

**Modelling of Total Claim Amounts with Regime  
Switching in the Framework of Generalized Cox  
Processes**

**Jonas Christensen**

Master's Thesis, Spring 2018





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

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

## Abstract

The main result in this thesis will be a new stochastic model for total claim amounts in non-life insurance which captures "regime switching" effects of data arising for example from natural disasters, regulatory changes or other "shocks" in non-life insurance markets:

We assume the total claim amount is modelled as

$$S(t) = \sum_{i=1}^{N(t)} X_i,$$

where

$$N(t) = \tilde{N}(\mu^*(t))$$

for a process

$$\mu^*(t) = \int_0^t \lambda^*(s, X_s) ds, \quad t \geq 0,$$

independent of a standard Poisson process  $\tilde{N}(t), t \geq 0$  and  $X_i, i \geq 1$ . The "intensity process"  $\lambda^*(t, X_t), t \geq 0$  is non-negative and depends on an unknown process  $X_t, 0 \leq t \leq T$ , which captures "regime switching" effects of data.

Chapter 2 introduces basic mathematical tools needed for the construction of our new stochastic model and estimation of parameters. Chapter 3 summarizes the most important theory needed concerning Lévy processes. Chapter 4 is arguably the core theory in this thesis; it introduces non-linear filtering theory, which will be heavily utilized in connection with our new model. Our new stochastic model fits into a non-linear filtering framework, which enables the estimation of the signal process  $X_t, t \geq 0$  from observed data. The latter is discussed in detail in Chapter 4. Chapter 5 introduces basic concepts of non-life insurance. Chapter 6 will study the new stochastic model which will be used for our simulations. The simulations at the end of Chapter 6 focus on an observation process with a pure jump component, and the signal process is constructed to capture mean reversion. Chapter 7 will contain extensions and future work to the model developed in Chapter 6.

## **Acknowledgement**

I would like to thank my supervisor Frank Proske, for providing me with a very interesting topic for my thesis, and for the great help and guidance.

I would like to thank my parents and my siblings for the continuous support. I would also like to thank my good friend Tobias Abrahamsen, for the insightful discussions throughout my master's degree. Finally, I would like to thank my girlfriend Maria for her endless encouragement and support.

# Notation

## Abbreviations

a.e. - Almost everywhere.

a.s. - Almost surely, with probability 1.

e.g. - For example.

i.i.d. - Independent and identically distributed.

w.r.t. - With respect to.

## Other notation

$\mathbb{1}_A = \mathbb{1}_A(x) := \begin{cases} 1 & \text{if } x \in A, \\ 0 & \text{if } x \notin A. \end{cases}$  - The indicator function.

$\emptyset$  - Empty set.

$\mathbb{N}$  - The set of all natural numbers.

$\mathbb{R}^n$  - For any natural number  $n$ , the set  $\mathbb{R}^n$  consists of all  $n$ -tuples of real numbers ( $\mathbb{R}$ ).

$\mathbb{R}_{\geq 0}$  - The set of all non-negative real numbers.

$\mathbb{R}^+$  - The set of all non-negative real numbers, with 0 excluded.

$\mathbb{R}^n - \{0\} := \{x \in \mathbb{R}^n, x \neq 0\}$ .

$\mathbb{R}^{d \times m}$  - All  $d \times m$  matrices with real entries.

$\mathbb{R}_0^m$  -  $\mathbb{R}^m$ , with the 0-vector excluded.

$C_b(X; S)$  - The set of all continuous functions  $f : X \rightarrow S$ , that are bounded.

$C(X \times Y; S)$  - The set of all continuous functions  $f : X \times Y \rightarrow S$ .

$\stackrel{d}{=}$  - Equal in distribution.

♣ - End of an example.

□ - End of a proof.

$t \wedge n := \min\{t, n\}$ .

$\xi \sim \Phi$  -  $\xi$  is  $\Phi$ -distributed.

$\|f\|_\infty = \|f\|_{\infty, S} := \sup\{|f(x)| : x \in S\}$  - The supremum norm, where  $f$  is a real- or complex-valued bounded function and  $S$  is a set.

$\langle x, y \rangle$  - The inner product in  $\mathbb{R}^d$ , where  $x = (x_1, x_2, \dots, x_d)$  with each  $x_i \in \mathbb{R}$  and  $y = (y_1, y_2, \dots, y_d)$  with each  $y_i \in \mathbb{R}$  for  $1 \leq i \leq d$ ; that is,  $\langle x, y \rangle = \sum_{i=1}^d x_i y_i$ , where  $x, y \in \mathbb{R}^d$ .

$\langle x, Ay \rangle := \sum_{i,j=1}^d A_j^i x_i y^j$  - The inner product where  $x, y \in \mathbb{R}^d$  and  $A = (A_j^i)$  is a  $d \times d$  matrix.

$\|x\|_1 := \sum_{i=1}^d |x_i|, x \in \mathbb{R}^d$  - The manhattan norm, where  $x$  is a vector.

$L^p(S, \mu)$  - The Banach space of all equivalence classes of mappings  $f : S \rightarrow \mathbb{R}^d$  which agree a.e. (with respect to  $\mu$ ) and for which  $\|f\|_p < \infty$ , where  $\|\cdot\|_p$  denotes the norm

$$\|f\|_p = \left[ \int_S |f(x)|^p \mu(dx) \right]^{1/p}.$$

# Contents

<b>1</b>	<b>Introduction</b>	<b>7</b>
<b>2</b>	<b>Preliminaries/basic mathematical tools</b>	<b>8</b>
2.1	Measure theory . . . . .	8
2.2	Probability Theory . . . . .	9
2.3	Essential stochastic processes . . . . .	11
2.4	Martingale theory . . . . .	14
2.5	Stochastic analysis . . . . .	16
2.5.1	Itô integration and Itô's formula . . . . .	16
2.6	Generalized linear models (GLMs) . . . . .	19
2.6.1	The exponential dispersion family . . . . .	20
2.6.2	The canonical link function of a GLM . . . . .	21
<b>3</b>	<b>Lévy processes</b>	<b>22</b>
3.1	Lévy processes . . . . .	22
3.1.1	Examples of Lévy processes . . . . .	23
3.1.2	Random measures . . . . .	23
3.1.3	Jumps of Lévy processes . . . . .	24
3.1.4	Poisson integration . . . . .	26
3.1.5	The Lévy-Itô decomposition . . . . .	26
3.1.6	Properties of Lévy processes . . . . .	27
3.1.7	Moments of Lévy processes. . . . .	29
3.1.8	Lévy processes and Martingales. . . . .	29
<b>4</b>	<b>(Non-linear) Filtering theory</b>	<b>31</b>
4.1	Non-linear filtering theory - an introduction . . . . .	31
4.1.1	Filtering problems . . . . .	32
4.2	The optimal filter . . . . .	33
<b>5</b>	<b>Non-life insurance</b>	<b>37</b>
5.1	Basic concepts . . . . .	37
5.1.1	Pricing insurance risk. . . . .	37
5.1.2	Solvency . . . . .	38
5.2	Monte Carlo simulation . . . . .	38
5.2.1	Percentiles . . . . .	39
5.2.2	Sampling random variables . . . . .	39
5.3	Modelling claim frequency . . . . .	41
5.3.1	Claim intensities . . . . .	41
5.3.2	Random claim intensities . . . . .	42
5.3.3	The mixed Poisson process. . . . .	43
5.3.4	Poisson regression . . . . .	46
5.4	Modelling claim size . . . . .	47
5.4.1	Introduction . . . . .	47
5.4.2	Parametric modelling . . . . .	48
5.4.3	Non-parametric modelling . . . . .	49

5.4.4	The log-normal family . . . . .	49
5.4.5	The Gamma family . . . . .	50
5.4.6	The Pareto family . . . . .	51
<b>6</b>	<b>Our new stochastic model for total claim amounts</b>	<b>53</b>
6.1	Building our model in the framework of generalized Cox processes . . . . .	53
6.1.1	Choosing the distribution for claim sizes . . . . .	57
6.1.2	A general approach to our model . . . . .	58
6.2	Simulating our model . . . . .	60
6.2.1	Numerical results . . . . .	64
6.3	Conclusion . . . . .	69
<b>7</b>	<b>Extensions</b>	<b>71</b>
7.1	The signal process . . . . .	71
7.1.1	Choosing a different coefficient for $b$ . . . . .	71
7.1.2	Choosing a different coefficient for $\sigma$ . . . . .	71
7.2	The observation process . . . . .	72
7.3	Claim size distribution . . . . .	72
7.4	Calculating the reserve . . . . .	72
<b>8</b>	<b>Appendix A - Calculations</b>	<b>74</b>
8.1	A.1 - Chapter 3 . . . . .	74
8.2	A.2 - Chapter 6 . . . . .	75
<b>9</b>	<b>Appendix B - Distributions</b>	<b>77</b>
9.1	B.1 - The normal distribution . . . . .	77
9.2	B.2 - Poisson distribution . . . . .	77
9.3	B.3 - Gamma distribution . . . . .	77
9.4	B.4 - Pareto distribution . . . . .	77
<b>10</b>	<b>Appendix C - R-code</b>	<b>78</b>
10.1	C.1 - Chapter 2 . . . . .	78
10.2	C.2 - Chapter 5 . . . . .	78
10.3	C.3 - Chapter 6 . . . . .	78
	<b>Bibliography</b>	<b>86</b>



# Chapter 1

## Introduction

An important challenge in the insurance world is accurately modelling the *total claim amount* in insurance portfolios. Existing models in non-life insurance are currently able to capture different phenomenons such as fluctuating properties of the insurance portfolio. As an example, the properties may include driving skill, the age and health state of the individual drivers in a car insurance portfolio.

However, in the world of non-life insurance, there does not exist a model which allows for the modelling of "regime switching" effects. What if a natural disaster occurred? What if regulatory changes completely disrupt the insurance markets? These are examples of "regime switching" effects, and our goal is to study a model which captures such shocks in non-life insurance markets.

Models which capture "regime switching" effects have been examined in Bølviken, Duedahl, Proske [6], applied in the realm of *life insurance*. In this thesis we will study a new stochastic model for the total claim amounts in the realm of *non-life insurance*. A general model for the dynamics of such aggregated claims is presented in Chapter 5, and we will use this general model as a framework when studying our new model.

This thesis is built in such a way that it is beneficial, but not necessary, to have prior knowledge about the theory and tools used throughout the thesis. We will present theory and develop tools that are required for the results as we go along.

The objectives of this thesis are the following:

- Introduce basic mathematical tools needed for the development of our new stochastic model.
- Give an introduction to basic concepts of (non-linear) filtering theory, jump processes (specifically Lévy processes) and stochastic models for total claim amounts.
- Estimation of the unknown parameters/signal process (introduced in Chapter 4) from simulated insurance data, which will be done by using non-linear filtering techniques and Monte-Carlo simulation.
- Allow our new stochastic model to capture "regime switching" effects of data, which are due to "shocks" in non-life insurance markets.

At the end of this thesis we will have developed a new stochastic model that allows for capturing "regime switching" effects of data. In other words, our new model will attempt to accurately predict future values for total claim amounts based on insurance data given in the past, and it will take potential "shocks" from non-life insurance markets into account.

