sequence of stochastic variables, such that \(\lim_{n \to \infty} X_n\) exists a.s. If there exists a constant \(b \in (0, \infty)\) such that \(\|X_n\| \leq b\), then

\[
E \left[ \lim_{n \to \infty} X_n \right] = \lim_{n \to \infty} E[X_n].
\]

Let \(\eta = (\eta_t)_{0 \leq t \leq T} \subset \mathbb{R}^d\) be a Lévy process. Let \(\varphi_t(\cdot)\) be the characteristic function of \(\eta_t\), \(0 \leq t \leq T\). Define \(\psi_t(\cdot)\), such that \(\psi_t(\cdot) = \log\{\varphi_t(\cdot)\}, 0 \leq t \leq T\).

**Claim**: For any pair \((K, J) \in \mathbb{N} \times \mathbb{N}\), we have that, for all \(u \in \mathbb{R}^d\), \(\psi_K(u) = J\psi_{K/J}(u)\), such that \(\psi_K(u) = K\psi_1(u)\).

**Proof**. First we have, for all \(u \in \mathbb{R}^d\),

\[
\psi_K(u) = \log\{E[\exp\{i\langle u, \eta_K \rangle\}]\}
= \log\{E[\exp\{i\langle u, \eta_K - \eta_0 \rangle\}]\}
\]

\[
\overset{(*)}{=} \log \left\{ \mathbb{E} \left[ \exp \left\{ i \sum_{j=0}^{J-1} \eta \left\{ \frac{K_{j+1}}{J} - \frac{K_j}{J} \right\} \right\} \right] \right\}
\]

\[
= \log \left\{ \mathbb{E} \left[ \exp \left\{ \sum_{j=0}^{J-1} i \langle u, \eta \left\{ \frac{K_{j+1}}{J} - \frac{K_j}{J} \right\} \right\} \right] \right\}
\]

\[
= \log \left\{ \mathbb{E} \left[ \exp \left\{ i \langle u, \eta \left\{ \frac{K}{J} - \frac{K}{J} \right\} \right\} \right] \right\}
\]

\[
= J \log \left\{ \mathbb{E} \left[ \exp \left\{ i \langle u, \eta \frac{K}{J} \rangle \right\} \right] \right\}
\]

\[
= J\psi_{K/J}(u).
\]

Here we have used:
We have for all $(ii)$ which proves that

\[ H \]

Here we will use the independent increments of

\[ (ii) \]

By exchanging $J$ with $K$ in the above calculations, we obtain the relation $\psi_K = K\psi_{K/K(u)} = K\psi_1$.  

\[ \Box \]

\[ B.2. \quad \text{CHAPTER 4} \]

\[ B.1.2 \quad \text{Proof of Proposition 3.9} \]

Proof. (i) We have for all $0 \leq s \leq t \leq T$,

\[
\E[M_t|\mathcal{F}_s] = \E[\eta_t|\mathcal{F}_s] - \E[\eta_s] \\
= \E[\eta_t - \eta_s|\mathcal{F}_s] + \eta_s - \E[\eta_t] \\
= \E[\eta_t] - \E[\eta_s] + \eta_s - \E[\eta_t] \\
= \eta_s - \E[\eta_s],
\]

which proves that $M$ is a martingale.

(ii) Here we will use the independent increments of $M$ and that $M$ is a martingale. We have, for all $0 \leq s \leq t \leq T$,

\[
\E[(M_t)^2 - \E[(M_t)^2]|\mathcal{F}_s] = \E[(M_t - M_s + M_s)^2|\mathcal{F}_s] - \E[(M_t)^2] \\
= \E[(M_t - M_s)^2|\mathcal{F}_s] + 2\E[M_s(M_t - M_s)|\mathcal{F}_s] \\
+ \E[(M_s)^2|\mathcal{F}_s] - \E[(M_t)^2] \\
= \E[(M_t - M_s)^2] + 2M_s\E[M_t - M_s] \\
+ (M_s)^2 - \E[(M_t)^2] \\
= \E[(M_t)^2] - 2\E[M_tM_s] + \E[(M_s)^2] - \E[(M_t)^2] \\
+ 2M_s\{\E[E[M_t|\mathcal{F}_s]] - \E[M_s]\} + (M_s)^2 \\
=(M_s)^2 - 2\E[M_s\E[M_t|\mathcal{F}_s]] + \E[(M_s)^2] \\
=(M_s)^2 - \E[(M_t)^2].
\]