# Chapter 2

## Preliminaries/basic mathematical tools

### 2.1 Measure theory

This section will contain key notions of measure theory and probability that will be used throughout the thesis. We will introduce several important definitions that will be the foundation for problemsolving in future chapters. This chapter is targeted more towards the readers that are not familiar with probability theory and stochastic analysis. The references used in this chapter are Agresti [1], Applebaum [2], Benth [4], McDonald, Weiss [9] and Øksendal [11].

We will begin with introducing basic concepts of measure and probability theory, and then continue with some key properties of stochastic analysis. The properties and concepts listed in this chapter will be used throughout the thesis.

**Definition 2.1.**  $\sigma$ -Algebra of Sets.

Let  $\Omega$  be a non-empty set. A nonempty collection  $\mathcal{H}$  of subsets of  $\Omega$  is called a  $\sigma$ -algebra if the following conditions are satisfied:

- i)  $\Omega \in \mathcal{H}$ .
- ii)  $A \in \mathcal{H}$  implies  $A^c \in \mathcal{H}$ .
- iii)  $\{A_n\}_n \subset \mathcal{H}$  implies  $\bigcup_n A_n \in \mathcal{H}$ .

**Definition 2.2.** Borel  $\sigma$ -algebra.

The Borel  $\sigma$ -algebra of  $\mathbb{R}^d$  is the smallest  $\sigma$ -algebra of subsets of  $\mathbb{R}^d$  that contains all the open sets. We denote it by  $\mathcal{B}(\mathbb{R}^d)$ . If  $S \in \mathcal{B}(\mathbb{R}^d)$  we define its Borel  $\sigma$ -algebra on  $S$  as

$$\mathcal{B}(S) = \{E \cap S; E \in \mathcal{B}(\mathbb{R}^d)\}. \quad (2.1)$$

**Definition 2.3.** Measure, measurable space and measure space.

Let  $\Omega$  be a set and  $\mathcal{H}$  a  $\sigma$ -algebra of subsets of  $\Omega$ . A measure  $\mu$  on  $\mathcal{H}$  is an extended real valued function satisfying the following conditions:

- i)  $\mu(A) \geq 0$  for all  $A \in \mathcal{H}$ .
- ii)  $\mu(\emptyset) = 0$ .
- iii) If  $A_1, A_2, \dots$  are in  $\mathcal{H}$ , with  $A_i \cap A_j = \emptyset$  for  $i \neq j$ , then

$$\mu\left(\bigcup_n A_n\right) = \sum_n \mu(A_n). \quad (2.2)$$

The pair  $(\Omega, \mathcal{H})$  is called a measurable space and the triple  $(\Omega, \mathcal{H}, \mu)$  is called a measure space.

**Definition 2.4.** Null set.

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

The set of all null sets will usually be denoted by  $\mathcal{N}$ .

Let  $(\Omega, \mathcal{H})$  be a measurable space. Let  $(\mathbb{S}, \mathcal{S})$  be another measurable space.

We say that a function  $f : \Omega \rightarrow \mathbb{S}$  is  $\mathcal{H}$ -measurable if

$$f^{-1}(A) = \{w \in \Omega : f(w) \in A\} \in \mathcal{H}, A \in \mathcal{S}. \quad (2.3)$$

## 2.2 Probability Theory

The triple  $(\Omega, \mathcal{F}, P)$  is called a probability space. It is called a complex probability space if  $\mathcal{F}$  contains all subsets  $G$  of  $\Omega$  with  $P$ -outer measure zero; that is, with

$$P * (G) = \inf\{P(F); F \in \mathcal{F}, G \subset F\} = 0. \quad (2.4)$$

Any probability space can be made complete simply by adding to  $\mathcal{F}$  all sets of outer measure 0 and by extending  $P$  accordingly. From now on we will assume that all our probability spaces are complete.

The subsets  $F$  of  $\Omega$  which belong to  $\mathcal{F}$  are called  $\mathcal{F}$ -measurable sets. In a probability context these sets are called events and we use the interpretation

$$P(F) = \text{"The probability that the event } F \text{ occurs"}. \quad (2.5)$$

If  $(\Omega, \mathcal{F}, P)$  is a given probability space, then a function  $Y : \Omega \rightarrow \mathbb{R}^n$  is called  $\mathcal{F}$ -measurable if

$$Y^{-1}(U) := \{w \in \Omega; Y(w) \in U\} \in \mathcal{F} \quad (2.6)$$

for all open sets  $U \in \mathbb{R}^n$ .

A random variable  $X$  is an  $\mathcal{F}$ -measurable function  $X : \Omega \rightarrow \mathbb{R}^n$ .

The following definition is an example of a transformation in probability theory. This kind of transformation can be useful when solving certain types of mathematical problems.

**Definition 2.5.** Characteristic function.

Let  $X$  be a random variable defined on  $(\Omega, \mathcal{F}, P)$  and taking values in  $\mathbb{R}^d$  with probability law  $p_X$ . Its characteristic function  $\phi_X : \mathbb{R}^d \rightarrow \mathbb{C}$  is defined by

$$\phi_X(u) = \mathbb{E}(e^{i\langle u, X \rangle}) = \int_{\Omega} e^{i\langle u, X(w) \rangle} P(dw) = \int_{\mathbb{R}^d} e^{i\langle u, y \rangle} p_X(dy) \quad (2.7)$$

for each  $u \in \mathbb{R}^d$ .

**Definition 2.6.** Absolutely continuous measure.

Let  $(S, \mathcal{F}, \mu)$  be an arbitrary measure space. A measure  $\nu$  on  $(S, \mathcal{F})$  is said to be absolutely continuous with respect to  $\mu$  if  $A \in \mathcal{F}$  and  $\nu(A) = 0 \implies \mu(A) = 0$ . We then write  $\nu \ll \mu$ .

The following theorem is very important in extending the ideas of probability theory from probability masses and probability densities defined over real numbers to probability measures defined over arbitrary sets. It tells if and how it is possible to change from one probability measure to another.

**Theorem 2.1.** Radon-Nikodým.

If  $\mu$  is  $\sigma$ -finite and  $\nu$  is  $\sigma$ -finite with  $\nu \ll \mu$ , then there exists a measurable function  $g : S \rightarrow \mathbb{R}^+$  such that, for each  $A \in \mathcal{F}$ ,

$$\nu(A) = \int_A g(x)\mu(dx). \quad (2.8)$$

The function  $g$  is unique up to  $\mu$ -almost-everywhere equality.

The functions  $g$  appearing in this theorem are sometimes denoted  $d\nu/d\mu$  and called the *Radon-Nikodým derivative of  $\nu$  with respect to  $\mu$* .

Now let  $(\Omega, \mathcal{F}, P)$  be a probability space and  $\mathcal{G}$  be a *sub- $\sigma$ -algebra* of  $\mathcal{F}$ .

Let  $X$  be an  $\mathbb{R}$ -valued random variable with  $\mathbb{E}(|X|) < \infty$ , and for now assume that  $X \geq 0$ . We define a finite measure  $\mathcal{Q}_X = \mathbb{E}(X \mathbb{1}_A)$  for  $A \in \mathcal{G}$ ; then  $\mathcal{Q}_X \ll P$ , and we write

$$\mathbb{E}(X|\mathcal{G}) = \frac{d\mathcal{Q}_X}{dP}. \quad (2.9)$$

We call  $\mathbb{E}(X|\mathcal{G})$  the *conditional expectation* of  $X$  with respect to  $\mathcal{G}$ . It is a random variable on  $(\Omega, \mathcal{G}, P)$  and is uniquely defined up to sets of  $P$ -measure zero.

The concept of conditional expectation is very useful in probability theory, which we will apply later on in this thesis. We will now state Bayes' rule for conditional expectation. When dealing with nonlinear filtering problems (introduced later in this thesis), we will use a formula which is a direct consequence of the Bayes' rule.

**Theorem 2.2.** Bayes' rule.

Let  $\mu$  and  $\nu$  be two probability measures on a measure space  $(\Omega, \mathcal{G})$  such that  $d\nu(w) = f(w)d\mu(w)$  for some  $f \in L^1(\mu)$ . Let  $X$  be a random variable on  $(\Omega, \mathcal{G})$  such that

$$\mathbb{E}_\nu[|X|] = \int_\Omega |X(w)|f(w)d\mu(w) < \infty. \quad (2.10)$$

Let  $\mathcal{H}$  be a  $\sigma$ -algebra,  $\mathcal{H} \subset \mathcal{G}$ . Then

$$\mathbb{E}_\nu[X|\mathcal{H}]\mathbb{E}_\mu[f|\mathcal{H}] = \mathbb{E}[fX|\mathcal{H}] \text{ a.s.} \quad (2.11)$$

*Proof.* See e.g. Lemma 8.6.2 in Øksendal [11]. □



For convenient notation, we note the following:  
Let  $Q$  be another probability measure on  $\mathcal{F}_T$  ( $\mathcal{F}_T$  is defined in Definition 2.17). By Definition 2.6, we know that  $Q$  is absolutely continuous w.r.t.  $P|_{\mathcal{F}_T}$  (the restriction of  $P$  to  $\mathcal{F}_T$ ) if

$$P(H) = 0 \Rightarrow Q(H) = 0 \text{ for all } H \in \mathcal{F}_T. \quad (2.12)$$

By the Radon-Nikodým theorem this occurs if and only if there exists an  $\mathcal{F}_T$ -measurable random variable  $Z_T(w) \geq 0$  such that

$$dQ(w) = Z_T(w)dP(w) \text{ on } \mathcal{F}_T. \quad (2.13)$$

In this case we write

$$\frac{dQ}{dP} = Z_T \text{ on } \mathcal{F}_T \quad (2.14)$$

and we call  $Z_T$  the *Radon-Nikodým derivative* of  $Q$  with respect to  $P$ .

A random variable  $X$  is a function from  $\Omega$  into  $\mathbb{R}$ , the set of real numbers. This means that for each outcome  $w \in \Omega$ ,  $X(w)$  is a real number. We are now ready to define a *stochastic process*.

**Definition 2.7.** Stochastic process.

*A stochastic process  $\{X(t)\}_{t \in [0, T]}$  is a family of random variables parametrized by time  $t$ ; that is, for each given  $t \in [0, T]$ ,  $X(t)$  is a random variable.*

## 2.3 Essential stochastic processes

Two important properties that some classes of stochastic processes 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.

**Property 1:** 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  $(X_s)_{0 \leq s \leq t}$ .

**Property 2:** 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.

An extremely important process, which is used in many different fields such as finance and physics, is 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't be certain of how it evolves over time. In finance, this is typically represented by the evolution of a stock. We can't possibly predict the price of a stock in the future, which is where Brownian motion enters the picture.

**Definition 2.8.** Brownian motion.

*Brownian motion  $B(t)$  is a stochastic process starting at zero; that is,  $B(0) = 0$ , and which satisfies the following three properties:*

1. *Independent increments:* The random variable  $B(t) - B(s)$  is independent of  $(B_u)_{0 \leq u \leq s}$ .
2. *Stationary increments:* The distribution of  $B(t) - B(s)$  for  $t > s \geq 0$  is the same as  $B_{t-s}$ .
3. *Normal increments:* The distribution of  $B(t) - B(s)$  for  $t > s \geq 0$  is normal with expectation 0 and variance  $t - s$ .

**Simulation of Brownian motion:**

Brownian motion is quite simple to simulate. When we want to simulate the paths Brownian motion, we can construct a short algorithm that creates a process which is normally distributed:

---

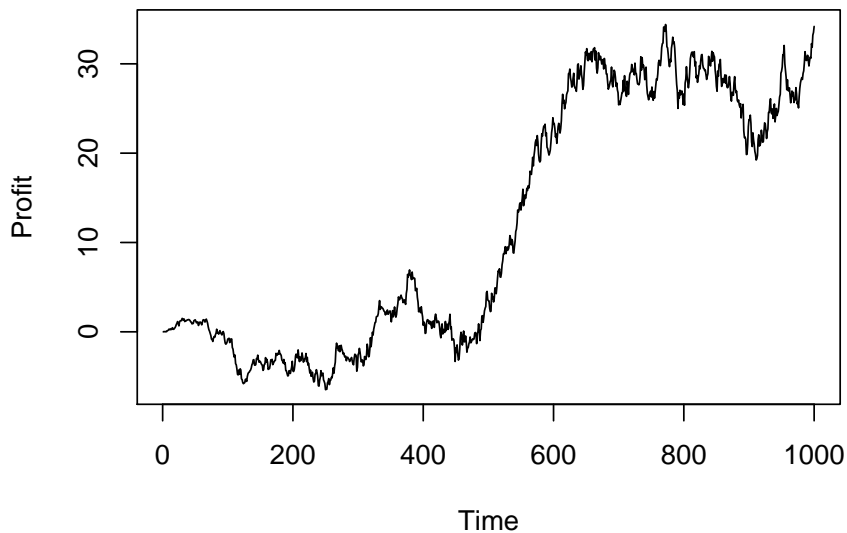
**Algorithm 1** Path of Brownian motion

---

Input: Time horizon  $T$ ; use  $n$  for partitioning

- 1:  $\Delta t \leftarrow T/n$
  - 2: Draw (generate)  $Z_i \sim N(0, 1)$ ,  $i = 0, \dots, n - 1$
  - 3:  $B_0 \leftarrow 0$
  - 4: **for**  $i = 0, \dots, n - 1$  **do**
  - 5:  $B_{t_{j+1}} \leftarrow B_{t_j} + Z_i \sqrt{\Delta t}$
  - 6: **end for**
  - 7: **return**  $(B_{t_j})_{i=0}^n$
- 

**Evolution of a stock price with Brownian motion**



**Figure 1:** Example of a stock price with Brownian motion.

As mentioned before, Brownian motion is very commonly used to model phenomenons such as stocks, as illustrated above. This is mostly due to the fact that Brownian motion is continuous. However, if we wish to model other

things, such as insurance claims, we need a process that does not depend on being continuous. This is because when a new claim occurs, there is a "jump".

This leads us to the Poisson process, which can model jumps or counting. Note that it is a Lévy process, whose definition will be introduced later in the thesis.

**Definition 2.9.** Poisson process.

The Poisson process of intensity  $\lambda$  is a Lévy process  $N$  taking values in  $\mathbb{N} \cup \{0\}$  wherein each  $N(t) \sim \pi(\lambda t)$ , so that we have

$$P(N(t) = n) = \frac{\lambda^n t^n}{n!} e^{-\lambda t} \quad (2.15)$$

for each  $n = 0, 1, 2, \dots$

Another useful process is called the *compound Poisson process*. This process is useful when we want to observe the jumps of a Poisson process, and then choose a distribution for the sizes of the jumps.

**Definition 2.10.** Compound Poisson process.

Let  $(Z(n), n \in \mathbb{N})$  be a sequence of i.i.d. random variables taking values in  $\mathbb{R}^d$  with common law  $\mu_Z$  and let  $N$  be a Poisson process of intensity  $\lambda$  that is independent of all the  $Z(n)$ .

The compound Poisson process  $X$  is defined as follows:

$$X(t) = \begin{cases} 0 & \text{when } N(t) = 0, \\ Z_1 + Z_2 + \dots + Z(N(t)) & \text{when } N(t) > 0, \end{cases}$$

for each  $t \geq 0$ , so each  $X_t \sim \pi(\lambda t, \mu_Z)$ .

We now move on to an even more general class of processes. The generalization of the Poisson process can be defined by introducing a stochastic intensity instead of the deterministic intensity that the regular Poisson process has. The new, generalized, process is called a Cox process.

We define the Cox process as in Bening, Korolov [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

$$\mu_0 = 0 \text{ and } \mathbb{P}[\mu_t < \infty] = 1, 0 \leq t \leq T. \quad (2.16)$$

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

$$N^\mu = (N_t^\mu)_{0 \leq t \leq T} = (N_1(\mu_t))_{0 \leq t \leq T} \quad (2.17)$$

is a Cox process.

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

$$X_t = \sum_{i=1}^{N_t^\mu} Z_i, \quad 0 \leq t \leq T, \quad (2.18)$$

where  $(Z_i)_{i \in \mathbb{N}} \subset \mathbb{R}^d$  is a sequence of i.i.d stochastic variables (see Definition 2.10), and  $N^\mu = (N_t^\mu)_{0 \leq t \leq T} \subset \mathbb{N}_0$  is a Cox process.

## 2.4 Martingale theory

Martingale theory is important with regards to many different subjects. For instance, it can be critical in models of gambling, but most importantly, it is important with regards to mathematical finance and economics. If we were to model the price dynamics of a financial asset as a stochastic (random) process, we demand pricing rules under which the discounted price asset is a martingale. In the context of assets, the martingale property is equivalent to not being able to conduct arbitrage through trades in that asset.

As we will see later on, martingales are essential to stochastic integration.

**Definition 2.12.** Filtration and martingale.

A filtration (on  $(\Omega, \mathcal{F})$ ) is a family  $\mathcal{M} = \{\mathcal{M}_t\}_{t \geq 0}$  of  $\sigma$ -algebras  $\mathcal{M}_t \subset \mathcal{F}$  such that

$$0 \leq s < t \Rightarrow \mathcal{M}_s \subset \mathcal{M}_t \quad (2.19)$$

(that is  $\{\mathcal{M}_t\}$  is increasing). An  $n$ -dimensional stochastic process  $\{M_t\}_{t \geq 0}$  on  $(\Omega, \mathcal{F}, P)$  is called a martingale with respect to a filtration  $\{\mathcal{M}_t\}_{t \geq 0}$  (and with respect to  $P$ ) if

- 1)  $M_t$  is  $\mathcal{M}_t$ -measurable for all  $t$ ,
- 2)  $E[|M_t|] < \infty$  for all  $t$  and
- 3)  $E[M_s | \mathcal{M}_t] = M_t$  for all  $s \geq t$ .

**Definition 2.13.** Adapted with respect to a filtration.

A stochastic process  $X_s$  is called adapted if for each time  $s \in [0, t]$  the random variable  $X_s$  is  $\mathcal{F}_s$ -measurable.

**Example 2.1.** Brownian motion  $B_t$  in  $\mathbb{R}^n$  is a martingale with respect to the  $\sigma$ -algebras  $\mathcal{F}_t$  generated by  $\{B_s; s \leq t\}$ .



*Proof.* We will use the following facts:

- 1)  $B_t$  is  $\mathcal{F}_t$ -measurable (see Theorem B.2.c in Øksendal [11].)
- 2)  $B_s - B_t$  is independent of  $\mathcal{F}_t$  (see (2.2.11) and Theorem B.2.d in Øksendal [11]).

We have:

$$E[|B_t|]^2 \leq E[|B_t|^2] = |B_0|^2 + nt$$



and if  $s \geq t$ , then

$$\begin{aligned}
E[B_s|\mathcal{F}_t] &= E[B_s - B_t + B_t|\mathcal{F}_t] \\
&= E[B_s - B_t|\mathcal{F}_t] + E[B_t|\mathcal{F}_t] \\
&= 0 + B_t \\
&= B_t.
\end{aligned}$$

Here we have used that  $E[(B_s - B_t)|\mathcal{F}_t] = E[B_s - B_t] = 0$  since  $B_s - B_t$  is independent of  $\mathcal{F}_t$ , and that  $E[B_t|\mathcal{F}_t] = B_t$  since  $B_t$  is  $\mathcal{F}_t$ -measurable.  $\square$

**Example 2.2.** Let  $M_t = B_t^2 - t$ . The process  $M_t$  is then a martingale with respect to the filtration  $\mathcal{F}_t$ .

♣

*Proof.* We need to fulfill the three points listed in Definition 2.12, in order to prove that  $M_t$  is a martingale. The first two properties are clearly already fulfilled.

The third one isn't as intuitive, so we need to prove it:

Let  $t > s$ .

$$\begin{aligned}
E[M_t|\mathcal{F}_s] &= E[B_t^2 - t|\mathcal{F}_s] = E[B_t^2|\mathcal{F}_s] - t \\
&\stackrel{(*)}{=} E[(B_t - B_s)^2 - B_s^2 + 2B_tB_s|\mathcal{F}_s] - t \\
&= [(B_t - B_s)^2] - E[B_s^2|\mathcal{F}_s] + 2E[B_tB_s|\mathcal{F}_s] - t \\
&= t - s - B_s^2 + 2B_s^2 - t \\
&= B_s^2 - s \\
&= M_s,
\end{aligned}$$

where (\*) uses the equality  $B_t = B_t - B_s + B_s$ . Now properties 1), 2) and 3) from Definition 2.12 are met, so  $M_t$  is a martingale with respect to the filtration  $\mathcal{F}_t$ .  $\square$

**Definition 2.14.** Stopping times.

A stopping time is a random variable  $T : \Omega \rightarrow [0, \infty]$  for which the event  $\{T \leq t\} \in \mathcal{F}_t$  for each  $t \geq 0$ .

An example of a stopping time is any ordinary deterministic time. Stopping time can be found in many different applications and definitions. An example of such a definition is a local martingale.

**Definition 2.15.** Local martingale.

A local martingale is a measurable adapted process

$$M = (M_t, t \geq 0) \tag{2.20}$$

for which there exists a sequence of stopping times  $\tau_1 \leq \dots \leq \tau_n \rightarrow \infty$  (a.s.) such that each of the processes  $(M(t \wedge \tau_n), t \geq 0)$  is a martingale.

We will in later chapters discuss filtering problems, which will be an essential part of finding the stochastic model for total claim amounts in this thesis. In order to solve filtering problems, we will apply the Girsanov theorem in connection with semimartingales to perform a change of measure. A semimartingale is defined below.

**Definition 2.16.** Semimartingale.

A (continuous) semimartingale with respect to the filtration  $\mathcal{F}_t$  is an  $\mathcal{F}_t$ -adapted process  $X_t, 0 \leq t \leq T$  such that

$$X_t = M_t + A_t, \quad 0 \leq t \leq T \text{ a.e.},$$

where  $M_t, 0 \leq t \leq T$  is a continuous local  $\mathcal{F}_t$ -martingale and  $A_t, 0 \leq t \leq T$  is an  $\mathcal{F}_t$ -adapted continuous bounded variation process.

## 2.5 Stochastic analysis

We have now arrived at stochastic analysis. This chapter will introduce the basics ideas and concepts of stochastic analysis with respect to Brownian motion, and is based on Chapter 5 in Øksendal [11].

### 2.5.1 Itô integration and Itô's formula

A very important part of stochastic analysis is Itô integration. Itô integration here deals with stochastic integrals with respect to Brownian motion; we want to define the integral

$$\int_S^T f(t, w) dB_t(w). \quad (2.21)$$

Before we can define this important integral, we need some definitions.