Hence the process $\{\{M_t\}^2 - \E[(M_t)^2]\}_{0 \leq t \leq T}$ is a martingale.  

\[ \Box \]

\[ B.2 \quad \text{Chapter 4} \]

The following definition is borrowed from Appendix A, page 294, in Bain and Crisan [2].

\[ B.1 \quad \text{Definition B.1 (Regular conditional probability).} \quad \text{Let} \ (\Omega, \mathcal{F}, \mathbb{P}) \ \text{be a complete probability space. Let} \ (\mathcal{S}, \mathcal{S}) \ \text{be a measurable space. Let} \ \mathcal{H} \ \text{be a sub-}\sigma\text{-algebra of} \ \mathcal{F}. \ \text{Let} \ X \ \text{be a stochastic variable on} \ (\Omega, \mathcal{F}, \mathbb{P}) \ \text{with values in} \ \mathcal{S}. \ \text{A function} \ \mathbb{Q}[\omega, A] \ \text{defined for all} \ \omega \in \Omega \ \text{and} \ A \in \mathcal{S} \ \text{is called a regular conditional distribution of} \ X \ \text{with respect to} \ \mathcal{H} \ \text{if} \\

(i) \quad \text{For each} \ A \in \mathcal{S}, \ \Omega \ni \omega \mapsto \mathbb{Q}[\omega, A] \ \text{is} \ \mathcal{H}-\text{measurable.} \\

(ii) \quad \text{For each} \ \omega \in \Omega, \ \mathcal{S} \ni A \mapsto \mathbb{Q}[\omega, A] \ \text{is a probability measure on} \ (\mathcal{S}, \mathcal{S}). \\

(iii) \quad \text{For any} \ A \in \mathcal{S}, \ \mathbb{Q}[\omega, A] = \mathbb{P}[X \in A|\mathcal{H}](\omega), \ \mathbb{P}-\text{a.s.} \]
B.3 Chapter 6

B.3.1 Lee-Carter: MLE estimates for $\sigma_\xi$ and $\theta$

Assume that the sequence $(k_t)_{t=1}^T$ follows a random walk with drift, given by

$$k_t = k_{t-1} + \theta + \xi, \quad t = 2, \ldots, T,$$

where $\xi \sim N(0, \sigma_\xi^2)$, such that $k_t \sim N(k_{t-1} + \theta, \sigma_\xi^2)$, $t = 2, \ldots, T$. Then the maximum likelihood estimates for $\theta$ and $\sigma_\xi^2$ are, respectively, given by

$$\hat{\theta} = \frac{k_T - k_1}{T - 1} \quad \text{and} \quad \hat{\sigma}_\xi^2 = \frac{1}{T - 1} \sum_{j=1}^{T-1} (k_{j+1} - k_{j} - \hat{\theta})^2.$$

Proof. Since $k_t \sim N(k_{t-1} + \theta, \sigma_\xi^2)$, $t = 2, \ldots, T$, we have the log-likelihood function

$$\ell(\sigma_\xi, \theta) = \log \left\{ \prod_{j=1}^{T-1} \frac{1}{\sigma_\xi \sqrt{2\pi}} \exp \left\{ -\frac{(k_{j+1} - (k_j + \theta))^2}{2\sigma_\xi^2} \right\} \right\}$$

$$= \log \left\{ \left( \frac{1}{\sigma_\xi \sqrt{2\pi}} \right)^{T-1} \right\} + \log \left\{ \exp \left\{ \frac{1}{2\sigma_\xi^2} \sum_{j=1}^{T-1} (k_{j+1} - (k_j + \theta))^2 \right\} \right\}$$

$$= -(T - 1) \log \{\sigma_\xi \sqrt{2\pi}\} + \frac{1}{2\sigma_\xi^2} \sum_{j=1}^{T-1} (k_{j+1} - k_j - \theta)^2.$$

Now by differentiating $\ell(\cdot)$, with respect to $\theta$, and putting the resulting expression equal to zero, we have

$$0 = \frac{\partial}{\partial \theta} \ell(\sigma, \theta) = -\frac{1}{2\sigma_\xi^2} \sum_{j=1}^{T-1} (k_{j+1} - k_j - \theta) \iff \hat{\theta} = \frac{k_T - k_1}{T - 1}.$$