**Definition 2.17.** The filtration  $\mathcal{F}_t$ .

Let  $B_t(w)$  be  $n$ -dimensional Brownian motion. Then we define  $\mathcal{F}_t = \mathcal{F}_t^{(n)}$  to be the  $\sigma$ -algebra generated by the random variables  $\{B_i(s)\}_{1 \leq i \leq n, 0 \leq s \leq t}$ . In other words,  $\mathcal{F}_t$  is the smallest  $\sigma$ -algebra containing all sets of the form

$$\{w; B_{t_1}(w) \in F_1, \dots, B_{t_k}(w) \in F_k\}, \quad (2.22)$$

where  $t_j \leq t$  and  $F_j \subset \mathbb{R}^n$  are Borel sets,  $j \leq k = 1, 2, \dots$  (We assume that all sets of measure zero are included in  $\mathcal{F}_t$ ).

One often thinks of  $\mathcal{F}_t$  as "the history of  $B_s$  up to time  $t$ ".

**Definition 2.18.**

Let  $\mathcal{V} = \mathcal{V}(S, T)$  be the class of functions

$$f(t, w) : [0, \infty) \times \Omega \rightarrow \mathbb{R} \quad (2.23)$$

such that

- 1)  $(t, w) \rightarrow f(t, w)$  is  $\mathcal{B} \times \mathcal{F}$ -measurable, where  $\mathcal{B}$  denotes the Borel  $\sigma$ -algebra on  $[0, \infty)$ .
- 2)  $f(t, w)$  is  $\mathcal{F}_t$ -adapted.
- 3)  $E[\int_S^T f(t, w)^2 dt] < \infty$ .

For functions  $f \in \mathcal{V}$  we are now ready to define the Itô integral mentioned above.

Let the Itô integral be defined by

$$\mathcal{I}[f](w) = \int_S^T f(t, w) dB_t(w), \quad (2.24)$$

where  $B_t$  is a 1-dimensional Brownian motion.

First, we define  $\mathcal{I}[\phi]$  for a simple class of functions  $\phi$ . Then we show that each  $f \in \mathcal{V}$  can be approximated by such  $\phi$ 's and we use this to define  $\int f dB$  as the limit of  $\int \phi dB$  as  $\phi \rightarrow f$ .

The following definition will be of great use:

**Definition 2.19.** Elementary function.

A function  $\phi \in \mathcal{V}$  is called elementary if it has the form

$$\phi(t, w) = \sum_j e_j(w) \cdot \mathbb{1}_{[t_j, t_{j+1})}(t). \quad (2.25)$$

Note that since  $\phi \in \mathcal{V}$  each function  $e_j$  must be  $\mathcal{F}_{t_j}$ -measurable.

Finally, for elementary functions  $\phi(t, w)$ , we define the integral:

$$\int_S^T \phi(t, w) dB_t(w) = \sum_{j \geq 0} e_j(w) [B_{t_{j+1}} - B_{t_j}](w). \quad (2.26)$$

An important observation for the Itô integral is the Itô isometry.

**Lemma 2.1.** The Itô isometry.

If  $\phi(t, w)$  is bounded and elementary then

$$E \left[ \left( \int_S^T \phi(t, w) dB_t(w) \right)^2 \right] = E \left[ \int_S^T \phi(t, w)^2 dt \right]. \quad (2.27)$$

*Proof.* For proof see Øksendal [11]. □

We will now look at some properties for the Itô integral.

**Theorem 2.3.** Properties of the Itô integral.

Let  $f, g \in \mathcal{V}(0, T)$  and let  $0 \leq S < U < T$ . Then

i)  $\int_S^T f dB_t = \int_S^U f dB_t + \int_U^T f dB_t$  for almost all  $w$ .

ii)  $\int_S^T (cf + g) dB_t = c \cdot \int_S^T f dB_t + \int_S^T g dB_t$  ( $c$  constant) for almost all  $w$ .

iii)  $E[\int_S^T f dB_t] = 0$ .

iv)  $\int_S^T f dB_t$  is  $\mathcal{F}_T$ -measurable.

*Proof.* This clearly holds for all elementary functions, so by taking limits we obtain this for all  $f, g \in \mathcal{V}(0, T)$ .  $\square$

As mentioned earlier, we will use the Girsanov theorem to introduce a change of measure to construct a new Brownian motion under the new measure. We now have enough background information to introduce the Girsanov theorem.

**Theorem 2.4.** The Girsanov theorem.

Let  $Y(t) \in \mathbb{R}^n$  be an Itô process of the form

$$dY(t) = a(t, w)dt + dB(t); t \leq T, Y_0 = 0. \quad (2.28)$$

where  $T \leq \infty$  is a given constant and  $B(t)$  is  $n$ -dimensional Brownian motion.

Put

$$M_t = \exp\left(-\int_0^t a(s, w)dB_s - \frac{1}{2}\int_0^t a^2(s, w)ds\right); 0 \leq t \leq T. \quad (2.29)$$

Assume that  $M_t$  is a martingale with respect to  $\mathcal{F}_t^{(n)}$  and  $P$ . Define the measure  $Q$  on  $\mathcal{F}_T^{(n)}$  by

$$dQ(w) = M_T(w)dP(w). \quad (2.30)$$

Then  $Q$  is a probability measure on  $\mathcal{F}_T^{(n)}$  and  $Y(t)$  is an  $n$ -dimensional Brownian motion with respect to  $Q$ , for  $0 \leq t \leq T$ .

*Proof.* For a detailed proof, see Øksendal [11].  $\square$

**Example 2.3.** Using the Girsanov theorem.

Let  $Y(t) \in \mathbb{R}^n$  be given by

$$dY(t) = g(t)dt + dB(t), \quad 0 \leq t \leq T, \quad (2.31)$$

where  $g : [0, T] \rightarrow \mathbb{R}^n$  is a continuous deterministic function.

Then  $Y(t)$  is a Brownian motion with respect to  $Q$ , where

$$dQ(w) = \exp\left(-\int_0^T g(s)dB(s) - \frac{1}{2}\int_0^T g^2(s)ds\right)dP(w) \text{ on } \mathcal{F}_T^{(n)}. \quad (2.32)$$

♣

The following theorem can be used to solve certain types of stochastic differential equations. In our case, this will specifically be applied to the Vasicek model when creating our new model in Chapter 6.

**Theorem 2.5.** The 1-dimensional Itô formula.

Let  $X_t$  be an Itô process given by

$$dX_t = udt + vdB_t.$$



Let  $g(t, x) \in C^2([0, \infty) \times \mathbb{R})$ . Then

$$Y_t = g(t, X_t)$$

is again an Itô process, and

$$\begin{aligned} dY_t &= \frac{\partial f(t, X_t)}{\partial t} dt + \frac{\partial f(t, X_t)}{\partial x} dX_t \\ &\quad + \frac{1}{2} \frac{\partial^2 f(t, X_t)}{\partial x^2} (dX_t)^2, \end{aligned} \tag{2.33}$$

with the calculation rules  $dt \cdot dt = 0$ ,  $dt \cdot dB_t = dB_t \cdot dt = 0$  and  $(dB_t)^2 = dt$ .

*Proof.* See the proof for Theorem 4.1.2 in Øksendal [11].  $\square$

**Example 2.4.** The process

$$X(t) := X(0)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB(s),$$

where  $X(0)$  is the initial value of  $X(t)$ , has the dynamics

$$dX(t) = -\alpha X(t)dt + \sigma dB(t).$$



*Proof.* Let  $f(t, x) = tx$ . Using Itô's formula, we obtain

$$\begin{aligned} dX(t) &= d\left(X(0)e^{-\alpha t} + e^{-\alpha t} \sigma \int_0^t e^{\alpha s} dB(s)\right) \\ &= e^{-\alpha t} \sigma e^{\alpha t} dB(t) - \alpha e^{-\alpha t} \left(X(0) + \sigma \int_0^t e^{\alpha s} dB(s)\right) dt \\ &= \sigma dB(t) - \alpha \underbrace{\left(X(0)e^{-\alpha t} + \sigma e^{-\alpha t} \int_0^t e^{\alpha s} dB(s)\right)}_{=X(t)} dt \\ &= -\alpha X(t)dt + \sigma dB(t). \end{aligned}$$

$\square$

## 2.6 Generalized linear models (GLMs)

This section will briefly introduce the key concepts in the theory of generalized linear models. When attempting to accurately model the number of incidents that will occur in a given time interval in non-life insurance, one usually turns to the Poisson distribution. In practice however, the Poisson distributions might not be sufficient, as the variance may exceed the mean. We will propose an alternative method to the Poisson distribution in Chapter 5. In order to use the alternative method, we need the basic concepts of what generalized linear models are.

### 2.6.1 The exponential dispersion family

**Definition 2.20.** Random component.

The random component of a GLM consists of a response variable  $y$  with independent observations  $(y_1, \dots, y_n)$  from a distribution having probability density or mass function for  $y_i$  of the form

$$f(y_i; \theta_i, \phi) = \exp \left\{ [y_i \theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi) \right\}. \quad (2.34)$$

Here  $f(y_i; \theta_i, \phi)$  is called the *exponential dispersion family*. The parameter  $\theta_i$  is called the natural parameter, and  $\phi$  is called the dispersion parameter. We can derive  $\mathbb{E}(y_i)$  and  $\text{var}(y_i)$  using the quantities in  $f(y_i; \theta_i, \phi)$ . Let  $L_i = \log f(y_i; \theta_i, \phi)$ . It is clear that

$$L_i = [y_i \theta_i - b(\theta_i)]/a(\phi) + c(y_i, \phi). \quad (2.35)$$

The derivative of  $L_i$  with respect to  $\theta_i$  is now given by

$$\frac{\partial L_i}{\partial \theta_i} = \frac{y_i - b'(\theta_i)}{a(\phi)}, \quad (2.36)$$

and the second derivative with respect to  $\theta$  is given by

$$\frac{\partial^2 L_i}{\partial \theta_i^2} = -\frac{b''(\theta_i)}{a(\phi)}. \quad (2.37)$$

Before deriving the expectation and variance itself, we need the following likelihood results:

$$\mathbb{E}\left(\frac{\partial L}{\partial \theta}\right) = 0 \quad \text{and} \quad -\mathbb{E}\left(\frac{\partial^2 L}{\partial \theta^2}\right) = \mathbb{E}\left(\frac{\partial L}{\partial \theta}\right)^2. \quad (2.38)$$

Further details on why these likelihood results hold can be found in Chapter 4 in Agresti [1].

Applying the likelihood results, we obtain the following:

$$\mathbb{E}[y_i - b'(\theta_i)]/a(\phi) = 0, \quad \text{so that} \quad \mu_i = \mathbb{E}(y_i) = b'(\theta_i). \quad (2.39)$$

Using the results connected to the second derivative of  $L_i$  with respect to  $\theta_i$ , we obtain

$$b''(\theta_i)/a(\phi) = \mathbb{E}\left[(y_i - b'(\theta_i))/a(\phi)\right]^2 = \text{var}(y_i)/[a(\phi)]^2,$$

so that

$$\text{var}(y_i) = b''(\theta_i)a(\phi). \quad (2.40)$$

## 2.6.2 The canonical link function of a GLM

**Definition 2.21.** Linear predictor of a GLM.

For observation  $i, i = 1, \dots, n$ , let  $x_{ij}$  denote the value of explanatory variable  $x_j, j = 1, \dots, p$ . The linear predictor of a GLM relates parameters  $\{\eta\}$  relating to  $\{\mathbb{E}(y_i)\}$  to the explanatory variables  $x_i, \dots, x_p$  using a linear combination of them,

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij}, \quad i = 1, \dots, n. \quad (2.41)$$

The *link function* of a GLM connects the linear predictor and the random component. In other words, a GLM states that a linear predictor

$$\eta_i = \sum_{j=1}^p \beta_j x_{ij}$$

relates to  $\mu_i$  by  $\eta_i = g(\mu_i)$ , for a link function  $g$ .

The link function  $g$  transform the mean  $\mu_i$  to the natural parameter  $\theta_i$  in (2.34) is called the *canonical link*. For the canonical link, the direct relationship

$$\eta_i = g(\mu_i) = \theta_i = \sum_{j=1}^p \beta_j x_{ij}$$

equates the natural parameter to the linear predictor.

# Chapter 3

## Lévy processes

In Chapter 2, we introduced stochastic processes such as the Poisson process, the compound Poisson process and Brownian motion. As it turns out, these processes are a subset of a more general class of processes, called the class of Lévy processes, which allows for processes with jumps.

This is of very great interest, as it has a wide field of real world applications. Insurance and finance are great examples of this. When modelling the number of claims for insurance, a continuous process does not reflect the real world. In the world of insurance, once a claim occurs, you get a "jump" in the process. This is because accidents occur suddenly and unexpectedly and may take place at any point in time.

In finance, a stock may look continuous when observing the historical prices, but sudden changes may happen at any point in time. It may increase or decrease tremendously in a small amount of time. The classic way of modelling the price of a stock is by the Black-Scholes model. In the Black-Scholes model, one assumes that the stock price evolves continuously, where the volatility is random and moves according to Brownian motion.

In this chapter, we will look at Lévy processes, which is a more general type of stochastic process, that will allow to capture sudden and unexpected movements. The references used in this chapter are Applebaum [2], Korolev [3] and Cont, Tankov [7].

### 3.1 Lévy processes

**Definition 3.1.** Lévy process.

Let  $X$  be a stochastic process. We say that  $X$  is a Lévy process if:

- (L1)  $X(0) = 0$  (a.s.);
- (L2)  $X$  has independent and stationary increments (see properties 1 and 2 in Section 2.3);
- (L3)  $X$  is stochastically continuous; that is, for all  $a > 0$  and for all  $s \geq 0$

$$\lim_{t \rightarrow s} P(|X(t) - X(s)| > a) = 0. \quad (3.1)$$

Note that in the presence of (L1) and (L2), (L3) is equivalent to the condition

$$\lim_{t \downarrow 0} P(|X(t)| > a) = 0. \quad (3.2)$$

for all  $a > 0$ .

Note that by (L3) in Definition 3.1, it states that  $X$  is stochastically continuous. This means that the probability of a discontinuity (jump) occurring at a given point in time is always zero, which again means that all discontinuities of Lévy processes occur at random times.

### 3.1.1 Examples of Lévy processes

**Example 3.1.** Brownian motion.

A (standard) Brownian motion in  $\mathbb{R}^d$  is a Lévy process  $B = (B(t), t \geq 0)$  for which

- (B1)  $B(t) \sim N(0, tI)$  for each  $t \geq 0$ ,
- (B2)  $B$  has continuous sample paths.

It follows immediately from (B1) that if  $B$  is a standard Brownian motion then its characteristic function is given by

$$\phi_{B(t)}(u) = \exp\left(-\frac{1}{2}t|u|^2\right) \quad (3.3)$$

for each  $u \in \mathbb{R}^d, t \geq 0$ . ♣

**Example 3.2.** The Poisson process.

The Poisson process in Definition 2.9 is a Lévy process.

*Claim:* The compound Poisson process (Definition 2.10) is a Lévy process.

*Proof of claim.* See Proposition 1.3.11 in Applebaum [2]. ♣

### 3.1.2 Random measures

**Definition 3.2.** Ring of subsets.

Let  $S$  be a set and  $\mathcal{A}$  be a ring of subsets of  $S$ ; that is,  $\emptyset \in \mathcal{A}, A \cup B \in \mathcal{A}$  and  $A - B \in \mathcal{A}$  (where  $A - B = A \cap B^c$ ). If  $A, B \in \mathcal{A}$ , we have  $A \cap B \in \mathcal{A}$  since  $A \cap B = A - (A - B)$ . Clearly if  $\mathcal{F}$  is a  $\sigma$ -algebra then it is also a ring.

**Definition 3.3.** Random measure.

Let  $(\Omega, \mathcal{F}, P)$  be a probability space. A random measure  $M$  on  $(S, \mathcal{A})$  is a collection of random variables  $(M(B), B \in \mathcal{A})$  such that:

- (i)  $M(\emptyset) = 0$ ;
- (ii) (finite additivity). Given any disjoint  $A, B \in \mathcal{A}$ ,

$$M(A \cup B) = M(A) + M(B). \quad (3.4)$$

A random measure is said to be  $\sigma$ -additive if (ii) can be strengthened to (ii)'.

- (ii)' ( $\sigma$ -additivity). Given any sequence  $(A_n, n \in \mathbb{N})$  of mutually disjoint sets in  $\mathcal{A}$  which are such that  $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$ ,

$$M\left(\bigcup_{n \in \mathbb{N}} A_n\right) = \sum_{n \in \mathbb{N}} M(A_n). \quad (3.5)$$


**Definition 3.4.** Independent scattering.

A random measure is said to be independently scattered if for each disjoint family  $\{B_1, \dots, B_n\}$  in  $\mathcal{A}$ , the random variables  $M(B_1), \dots, M(B_n)$  are independent.

**Example 3.3.**

Let  $X = (X(t), t \geq 0)$  be a Lévy process and choose  $S = [0, T]$  for some  $T > 0$ . Take  $\mathcal{A}$  to be the smallest ring that contains all finite unions of disjoint intervals in  $S$ . These intervals may be open, closed or half-open so that  $\mathcal{A}$  also contains isolated points. If  $A = (s_1, t_1) \cup \dots \cup (s_n, t_n)$ , define

$$M(A) = \sum_{j=1}^n X(t_j) - X(s_j), \quad (3.6)$$

with  $M(\{t\}) = 0$  if  $t \in [0, T]$ . Then  $M$  is an independently scattered random measure on  $(S, \mathcal{A})$ . 

We are now ready to define a Poisson random measure, which is an important tool for modelling jump behavior. It will be used relatively often in this thesis.

**Definition 3.5.** Poisson random measure.

Let  $\mathcal{S}$  be a  $\sigma$ -algebra of subsets of  $S$ . Fix a non-trivial ring  $\mathcal{A} \subseteq \mathcal{S}$ . An independently scattered  $\sigma$ -additive random measure  $M$  on  $(S, \mathcal{S})$  is called a Poisson random measure if  $M(B) < \infty$  for each  $B \in \mathcal{A}$  and each such  $M(B)$  has a Poisson distribution.

In many cases, the prescription  $\lambda(A) = \mathbb{E}(M(A))$  for all  $A \in \mathcal{A}$  extends to a  $\sigma$ -finite measure  $\lambda$  on  $(S, \mathcal{S})$ .

**Example 3.4.**

Let  $U = \mathbb{R}^d - \{0\}$  and  $\mathcal{U}$  be its Borel  $\sigma$ -algebra. Let  $\mathcal{A}$  be the ring of all sets in  $\mathcal{U}$  which are bounded below. Let  $X$  be a Lévy process; then  $\Delta X$  is a Poisson point process and  $N$  is its associated Poisson random measure. For each  $t \geq 0$  and  $A$  bounded below, we define the *compensated Poisson random measure* by

$$\tilde{N}(t, A) = N(t, A) - t\mu(A). \quad (3.7)$$

**3.1.3 Jumps of Lévy processes**

As explained in the introduction to this chapter, it is very useful being able to model jumps (discontinuities, unexpected and surprising movement) when working with insurance. This subsection lays the foundation for being able to model discontinuous movements, which will be introduced in the next subsection.

**Definition 3.6.** Jump size of a Lévy process.

Let  $X = X(t)$  be a Lévy process. The jump at time  $t$  is defined by

$$\Delta X(t) = X(t) - X(t^-), \quad (3.8)$$

where  $X(t^-)$  is the left limit of  $X(t)$ .

**Definition 3.7.** Jump measure of a Lévy process.

Let  $0 \leq t < \infty$  and  $A \in \mathcal{B}(\mathbb{R}^d - \{0\})$ . The jump measure of a Lévy process  $X(t)$  is given by

$$N(t, A) = \sum_{0 \leq s \leq t} \mathbb{1}_A(\Delta X(s)). \quad (3.9)$$

**Definition 3.8.** Lévy measure.

Let  $\nu$  be a Borel measure defined on  $\mathbb{R}^d - \{0\} = \{x \in \mathbb{R}^d, x \neq 0\}$ . We say that it is a Lévy measure if

$$\int_{\mathbb{R}^d - \{0\}} (|y|^2 \wedge 1) \nu(dy) < \infty. \quad (3.10)$$

Note: We say that  $A \in \mathcal{B}(\mathbb{R}^d - \{0\})$  is bounded below if  $0 \notin \bar{A}$ .

**Example 3.5.** Jump measure of a Poisson process.

The jump measure of the Poisson process (Definition 2.9) is given by  $J_N = \sum_{n \geq 1} \delta_{(T_n, 1)}$ :

$$J_N([0, t] \times A) = \begin{cases} \#\{i \geq 1, T_i \in [0, t]\} & \text{if } 1 \in A, \\ 0 & \text{if } 1 \notin A. \end{cases}$$

♣

Next we will look at the finite variation of a Lévy process. Before we can do this, we need to define total variation.

The total variation of a function  $f : [a, b] \rightarrow \mathbb{R}^d$  is defined by

$$\sup \sum_{i=1}^n |f(t_i) - f(t_{i-1})|, \quad (3.11)$$

where the supremum is taken over all finite partitions  $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$  of the interval  $[a, b]$ .

**Proposition 3.1.** Finite variation Lévy process.

A Lévy process is of finite variation if and only if its characteristic triplet  $(A, \nu, \gamma)$  satisfies:

$$A = 0 \text{ and } \int_{|x| \leq 1} |x| \nu(dx) < \infty. \quad (3.12)$$

*Proof.* See Proposition 3.9 in Cont, Tankov [7]. □

**Theorem 3.1.**

(i) If  $A$  is bounded below, then  $(N(t, A), t \geq 0)$  is a Poisson process with intensity  $\mu(A)$ .

(ii) If  $A_1, \dots, A_m \in \mathcal{B}(\mathbb{R}^d - \{0\})$  are disjoint and bounded below and if  $s_1, \dots, s_m \in \mathbb{R}^+$  are distinct, then the random variables  $N(s_1, A_1), \dots, N(s_m, A_m)$  are independent.

*Proof.* See Applebaum [2]. □

### 3.1.4 Poisson integration

When modelling jump behavior, it is very common to use integrals with respect to a random measure. In particular, when working with Lévy processes, one applies integrals with respect to Poisson random measures.

Let  $N$  be the Poisson random measure associated to a Lévy process  $X = (X(t), t \geq 0)$ .