By a similar argument for $\sigma_\xi$, we have

$$0 = \frac{\partial}{\partial \sigma_\xi} \ell(\sigma, \theta) = -\frac{T}{\sigma_\xi} + \frac{1}{\sigma_\xi^2} \sum_{i=1}^{T} (k_i - k_{i-1} - \hat{\theta})^2$$

$$\iff \hat{\sigma}_\xi^2 = \frac{1}{T} \sum_{i=1}^{T} (k_i - k_{i-1} - \hat{\theta})^2.$$

We have now arrived at the desired result. $\blacksquare$

B.3.2 The Vasicek model

Let $r = (r_t)_{0 \leq t \leq T}$ be a stochastic process with values in $\mathbb{R}$, and assume that $r$ is described by the following SDE:

$$r_t = r_0 + \int_0^t a(b - r_s) \, ds + \int_0^t \sigma \, dB_s, \quad 0 \leq t \leq T,$$

\hspace{4cm} (*)
where \( r_0, \sigma \in \mathbb{R} \) and \( a_0, b_0 > 0 \) are constants. Then
\[
\mathbb{E}[r_t] \xrightarrow{t \to \infty} b, \quad \text{and} \quad \left\{ \begin{array}{ll}
\mathbb{E}[r_t] < b, & \text{as } t \to \infty, \quad \text{if } b < r_0, \\
\mathbb{E}[r_t] > b, & \text{as } t \to \infty, \quad \text{if } b > r_0.
\end{array} \right.
\]

**Proof.** By defining \( Z = (Z_t)_{0 \leq t \leq T} \subset \mathbb{R} \) to be the stochastic process given by
\[
Z_t = r_t - b, \quad Z_0 := z_0 := r_0 - b, \quad 0 \leq t \leq T,
\]
and substituting into (\( * \)), we obtain
\[
Z_t = z_0 - a \int_0^t Z_s \, ds + \int_0^t \sigma \, dB_s, \quad 0 \leq t \leq T.
\]

Before we continue, we need the following theorem.

**Theorem B.2 (Itô’s Lemma).** Let \( X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R} \) be a stochastic process on the filtered probability space \((\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_t)_{0 \leq t \leq T}, \mathbb{P})\), with the representation
\[
X_t = X_0 + \int_0^t F(s) \, ds + \int_0^t G(s) \, dB_s, \quad 0 \leq t \leq T,
\]
where \( F(t) \in \mathbb{R} \) and \( G(t) \in \mathbb{R} \), \( 0 \leq t \leq T \), are \( \mathbb{F} \)-adapted and satisfy the following conditions:
\[
\mathbb{E} \left[ \int_0^t |F(s)|^2 \, ds \right] < \infty \quad \text{and} \quad \mathbb{E} \left[ \int_0^t |G(s)|^2 \, ds \right] < \infty, \quad 0 \leq t \leq T.
\]
Then we say that \( X \) is an Itô process. Now for \( g \in C^{1,2}([0, T] \times \mathbb{R}; \mathbb{R}) \), the following is also an Itô process:
\[
Y_t = g(t, X_t), \quad 0 \leq t \leq T,
\]
and we have
\[
Y_t = g(0, X_0) + \int_0^t \left\{ \frac{\partial}{\partial t} g(s, X_s) + F(s) \frac{\partial}{\partial x} g(s, X_s) + \frac{1}{2} G(s) \frac{\partial^2}{\partial x^2} g(s, X_s) \right\} \, ds \\
+ \int_0^t G(s) \frac{\partial}{\partial x} g(s, X_s) \, dB_s, \quad 0 \leq t \leq T.
\]

Now we wish to apply Itô’s lemma to our problem. Let \( g \in C^{1,2}([0, T] \times \mathbb{R}; \mathbb{R}) \) be the function
\[
g(t, x) = x \exp\{at\}, \quad 0 \leq t \leq T, \quad x \in \mathbb{R}.
\]
By Itô’s Lemma we obtain the following:
\[
Z_t \exp\{at\} = g(t, Z_t)
\]
\[
= g(0, Z_0) + \int_0^t \left\{ \frac{\partial}{\partial t} g(s, Z_s) + (-a Z_s) \frac{\partial}{\partial x} g(s, Z_s) \\
+ \frac{1}{2} \sigma^2 \frac{\partial^2}{\partial x^2} g(s, X_s) \right\} \, ds + \int_0^t \sigma \frac{\partial}{\partial x} g(s, X_s) \, dB_s \\
\iff Z_t = z_0 \exp\{-at\} + \sigma \int_0^t \exp\{a(s-t)\} \, dB_s, \quad 0 \leq t \leq T.
\]
Now by how we have defined the process $Z$, we find that
\[ r_t = Z_t + b \]
\[ = b + (r_0 - b) \exp(-at) + \sigma \int_0^t \exp(a(s-t)) \, dB_s, \quad 0 \leq t \leq T. \]

Now we find the expected value of $r_t$ for all $0 \leq t \leq T$, by
\[ \mathbb{E}[r_t] = \mathbb{E} \left[ b + (r_0 - b) \exp(-at) + \sigma \int_0^t \exp(a(s-t)) \, dB_s \right] \]
\[ = b + (r_0 - b) \exp(-at) + \sigma \mathbb{E} \left[ \int_0^t \exp(a(s-t)) \, dB_s \right] \]
\[ = b + (r_0 - b) \exp(-at), \]

where by Proposition 2.3 the expectation of the integral, with respect to the Brownian motion, is equal to zero. Now we see that
\[ \mathbb{E}[r_t] \xrightarrow{t \to \infty} b. \]

More specifically, if $r_0 > b$, then $r_0 - b > 0$, such that
\[ \mathbb{E}[r_t] \searrow b, \quad \text{as} \ t \to \infty, \]

and if $r_0 < b$, then $r_0 - b < 0$, such that
\[ \mathbb{E}[r_t] \nearrow b, \quad \text{as} \ t \to \infty. \]

Now we have arrived at the desired result. \[ \blacksquare \]
Appendix C

Code

C.1 Programs for Chapter 3

C.1.1 Program for plots in Figure 3.1

R program.

```r
1 %-----------------------------------
2 % Plot path of B.m. and Poi. proc.:
3 %-----------------------------------
4 5 n = 1000 % Number of time points
6 d = 1000 % Number of r.v.'s to generate
7 h = 10/n % subinterval width
8 lambda = 0.5 % Poisson parameter
9 10 %---------------------
11 % Simulate paths:
12 %---------------------
13 14 N = cumsum(rpois(d,h*lambda)) % Poi. proc.
15 B = cumsum(sqr(h)*rnorm(d,0,1)) % B.m.
16 17 %--------
18 % Plot:
19 %--------
20 21 x = seq(0,10,length=d)
22 23 % Plot Poi. proc.:
24 plot(x,N,xlab='',ylab='',pch=',xlim=c(0,10),ylim=c(min(c(B,N)),max(c(B,N))
25 % Plot B.m.:
26 lines(x,B,lty=4,lwd=2)
```

Programs for Chapter 5

Programs for plots in Figure 5.1

MATLAB program to obtain coefficients in the Gompertz-Makeham-type models.

```matlab
1 %---------------------------------------
2 % Extract central death rates from data:
3 %---------------------------------------
```
% Central death rates 1947:
mtx_1947 = textread('/uio/hume/student-u73/knutgaas/master/thesis/simulated_examples/data/austria_life_table_fm_1947.txt','%*s %*s %f %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s', 'headerlines', 3);

% Central death rates 1981:
mtx_1981 = textread('/uio/hume/student-u73/knutgaas/master/thesis/simulated_examples/data/austria_life_table_fm_1981.txt','%*s %*s %f %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s', 'headerlines', 3);

% Central death rates 2007:
mtx_2007 = textread('/uio/hume/student-u73/knutgaas/master/thesis/simulated_examples/data/austria_life_table_fm_2007.txt','%*s %*s %f %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s %*s', 'headerlines', 3);

age = 0:110;
age = age.';
I = 111; % Number of ages per year
mtx = zeros(I,3);
mtx(:,1) = mtx_1947(:,1);
mtx(:,2) = mtx_1981(:,1);
mtx(:,3) = mtx_2007(:,1);

%---------------------------------
% Fit a standard Gompertz-Makeham:
%---------------------------------
f = fittype( 'g_m_standard(x,a1,a2,a3)' );