Let  $f$  be a Borel measurable function from  $\mathbb{R}^d$  to  $\mathbb{R}^d$  and let  $A$  be bounded below; then for each  $t > 0$ ,  $\omega \in \Omega$ , we may define the *Poisson integral* of  $f$  as a random finite sum by

$$\int_A f(x)N(t, dx)(\omega) = \sum_{x \in A} f(x)N(t, \{x\})(\omega). \quad (3.13)$$

Now, since  $N(t, \{x\}) \neq 0 \iff X(u) = x$  for at least one  $0 \leq u \leq t$ , we have

$$\int_A f(x)N(t, dx) = \sum_{0 \leq u \leq t} f(\Delta X(u)) \mathbb{1}_A(\Delta X(u)). \quad (3.14)$$

#### Theorem 3.2.

Let  $A$  be bounded below. Then:

(i) for each  $t \geq 0$ ,  $\int_A f(x)N(t, dx)$  has compound Poisson distributions such that, for each  $u \in \mathbb{R}^d$ ,

$$\mathbb{E} \left( \exp \left[ i \left\langle u, \int_A f(x)N(t, dx) \right\rangle \right] \right) = \exp \left[ t \int_{\mathbb{R}^d} (e^{i \langle u, x \rangle} - 1) \mu_{f,A}(dx) \right], \quad (3.15)$$

where  $\mu_{f,A}(B) = \mu(A \cap f^{-1}(B))$ , for each  $B \in \mathcal{B}(\mathbb{R}^d)$ .

(ii) If  $f \in L^1(A, \mu_A)$ , we have

$$\mathbb{E} \left( \int_A f(x)N(t, dx) \right) = t \int_A f(x) \mu(dx). \quad (3.16)$$

(iii) If  $f \in L^2(A, \mu_A)$ , we have

$$\text{Var} \left( \int_A f(x)N(t, dx) \right) = t \int_A |f(x)|^2 \mu(dx). \quad (3.17)$$

*Proof.* See Applebaum [2]. □

### 3.1.5 The Lévy-Itô decomposition

There's a very useful theorem showing that a Lévy process can be decomposed into a drift component, a Brownian motion, a compound Poisson process, and a jump process which is an  $L^2$ -martingale.



**Theorem 3.3.** The Lévy-Itô decomposition.

If  $X$  is a Lévy process, then there exists  $b \in \mathbb{R}^d$ , a Brownian motion  $B_A$  with covariance matrix  $A$  and an independent Poisson random measure  $N$  on  $\mathbb{R}^+ \times (\mathbb{R}^d - \{0\})$  such that, for each  $t \geq 0$ ,

$$X(t) = bt + B_A(t) + \int_{|x|<1} x\tilde{N}(t, dx) + \int_{|x|\geq 1} xN(t, dx). \quad (3.18)$$

*Proof.* See Theorem 2.4.16 in Applebaum [2]. □

In Theorem 3.3,

$$B_A(t) = (B_A^1(t), \dots, B_A^d(t)) \quad (3.19)$$

can also be written on the form

$$B_A^i(t) = \sum_{j=1}^m \sigma_j^i B^j(t), \quad (3.20)$$

where  $B^1, \dots, B^m$  are standard one-dimensional Brownian motions and  $\sigma$  is a  $d \times m$  real valued matrix for which  $\sigma\sigma^T = A$ .

Consider the Lévy process  $X = (X_t)_{0 \leq t \leq T} \subset \mathbb{R}^d$ . The jumps of  $X$  has finite variations. Using Theorem 3.3, we see that  $X$  can be decomposed in this way:

$$X_t = bt + \sigma B_t + \int_{|x|\geq 1} xN(t, dx) = bt + \sigma B_t + \sum_{0 \leq s \leq t} \Delta X(s), \quad 0 \leq t \leq T, \quad (3.21)$$

where  $b \in \mathbb{R}^d$  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  $N(t, dx)$  is the jump measure of  $X$ .

**Example 3.6.** Let  $X = (X_t)_{0 \leq t \leq T}$  be a compound Poisson process where  $X \in \mathbb{R}$  and Lévy measure  $\nu$ . If  $N(t, dx), 0 \leq t \leq T, x \in \mathbb{R}_0^d$  is the jump measure of  $X$ , then

$$X_t = \sum_{0 \leq s \leq t} \Delta X_s = \int_{|x|\geq 1} xN(t, dx), \quad 0 \leq t \leq T, \quad (3.22)$$

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

### 3.1.6 Properties of Lévy processes

**Definition 3.9.** Infinite divisibility.

A probability distribution  $F$  on  $\mathbb{R}^d$  is said to be infinitely divisible if for any integer  $n \geq 2$ , there exists  $n$  i.i.d. random variables  $Y_1, \dots, Y_n$  such that  $Y_1 + \dots + Y_n$  has distribution  $F$ .

*Claim:* Any Lévy process  $X$  has an infinitely divisible distribution.

*Proof of claim:* Recall the information from Definition 3.1. Using the properties of stationary and independent increments, we acquire the following:

Let  $X = (X_t)_{0 \leq t \leq T}$  be a Lévy process with values in  $\mathbb{R}^d$ . Let  $(0 = t_0 < t_1 < \dots < t_n = t)$  be the partitioning of  $[0, t], t \leq T$ . Put  $t_j = \frac{tj}{n}$  for  $0 \leq j \leq n$ , and where  $n \geq 2$  is an integer. We then have

$$X_t = X_t - X_0 = \sum_{j=0}^{n-1} (X_{t_{j+1}} - X_{t_j}). \quad (3.23)$$

Let  $Y_{j+1}^{(n)} = X_{t_{j+1}} - X_{t_j}, 0 \leq j \leq n-1$ . Using the propositionerties of stationary and independent increments of Lévy processes, we see that  $(Y_j^{(n)})$  are i.i.d. and we have

$$X_t = \sum_{j=1}^n Y_j^{(n)}. \quad (3.24)$$

□

**Corollary 3.1.**

If  $X$  is a Lévy process then for each  $u \in \mathbb{R}^d, t \geq 0$ ,

$$\begin{aligned} \mathbb{E}(e^{i\langle u, X(t) \rangle}) &= \exp \left( \left\{ i\langle b, u \rangle - \frac{1}{2} \langle u, Au \rangle \right. \right. \\ &\quad \left. \left. + \int_{\mathbb{R}^d - \{0\}} [e^{i\langle u, y \rangle} - 1 - i\langle u, y \rangle \mathbb{1}_B(y)] \mu(dy) \right\} \right). \end{aligned}$$

*Proof.* See Corollary 2.4.20 in Applebaum [2].

□

This function is known as the characteristic function of a Lévy process, and it is actually an exponential function.

That is, the function  $\phi_t(u) = \mathbb{E}[\exp\{i\langle u, X(t) \rangle\}], u \in \mathbb{R}^d, t \geq 0$  is an exponential function.

Now we will state the Lévy-Khintchine theorem, which gives a characterisation of infinitely divisible random variables through their characteristic functions.

**Theorem 3.4.** Lévy-Khintchine.

$\mu \in \mathcal{M}_1(\mathbb{R}^d)$  is infinitely divisible if there exists a vector  $b \in \mathbb{R}^d$ , a positive definite symmetric  $d \times d$  matrix  $A$  and a Lévy measure  $\nu$  on  $\mathbb{R}^d - \{0\}$  such that, for all  $u \in \mathbb{R}^d$ ,

$$\begin{aligned} \phi_\mu(u) &= \exp \left( \left\{ i\langle b, u \rangle - \frac{1}{2} \langle u, Au \rangle \right. \right. \\ &\quad \left. \left. + \int_{\mathbb{R}^d - \{0\}} [e^{i\langle u, y \rangle} - 1 - i\langle u, y \rangle \mathbb{1}_{\hat{B}}(y)] \nu(dy) \right\} \right), \end{aligned}$$

where  $\hat{B} = B_1(0)$ . Conversely, any mapping of the form of  $\phi_\mu(u)$  is the characteristic function of an infinitely divisible probability measure on  $\mathbb{R}^d$ .

### 3.1.7 Moments of Lévy processes.

In this section we will look at the moments of Lévy processes, which will be a useful tool for the final section of this chapter. We will give necessary and sufficient conditions for a Lévy process  $X = (X(t), t \geq 0)$  to have a finite moment.

Recall from Section 2.3 that if  $Y$  is a compound Poisson variable, then

$$Y = W_1 + \cdots + W_N, \quad (3.25)$$

where  $(W_n, n \in \mathbb{N})$  is a sequence of i.i.d. random variables and  $N$  is an independent Poisson random variable of intensity  $\lambda > 0$ .

**Lemma 3.1.**

*If  $Y$  is a compound Poisson random variable then for each  $n \in \mathbb{N}$ ,  $\mathbb{E}(|Y|^n) < \infty$  if and only if  $\mathbb{E}(|W|^n) < \infty$ .*

*Proof.* See Lemma 2.5.1 in Applebaum [2]. □

**Theorem 3.5.**

*If  $X$  is a Lévy process and  $n \in \mathbb{N}$ ,  $\mathbb{E}(|X(t)|^n) < \infty$  for all  $t > 0$  if and only if  $\int_{|x| \geq 1} |x|^n \nu(dx) < \infty$ .*

*Proof.* See Theorem 2.5.2 in Applebaum [2]. □

### 3.1.8 Lévy processes and Martingales.

Recalling from Section 2.4, we know that martingale theory is crucial when applying stochastic processes to real world, practical problems. This section will give a summary of how Lévy processes are connected to martingales. The information in this subsection is mainly borrowed from Cont, Tankov [7].

**Proposition 3.2.**

*Let  $X = (X_t)_{t \geq 0}$  be a real-valued process with independent increments. Then*

(i)  $\left( \frac{e^{iuX_t}}{\mathbb{E}[e^{iuX_t}]} \right)_{t \geq 0}$  is a martingale  $\forall u \in \mathbb{R}$ .

(ii) If for some  $u \in \mathbb{R}$ ,  $\mathbb{E}[e^{uX_t}] < \infty \forall t \geq 0$  then  $\left( \frac{e^{uX_t}}{\mathbb{E}[e^{uX_t}]} \right)_{t \geq 0}$  is a martingale.

(iii) If  $\mathbb{E}[X_t] < \infty \forall t \geq 0$  then  $M_t = X_t - \mathbb{E}[X_t]$  is a martingale.

(iv) If  $\text{Var}[X_t] < \infty \forall t \geq 0$  then  $(M_t)^2 - \mathbb{E}[(M_t)^2]$  is a martingale, where  $M$  is the martingale defined in (iii).

*If  $(X_t)$  is a Lévy process, for all of the processes of this proposition to be martingales it suffices that the corresponding moments be finite for one value of  $t$ .*

*Proof.* Proof is located in Appendix A. □

**Proposition 3.3.**

Let  $(X_t)_{t \geq 0}$  be a Lévy process on  $\mathbb{R}$  with characteristic triplet  $(A, \nu, \gamma)$ .

(i)  $(X_t)$  is a martingale if and only if  $\int_{|x| \geq 1} |x| \nu(dx) < \infty$  and

$$\gamma + \int_{|x| \geq 1} x \nu(dx) = 0. \tag{3.26}$$

(ii)  $e^{(X_t)}$  is a martingale if and only if  $\int_{|x| \geq 1} e^x \nu(dx) < \infty$  and

$$\frac{A}{2} + \gamma + \int_{-\infty}^{\infty} (e^x - 1 - x \mathbb{1}_{|x| \leq 1}) \nu(dx) = 0. \tag{3.27}$$

Proposition 3.3 was borrowed from Proposition 3.18 in Cont, Tankov [7].