% Create matrix to store coefficients from curve fitting:
alpha = zeros(3,3);

% Set options for the fit of f:
options = fitoptions(f);
options.Lower = [0 0 -100];
options.Upper = [1 5 100];
options.StartPoint = [0.1 0.1 0];

% Obtain coefficients by curve fitting:
curve_1947 = fit(age,mtx_1947,f,options);
alpha(:,1) = [ curve_1947.a1 , curve_1947.a2 , curve_1947.a3 ];
curve_1981 = fit(age,mtx_1981,f,options);
alpha(:,2) = [ curve_1981.a1 , curve_1981.a2 , curve_1981.a3 ];
curve_2007 = fit(age,mtx_2007,f,options);
alpha(:,3) = [ curve_2007.a1 , curve_2007.a2 , curve_2007.a3 ];

%------------------------------------
% Fit a generalized Gompertz-Makeham:
%------------------------------------
g = fittype( 'g_m_general(x,b1,b2,b3,b4,b5)' );

% Create matrix to store coefficients from curve fitting:
beta = zeros(5,3);

% Set options for the fit of g:
options = fitoptions(g);
options.Lower = [0 0 -100 -100 -0.1];
options.Upper = [0.1 0.1 100 100 0.1];
options.StartPoint = [0 0 0 0 0];

% Obtain coefficients by curve fitting:
curve_1947 = fit(age,mtx_1947,g,options);
beta(:,1) = [ curve_1947.b1 , curve_1947.b2 , curve_1947.b3 , curve_1947.b4 , curve_1947.b5 ];
curve_1981 = fit(age,mtx_1981,g,options);
beta(:,2) = [ curve_1981.b1 , curve_1981.b2 , curve_1981.b3 , curve_1981.b4 , curve_1981.b5 ];
APPENDIX C. CODE

R program to create plots.

```
#----------------------------------------------------------------------
#----------------------------------------------------------------------
#---------------------------------------------------
# Read central death rates from data:  
#---------------------------------------------------
# mtx_1947 = read.table("/mn/anatu/studenter-u3/knutgaas/master/thesis/  
# names=NULL);  
# mtx_1981 = read.table("/mn/anatu/studenter-u3/knutgaas/master/thesis/  
# names=NULL);  
# mtx_2007 = read.table("/mn/anatu/studenter-u3/knutgaas/master/thesis/  
# names=NULL);  
# age=0:110;  
#---------------------------------------------------
# Coefficients in Gompertz-Makeham:  
#---------------------------------------------------
# alpha_1947_1 = 0.00000000571140683920  
# alpha_1947_2 = 0.00129478681287773802  
# alpha_1947_3 = 0.059480028444423518065  
#---------------------------------------------------
# alpha_1981_1 = 0.00000000262217223987  
# alpha_1981_2 = 0.0006055297381871402  
# alpha_1981_3 = 0.07970704193327543541  
```
# Coefficients in generalized Gompertz-Makeham:

```r
beta_1947_1 = 0.00000000020458415778;
beta_1947_2 = 0.00000000013611183451;
beta_1947_3 = -17.18360749103539575344;
beta_1947_4 = 0.28318068727031331377;
beta_1947_5 = -0.00117339517904918632;
```

```r
beta_1981_1 = 0.00000008178432929628;
beta_1981_2 = 0.00007657650852511533;
beta_1981_3 = -25.11417013419226762494;
beta_1981_4 = 0.431446162656342482874;
beta_1981_5 = -0.00186491786604405592;
```

```r
beta_2007_1 = 0.00000000013093486532;
beta_2007_2 = 0.00000000638427429841;
beta_2007_3 = -29.70704107813693184426;
beta_2007_4 = 0.50210800354194429218;
beta_2007_5 = -0.00212787084213699562;
```

### Generalized Gompertz-Makeham:

```r
GM_general = function(x,b1,b2,b3,b4,b5){
g = b1+b2*x+exp(b3+b4*x+b5*(x^2));
return(g)
}
```

### Standard Gompertz-Makeham:

```r
GM_standard = function(x,a1,a2,a3){
f = a1+a2*exp(a3*x);
return(f)
}
```

### Plot:

```r
x = seq(0,110,0.001)
plot(x mtx_1947[,3], xlab='', ylab='', pch=5, cex=0.75, ylim=c(0,1), xlim=c(0,110))
lines(x, GM_general(x, beta_1947_1, beta_1947_2, beta_1947_3, beta_1947_4, beta_1947_5), lty=5, lwd=2)
lines(x, GM_standard(x, alpha_1947_1, alpha_1947_2, alpha_1947_3), lty=1, lwd=2)
```

```r
plot(x mtx_1981[,3], xlab='', ylab='', pch=5, cex=0.75, ylim=c(0,1), xlim=c(0,110))
lines(x, GM_general(x, beta_1981_1, beta_1981_2, beta_1981_3, beta_1981_4, beta_1981_5), lty=5, lwd=2)
lines(x, GM_standard(x, alpha_1981_1, alpha_1981_2, alpha_1981_3), lty=1, lwd=2)
```

```r
plot(x mtx_2007[,3], xlab='', ylab='', pch=5, cex=0.75, ylim=c(0,1), xlim=c(0,110))
lines(x, GM_general(x, beta_2007_1, beta_2007_2, beta_2007_3, beta_2007_4, beta_2007_5), lty=5, lwd=2)
lines(x, GM_standard(x, alpha_2007_1, alpha_2007_2, alpha_2007_3), lty=1, lwd=2)
```
C.2 Programs for Chapter 6

R program for Vasicek plots in Figure 6.1 and Figure 6.2.

```r
# ------------------------------------------------------------------
# Vasicek - model with parameters a=(-1,-1), b=(-2,-1), r0=(-3,-1), sigma=1:
# ------------------------------------------------------------------
# Vasicek parameters:
# a_1 = -1
# b_1 = -2
# r0_1 = -3
# a_2 = -1
# b_2 = -1
# r0_2 = -1
#
# Partitioning:
# T = 20
# n = 100
# h = T/n
# Guassian increments of B.m.:
# xi_1 = rnorm(n,0,1)
# xi_2 = rnorm(n,0,1)
#
# Simulate Vasicek
# r_1 = c()
# r_2 = c()
# r_1[1] = r0_1
# r_2[1] = r0_2
# for (i in 2:(n+1)){
#   r_1[i] = r_1[i-1] + a_1*(b_1-r_1[i-1])*h + xi_1[i]*h
#   r_2[i] = r_2[i-1] + a_2*(b_2-r_2[i-1])*h + xi_2[i]*h
# }
# grid = seq(from=0, to=T, by=h)
# plot(grid, r_1[1:(n+1)], xlab='', ylab='', pch='*')
# lines(grid, r_1[1:(n+1)])
# plot(grid, r_2[1:(n+1)], xlab='', ylab='', pch='*')
# lines(grid, r_2[1:(n+1)])
# --------------------------------------------------
# Vasicek - model initialized by Gamma, and uniform:
# --------------------------------------------------
# Initial values --- Gamma:
# a_Gam = rgamma(1, shape=1, scale=2)
# b_Gam = rgamma(1, shape=1, scale=2)
# r0_Gam = rgamma(1, shape=1, scale=2)
# Initial values --- Uniform:
# a_Uni = runif(1, min=0, max=0.1)
# b_Uni = runif(1, min=0, max=0.1)
# r0_Uni = runif(1, min=0, max=0.5)
# Guassian increments of B.m.:
# xi_Gam = rnorm(n,0,1)
# xi_Uni = rnorm(n,0,1)
# Simulate Vasicek
# r_Gam = c()
# r_Uni = c()
# r_Gam[1] = r0_Gam
# r_Uni[1] = r0_Uni
# for (i in 2:(n+1)){
```
MATLAB program for simulation of the stochastic transition rates.

```matlab
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52
% Program:
% Monte Carlo simulation to obtain stochastic transition rates, from alive to dead.
% I---------------------------------------------
% grey_colour = [127 125 126] ./ 255; % Colour for plot of simulated results
% I---------------------------------------------
% Obtain observation process:
% I---------------------------------------------
% data_mtx = textread('..data/norway_life_table_both_sex.txt','%*s %*s %f %*s %*s %*s %*s %*s %*s %*s %*s','headerlines',3); % mx from data
% age = 0:110;
% I = 111; % Number of ages per year
% T = 2009-1846+1; % Number of years of data
% m = 5; % Number of components in Y(t)
% Create matrix of mortality rates (mx): (Each column represents a year)
% mtx = zeros(I,T);
% mtx(:,1) = data_mtx(1:111);
% for j=1:(T-1)
% mtx(:,j+1) = data_mtx((j*111+1):((j+1)*111));
% end
% Y = zeros(5,T);
% f = fittype('mu(x,a1,a2,a3,a4,a5)'); % Chosen function to fit
% options = fitoptions(f);
% Bounds for coefficients:
% options.Lower = [0 0 -100 -100 -0.1];
% options.Upper = [100 100 100 100 0.1];
% options.StartPoint = [0, 0, 0, 0, 0];
% Obtain observations for each year:
% Y = zeros(5,T);
% for j=1:T
% curve = fit(age,mtx(:,j),f,options); % Fit curve to obtain observation process
% Y(:,j) = [curve.a1,curve.a2,curve.a3+20,curve.a4,curve.a5];
% end
% I---------------------------------------------
% Define jumps:
% I---------------------------------------------
% V = zeros(m,T-1);
```
d = zeros(1, T-1);
for j = 1:(T-1);
    V(:, j) = Y(:, j+1) - Y(:, j);
    d(j) = sqrt(V(:, j)'*V(:, j));
end