# Chapter 4

## (Non-linear) Filtering theory

This chapter will give an introduction to the basic concepts of non-linear filtering theory. Our ultimate goal in this thesis is to model total claim amounts in non-life insurance. In order to accomplish this, a good way of estimating the process  $X_t, 0 \leq t \leq T$  from insurance data is to apply non-linear filtering techniques for Lévy processes. This chapter will give further details about this technique.

In addition, we will mention Monte Carlo when searching for the best possible estimate  $\hat{X}_t$  of the signal  $X_t$ . Essentially, Monte Carlo methods are a broad class of computational algorithms that rely on repeated random sampling to obtain numerical results. Monte Carlo is studied in detail in Chapter 5. The references used in this chapter are Applebaum [2], Bølviken, Duedahl, Proske [6], Øksendal [11] and Xiong [12].

### 4.1 Non-linear filtering theory - an introduction

When looking at stochastic processes, one often uses information collected from the past to make a prediction on what will happen in the future. This occurs constantly in the world of stocks, where one might construct a model in order to obtain financial profit.

However, this requires that the information from the past is accurate. How would one go about constructing a good predictive model if the information from the past is inaccurate?

Suppose we are looking at a random system that evolves over time. In this random system we know that there exists a process in which we are interested in. This might be stocks as mentioned above, or something different entirely, as long as it contains a random element. Our problem arises when we try to examine this process, as we are unable to observe the process directly. We can, however, collect partial observations of the process. In other words, we observe a distorted version of the real process, and we wish to replicate the real process as accurately as possible, with the information available.

There are two key processes we will consider when dealing with filtering problems. These processes are defined on a probability space  $(\Omega, \mathcal{F}_t, \mathbb{P})$ .

The two processes are:

1) The *observation process*.

The observation process is the process where we have known information. The information about the process may or may not be very distorted, but it is the only information we have about the process nonetheless.

The observation process will be denoted by  $Y_t, 0 \leq t \leq T$ , where  $Y$  takes

values in  $\mathbb{R}^m$ .

2) The *signal process*.

This is the process that we wish to estimate. We will estimate this process by extracting information from the observation process  $Y_t$ . The signal process will be denoted by  $X_t$ ,  $0 \leq t \leq T$ , where  $X$  takes values in  $\mathbb{R}^d$ .

Let's unpack the idea behind filtering problems with an example:

**Example 4.1.**

Wildlife tracking is a process where scientific researchers can remotely observe patterns of an animal using GPS. Consider the tracking of e.g. a bear, where the purpose is to study the behavior and patterns of said bear. The GPS tracker attached to the bear is unfortunately of poor quality, and does not represent its accurate position. This is a classic example of a filtering problem. Here the movement from the GPS tracker is represented as the observation process  $Y_t$ ,  $0 \leq t \leq T$ . The goal is to replicate the exact coordinates of the bear at time  $t \geq 0$ . This is done by collecting information from the GPS; that is, looking at the observation process  $Y_t$ . The information known is given by the filtration  $\mathcal{F}_t^Y$ . ♣

**4.1.1 Filtering problems**

The signal process  $X_t, 0 \leq t \leq T$  follows the dynamics given by the stochastic differential equation:

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t^X, \quad 0 \leq t \leq T, \quad (4.1)$$

where  $b, \sigma$  are Borel functions and  $B_t^X, 0 \leq t \leq T$  is a Brownian motion. The signal process  $X_t$  is not directly observable, hence we must turn to the observation process  $Y_t$  to acquire information about the signal  $X$ .

The observation process  $Y_t$  is described by the dynamics:

$$dY_t = h(t, X_t)dt + dB_t^Y + \int_{\mathbb{R}^m} zN_\lambda(dt, dz), \quad 0 \leq t \leq T, \quad (4.2)$$

where  $h$  is a Borel function,  $B_t^Y$  is a Brownian motion and  $N_\lambda$  is the jump measure of a generalized Cox process with a predictable compensator  $\hat{\mu}$  given by

$$\hat{\mu}(dt, dz, \omega) = \lambda(t, X_t, z)dt\nu(dz) \quad (4.3)$$

for a Lévy measure  $\nu$  and Borel function  $\lambda$ .

Consider the (complete filtered) probability space  $(\Omega, \mathcal{F}, \mathbb{F} := (\mathcal{F}_{0 \leq t \leq T}), \mathbb{P})$ , where  $T < \infty$ .

Let  $\mathbb{F}^X := (\mathcal{F}_t^X)_{0 \leq t \leq T}$  be the natural filter of the process  $X$ , and  $\mathbb{F}^Y := (\mathcal{F}_t^Y)_{0 \leq t \leq T}$  be the natural filter of the process  $Y$ .

One of our goals in this thesis is to estimate the signal  $X$  to the best of our

ability. In other words, we want to find the best estimate of  $X$  at a given time  $t \leq T$ , given the information (observations) in  $\mathcal{F}_t^Y$ .

How exactly do we do this? We find the least square estimate to the (possibly transformed) signal process  $X_t$ ; that is, we determine the conditional expectation

$$\mathbb{E}[f(X_t)|\mathcal{F}_t^Y],$$

where  $f$  is a given Borel function and where  $\mathcal{F}_t^Y$  is the  $\sigma$ -algebra generated by  $\{Y_s, 0 \leq s \leq t \leq T\}$ .

In order to guarantee a unique strong solution to the systems (4.1) and (4.2), we require that the continuous coefficients  $b : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\sigma : \mathbb{R}^n \rightarrow \mathbb{R}^{n \times n}$ ,  $h : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R}^n$  and  $\lambda : [0, T] \times \mathbb{R}^n \times \mathbb{R}_0^m \rightarrow \mathbb{R}$  fulfill a linear growth and Lipschitz condition; that is,

$$\|b(x)\| + \|\sigma(x)\| + \|h(t, x)\| + \int_{\mathbb{R}_0^m} |\lambda(t, x, z)|\nu(dz) \leq C(1 + \|x\|) \quad (4.4)$$

and

$$\begin{aligned} & \|b(x) - b(y)\| + \|\sigma(x) - \sigma(y)\| + \|h(t, x) - h(t, y)\| \\ & + \int_{\mathbb{R}_0^m} |\lambda(t, x, z) - \lambda(t, y, z)|\nu(dz) \\ & \leq C\|x - y\| \end{aligned} \quad (4.5)$$

for all  $x, y, t$  and a constant  $C < \infty$ , where  $\|\cdot\|$  stands for a vector or matrix norm.

## 4.2 The optimal filter

As mentioned in Section 4.1, our aim is to determine the expectation  $\mathbb{E}[f(X_t)|\mathcal{F}_t^Y]$ . We will now look at how the theory behind how this can be determined.

Denote by  $\pi_t : \Omega \times \mathcal{B}(\mathbb{R}^n) \rightarrow [0, \infty)$  the regular conditional probability measure of the signal process  $X_t$  given the  $\sigma$ -algebra  $\mathcal{F}_t^Y$ , generated by  $\{Y_s, 0 \leq s \leq t\}$  and the null sets  $\mathcal{N}$ . Then

$$\mathbb{E}[f(X_t)|\mathcal{F}_t^Y] = \langle \pi_t, f \rangle$$

for all  $f \in C_b(\mathbb{R}^n)$ , where  $\langle \pi_t, f \rangle := \int_{\mathbb{R}^n} f(x)\pi_t(\omega, dx)$ .

Suppose that the function  $\lambda : [0, T] \times \mathbb{R}^n \times \mathbb{R}_0^m \rightarrow \mathbb{R}$  is strictly positive and consider the density process

$$\begin{aligned} \Lambda_t := & \left\{ \exp \sum_{i=1}^m \int_0^t -h_i(s, X_s) dB_s^{Y,i} - \frac{1}{2} \int_0^t \|h(s, X_s)\|^2 ds \right. \\ & + \int_0^t \int_{\mathbb{R}_0^m} -\log \lambda(s, X_s, z) N_\lambda(ds, dz) \\ & \left. + \int_0^t \int_{\mathbb{R}_0^m} (\lambda(s, X_s, z) - 1) ds \nu(dz) \right\}, \quad 0 \leq t \leq T, \end{aligned} \quad (4.6)$$

where  $B_t^Y = (B_s^{Y,1}, \dots, B_s^{Y,m})^*$  and  $h(t, x) = (h_1(t, x), \dots, h_m(t, x))^*$  (\* transposition). Further, assume that

$$\mathbb{E}[\Lambda_T] = 1. \quad (4.7)$$

**Remark 1.** The following conditions are sufficient for (4.7) to hold (by Bølviken, Duedahl, Proske [6]):

$$\sup_{0 \leq t \leq T} \mathbb{E} \left[ \exp \left( 6 \int_0^t \|h(s, X_s)\|^2 ds \right) \right] \quad (4.8)$$

$$+ 4 \int_0^t \int_{\mathbb{R}_0^m} (1 - \lambda^{-1}(s, X_s, z)) \lambda(s, X_s, z) d\nu(dz) \\ - \int_0^t \int_{\mathbb{R}_0^m} (1 - \lambda^{-4}(s, X_s, z)) \lambda(s, X_s, z) \lambda(s, X_s, z) d\nu(dz) \Big]$$

$< \infty$

$$\mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_0^m} |(\lambda^{-4}(s, X_s, z) - 1) \lambda(s, X_s, z)| \nu(dz) ds \right] \quad (4.9)$$

$$+ \mathbb{E} \left[ \int_0^T \left( \int_{\mathbb{R}_0^m} |(\lambda(s, X_s, z) - 1)| \nu(dz) \right)^2 ds \right]$$

$< \infty$

$$\mathbb{E} \left[ \int_0^T \int_{\mathbb{R}_0^m} |\lambda(s, X_s, z) \log \lambda(s, X_s, z)| d\nu(dz) \right] \quad (4.10)$$

$< \infty$ .

When we looked at the Girsanov theorem in Chapter 2, we mentioned applying a change of measure in order to solve our non-linear filtering problem. We will now look at a new representation for the observation process  $(Y_t)_{0 \leq t \leq T}$  under  $\mathbb{Q}$ .

Define the probability measure  $\mathbb{Q}$  with Radon-Nikodým (see Theorem 2.1) derivative on  $(\Omega, \mathcal{F}_t)$  given by

$$\frac{d\mathbb{Q}}{d\mathbb{P}} \Big|_{\mathcal{F}_t} = \Lambda_t \quad (4.11)$$

and require that

$$\int_{\mathbb{R}_0^d} \|z\| \nu(dz) < \infty. \quad (4.12)$$

Then by Girsanov's theorem and the uniqueness of semimartingale characteristics (see Bølviken, Duedahl, Proske [6]), the observation process  $(Y_t)_{0 \leq t \leq T}$  becomes a Lévy process being independent of the signal process under the new probability measure  $\mathbb{Q}$ . That is, the systems (4.1) and (4.2) has the following representation under  $\mathbb{Q}$ :

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t^X \quad (4.13)$$

$$dY_t = dB_t + dL_t, \quad (4.14)$$



where  $(Y_t)_{0 \leq t \leq T}$  is a Lévy process independent of  $(X_t)_{0 \leq t \leq T}$  and where

$$B_t := B_t^Y - \int_0^t (-h(s, X_s)) ds, \quad 0 \leq t \leq T \quad (4.15)$$

is the Gaussian part and

$$L_t = \int_0^t \int_{\mathbb{R}_0^m} z N(ds, dz) \quad (4.16)$$

is the jump component with respect to the Poisson random measure

$$N(ds, dz) := N_\lambda(ds, dz), \quad 0 \leq s \leq t \leq T,$$

with compensator  $ds\nu(dz)$ . Since  $(Y_t)_{0 \leq t \leq T}$  is a Lévy process under  $\mathbb{Q}$ , we also observe that the filtration  $\mathcal{F}_t^Y, 0 \leq t \leq T$  is right-continuous.