Ybar = (1/(T-1))*sum(d);
epsilon_bar = 0.15;
epsilon_lambda = 2.3*0.0001;
jump = Ybar + epsilon_bar; % Jump if: Y(i+1) - Y(i) > jump

% Count number of jumps:
jump_counter = 0;
for j = 1:length(d);
    if d(j) > jump
        jump_counter = jump_counter + 1;
    end
end

% -----------------------------
% Simulate the Vasicek model:
% -----------------------------
N = 1000; % Partitioning of [0, T]
L = 1000; % Number of simulated paths
K = 100; % Number of times, one simulates a sequence of paths

pi2 = zeros(1, K);
pi3 = zeros(1, K);
h = T/N; % Step-size
X1 = zeros(N, L, K);
X2 = zeros(1, L, K);
X3 = zeros(1, L, K);
}

X1(1, :, :) = 0.1*rand(1, L, K); % Apriori
X2(1, :, :) = 0.1*rand(1, L, K); % Apriori
X3(1, :, :) = 0.5*rand(1, L, K); % Apriori

xi = normrnd(0, 1, [N-1, L, K]); % N(0, 1) to help simulate B(t)

% -----------------------------
% Simulate L paths of X1, K times:
% -----------------------------
for k = 1:K
    for l = 1:L
        for n = 1:(N-1)
            X1(n+1, l, k) = X1(n, l, k) + X3(1, l, k)*(X2(1, l, k) - X1(1, l, k))*h + xi(n, l, k)*sqrt(h);
        end
    end
end

% ---------------------------------
% Obtain X1(t), for t = 0, 1, 2, ..., 163 (by linear interpolation)
% ---------------------------------
X1year = zeros(T, L, K);
X1year(1, :, :) = X1(1, :, :);
for k=1:K
    for l=1:L
        for t=2:(T-1)
            t1 = floor(t/h +1);
            t2 = ceil(t/h +1);
            if t1==t2
                X1year(t,l,k) = X1(t1,l,k);
            else
                s = (t-t1)/(t2-t1);
                X1year(t,l,k) = X1(t1,l,k) + s*(X1(t2,l,k)-X1(t1,l,k));
            end
        end
    end
end

% -----------------------------------------------------------

% -----------------------------
% Compute density process Z(t):
% -----------------------------

I1 = zeros(1,L,K);
for k=1:K
    for l=1:L
        for j=1:(T-1)
            if d(j)>jump
                I1(1,l,k) = I1(1,l,k) + log(lambda(X1year(j+1,l,k),X2(1,l,k),X3(1,l,k)));
            end
        end
    end
end

% Second: Compute I2 by trapezoidal rule:
I2 = zeros(1,L,K);
for k=1:K
    for l=1:L
        I2(1,l,k) = (h/2)*(2-lambda(X1(1,l,k),X2(1,l,k),X3(1,l,k)) - lambda(X1(N,l,k),X2(1,l,k),X3(1,l,k))) + h*sum(1-lambda(X1(:,1,l),X2(1,l,k),X3(1,l,k)));
    end
end

% Third: Obtain M(t):
Z = zeros(1,L,K);
for k=1:K
    for l=1:L
        Z(1,l,k) = exp(I1(1,l,k) + I2(1,l,k));
    end
end

% Compute <pi(t),f> for X1(t),X2(t),X3(t), with f(x)=x:

Psi2_x = sum(Z(1,:,k).*X2(1,:,k));
Psi3_x = sum(Z(1,:,k).*X3(1,:,k));

% Compute Psi*L with f(x)=1:
APPENDIX C. CODE

```
183  psi_1 = sum( Z(:,1:k) );
185  % Compute hat(I)1(t) hat(I)2(t):
186  pi2(1,k) = psi_2_x/psi_1;
187  pi3(1,k) = psi_3_x/psi_1;
189  subplot(2,1,1)
190  plot(pi2(1,1:k))
192  subplot(2,1,2)
193  plot(pi3(1,1:k))
195  x2hat = sum(pi2)/length(pi2);
196  x3hat = sum(pi3)/length(pi3);
198  subplot(2,1,1)
199  plot(x2hat)
200  subplot(2,1,2)
201  plot(x3hat)
203  X2hat = sum(x2hat)/length(x2hat);
204  X3hat = sum(x3hat)/length(x3hat);
207  Ttilde = 26;
208  N = 1000;
209  h = Ttilde/N;
210  X1 = zeros(N+1,L);
211  xi = normrnd(0,1,[N,L]); % N(0,1) to help simulate dB(t)
212  for l = 1:L
213    for n = 1:N
214      X1(n+1,l) = X1(n,l) + X3hat*(X2hat-X1(1,l))*h + xi(n,l)*sqrt(h);
215    end
216  end
222  figure
223  for l = 1:L
224    plot(0:h:Ttilde,X1(:,l),'Color',greycolour);
225    hold('on')
227  end
232  % Simulate Yi(t), i = 1, 2, ..., 5, 0 <= t <= Ttilde (future time interval):
233  mu_ti = zeros(N+1,L); % Jump intensity
236  for i = 1:N
238    mu_ti(i+1,l) = mu_ti(i,l) + (h/2)*lambda(X1(i,l),X2hat,X3hat) + lambda(X1(i+1,l),X2hat,X3hat); %Integration by trapezoidal rule
240  end
242  figure
244  for l = 1:L
245    plot(0:h:Ttilde,mu_ti(:,l),'Color',greycolour);
246    hold('on')
247  end
250  axis([0 Ttilde -inf inf])
```
% Simulate the stochastic coefficients for the Gompertz-Makeham-type model:

```matlab
M = zeros(5, N+1, L);
intensity = zeros(N+1, L);
jumpy = zeros(L);
for l = 1:L
    for i = 1:N
        C = poissrnd(mu_ti(i+1, l) - mu_ti(i, l));
        jumpy(l) = jumpy(l) + C;
        M(1, i+1, l) = M(1, i, l) + sum((10^(-4)) * 0.5 * rand(C, 1));
        M(2, i+1, l) = M(2, i, l) + sum((10^(-5)) * (1.0 + (3.0 - 1.0) * rand(C, 1)));
        M(3, i+1, l) = M(3, i, l) + sum((-1.0 + (1.0 - (-1.0)) * rand(C, 1)));
        M(4, i+1, l) = M(4, i, l) + sum((10^(-6)) * (55815 + (55815 - 43055) * rand(C, 1)));
        M(5, i+1, l) = M(5, i, l) + sum((-10^(-7)) * (1405 + (1405 - 955) * rand(C, 1)));
    end
end

% Plot simulated Y at time \( t = 13 \) versus last year of data:
figure
for l = 1:L
    hold('on');
    plot(mu(age,'Y(1, end), Y(2, end), -20 + Y(3, end), Y(4, end), Y(5, end)), 'k', 'LineWidth', 3);
end
hold('off');

% Plot simulated Y at time \( t = 26 \) versus last year of data:
figure
for l = 1:L
    hold('on');
    plot(mu(age,'M(1,500, l), M(2,500, l), -20 + M(3,500, l), M(4,500, l), M(5,500, l)), 'Color', 'grey_colour');
end
axis([0 111 0 1])
hold('off');
```

C.2. PROGRAMS FOR CHAPTER 6
Bibliography


[9] Human Mortality Database. University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on 01.09.2015).