The following theorem is borrowed from Bølviken, Duedahl, Proske [6].

**Theorem 4.1.** Representation of the optimal filter.

*The optimal filter  $\pi_t$  has the representation*

$$\langle \pi_t, f \rangle = \frac{\langle \Psi_t, f \rangle}{\langle \Psi_t, 1 \rangle} \quad (4.17)$$

with

$$\langle \Psi_t, f \rangle := \mathbb{E}[Z_t f(X_t) | \mathcal{F}_t^Y] \quad (4.18)$$

for all  $f \in C_b(\mathbb{R}^n)$ , where  $\mathbb{E}_\pi$  denotes the expectation with respect to  $\pi$  and where

$$\begin{aligned} Z_t &:= \Lambda_t^{-1} \\ &= \exp \left\{ \sum_{i=1}^m \int_0^t h_i(s, X_s) dB_s^i - \frac{1}{2} \int_0^t \|h(s, X_s)\|^2 ds \right. \\ &\quad + \int_0^t \int_{\mathbb{R}_0^m} \log \lambda(s, X_s, z) N(ds, dz) \\ &\quad \left. + \int_0^t \int_{\mathbb{R}_0^m} (1 - \lambda(s, X_s, z)) ds\nu(dz) \right\}, \quad 0 \leq t \leq T, \end{aligned} \quad (4.19)$$

under  $\mathbb{Q}$ .

**Proposition 4.1.** Assume that the functions  $b, \sigma, h$  and  $\lambda$  are bounded and satisfy the conditions of linear growth and Lipschitz ((4.4) and (4.5)). Additionally, require that (4.8), (4.9) and (4.10) holds. Let  $X_t^i, 0 \leq t \leq T, i \geq 1$  be a sequence of i.i.d. copies of the solution  $X_t, 0 \leq t \leq T$  to (4.1) on our probability space, being independent of  $Y_t, 0 \leq t \leq T$ , and denote by

$Z_i, 0 \leq t \leq T$  the stochastic exponential in (4.19) based on  $X_t^i, 0 \leq t \leq T$  for all  $i \geq 1$ . Let  $f \in C_b(\mathbb{R}^n)$ . Then

$$\phi^l(f) := \frac{1}{l} \sum_{j=1}^l Z_t^j f(X_t^j) \xrightarrow{l \rightarrow \infty} \langle \Psi_t, f \rangle = \mathbb{E}_{\mathbb{Q}}[Z_t f(X_t) | \mathcal{F}_t^Y] \quad \text{a.e.} \quad (4.20)$$

for all  $t$ . Moreover, for all  $t$  there exists a constant  $C < \infty$  such that

$$\begin{aligned} & \mathbb{E}_{\mathbb{Q}} \left[ \left( \phi^l(f) - \langle \Psi_t, f \rangle \right)^2 \right] \\ &= \mathbb{E}_{\mathbb{Q}} \left[ \left( \phi^l(f) - \mathbb{E}_{\mathbb{Q}}[Z_t f(X_t) | \mathcal{F}_t^Y] \right)^2 \right] \\ &\leq \frac{1}{l} C \|f\|_{\infty}^2 \end{aligned} \quad (4.21)$$

for all  $l \geq 1$ .

*Proof.* For proof, see Proposition 13 in Bølviķen, Duedahl, Proske [6].  $\square$

# Chapter 5

## Non-life insurance

This chapter will introduce the basic concepts of non-life insurance. We will study the mathematics behind non-life insurance, and provide an intuition into why estimating the claim amount is an important part of managing risk for the insurance companies. When modelling claim amounts, it is pivotal to obtain the best estimate possible, as the insurance companies may actually go bankrupt if the claim amount is highly underestimated or overestimated. Before creating our own model (in the next chapter), we will examine the more classical ways of modelling claim frequency and claim amount. We begin this chapter by introducing basic concepts of risk in non-life insurance, followed by the introduction of Monte Carlo methods. The references used in this chapter are Agresti [1], Bølviken [5], Berk, Devore [8] and Mikosch [10].

### 5.1 Basic concepts

**Definition 5.1.** General insurance.

*General insurance is economic responsibility for incidents such as fires or accidents passed on to an insurer against a fee.*

One can intuitively think of insurance as transferring the risk from individual(s) or businesses to an insurer (the insurance company).

**Definition 5.2.** Claim.

*A claim is a formal request to an insurance company for coverage or compensation for a covered loss or policy event. The insurance company validates the claim and, once approved, issues payment to the insured or an approved interested party on behalf of the insured.*

Going forward, a claim will be denoted by  $X$ . The total claim  $X$  is a central quantity amassed during a certain period of time (typically a year). As the risk of an insurance claim is usually low, the claim amount  $X$  will often be zero. However, though rare,  $X$  can be very large.

**Definition 5.3.** Policy.

*The contract, known as a policy, releases claims when such events occur.*

The insurance company has a portfolio of risks and only a few of the risks materialize. This raises the issue of controlling the total uncertainty, which is a large subject in general insurance.

#### 5.1.1 Pricing insurance risk.

Naturally, the insured pays an agreed amount (normally monthly or yearly) to the insurer. This is because the insured is transferring risk to the insurer, which is not free.

**Definition 5.4.** Premium.

The pure premium is charged in advance and is given by

$$\pi^{\text{pu}} = \mathbb{E}(X). \quad (5.1)$$

The pure premium is a break-even situation. In practice, the insurance company will add a loading  $\gamma$  on top of  $\pi^{\text{pu}}$ , in order to make money. The premium charged by the insurer is then

$$\pi = (1 + \gamma)\pi^{\text{pu}}. \quad (5.2)$$

**5.1.2 Solvency**

A vital part in insurance is *control*. Insurance companies are obliged to set aside funds to cover future obligations.

**Definition 5.5.** Total claim amount.

Suppose a portfolio consists of  $J$  policies with claims  $X_1, \dots, X_J$ . The total claim amount is then

$$\mathcal{X} = X_1 + \dots + X_J = \sum_{i=1}^J X_i. \quad (5.3)$$

Another way of expressing the dynamics of aggregated claims over time is given by the random partial sum process

$$S(t) = \sum_{i=1}^{N(t)} X_i, \quad (5.4)$$

where  $N(t), t \geq 0$  is the claim number process and  $X_i$  are the claim sizes arriving by time  $T_i, i \geq 1$ . Generally, the claim numbers are modelled by a Poisson process, which is assumed to be independent of the i.i.d. claim sizes.

Regulators demand sufficient funds to cover  $\mathcal{X}$  with high probability. The mathematical formulation is in terms of the percentile  $q_\epsilon$ , which is the solution of the equation

$$\Pr(\mathcal{X} > q_\epsilon) = \epsilon, \quad (5.5)$$

where  $\epsilon$  is a small number. The amount  $q_\epsilon$  is known as the *solvency capital* or *reserve*.

**5.2 Monte Carlo simulation**

The risk variables  $X$  presented in 5.1 may have many random sources, and it may often be hard to find their density function  $f(x)$  or distribution function  $F(x)$ . Monte Carlo simulations allow distributions to be approximated. The techniques used approximates the solution to some problem by associating it with a probability distribution, and then sampling repeatedly from that distribution.

**Definition 5.6.** Mean and standard deviation in Monte Carlo simulations. Let  $\xi = \mathbb{E}(X)$  be the expectation and  $\sigma = \text{sd}(X)$  be the standard deviation of  $X$ . Their Monte Carlo estimates are

$$\bar{X}^* = \frac{1}{m}(X_1^* + \cdots + X_m^*) \quad \text{and} \quad s^* = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (X_i^* - \bar{X}^*)^2} \quad (5.6)$$

with the statistical properties for the mean being well known:

$$\mathbb{E}(\bar{X}^* - \xi) = 0 \quad \text{and} \quad \text{sd}(\bar{X}^*) = \frac{\sigma}{\sqrt{m}}. \quad (5.7)$$

### 5.2.1 Percentiles

The percentile  $q_\epsilon$  is the solution of either of the equations

$$F(q_\epsilon) = 1 - \epsilon \quad \text{or} \quad F(q_\epsilon) = \epsilon,$$

where  $1 - \epsilon$  represents the upper version and  $\epsilon$  represents the lower version. Monte Carlo approximations  $q_\epsilon^*$  are obtained by sorting the simulations, for example in descending order as  $X_{(1)}^* \geq \cdots \geq X_{(m)}^*$ . Then

$$q_\epsilon^* = X_{(\epsilon m)}^* \quad (5.8)$$

is the upper value and

$$q_\epsilon^* = X_{((1-\epsilon)m)}^* \quad (5.9)$$

is the lower value.

### 5.2.2 Sampling random variables

Below are some examples of algorithms that utilizes the Monte Carlo method.

Let  $F(x)$  be a strictly increasing distribution function with inverse  $x = F^{-1}(u)$  and let

$$X = F^{-1}(U) \quad \text{or} \quad X = F^{-1}(1 - U), \quad U \sim \text{uniform}. \quad (5.10)$$

We can now construct an algorithm for sampling random variables using the inversion method.

---

**Algorithm 2** Sampling by inversion

---

Input: The percentile function  $F^{-1}(u)$ .

1. **draw**  $U^* \sim \text{uniform}$
  2. **return**  $X^* \leftarrow F^{-1}(U^*)$  or  $X^* \leftarrow F^{-1}(1 - U^*)$
- 

Let's further demonstrate this method with an example.

**Example 5.1.** Using the inversion method on the Burr distribution.

Let

$$\mathcal{X} = \sum_{i=1}^{\mathcal{N}} Z_i$$

be the total pay-out in a general insurance portfolio under standard assumptions, where  $\mathcal{N} \sim \text{Poisson}(\lambda)$  is independent of the individual losses  $Z_i$ , which are i.i.d. with the Burr distribution. That is,  $Z_i$  has the distribution function

$$F(z) = 1 - \frac{1}{\left(1 + \left(\frac{z}{\beta}\right)^\theta\right)^\alpha}, \quad z > 0, \quad (5.11)$$

where  $\alpha, \theta, \beta \in \mathbb{R}^+$ . We can now sample the random variables  $Z_1, \dots, Z_m$  using the inversion method. Before we get to the algorithm, we need to find  $F^{-1}(u)$  by basic algebra:

$$F(z) = 1 - \frac{1}{\left(1 + \left(\frac{z}{\beta}\right)^\theta\right)^\alpha} = u \quad (5.12)$$

$$\Rightarrow z = \beta \left( (1 - u)^{-1/\alpha} - 1 \right)^{1/\theta} = F^{-1}(u). \quad (5.13)$$

When simulating  $\mathcal{X}$ , we obtain the following algorithm:

1. Input:  $\alpha, \theta, \beta, \lambda$
2. **for**  $i = 1, \dots, m$  **do**
3.     Draw  $\mathcal{N}^* \sim \text{Poisson}(\lambda)$
4.      $\mathcal{X}_i^* \leftarrow 0$
5.     **for**  $j = 1, \dots, \mathcal{N}^*$  **do**
6.         Draw  $Z^* = \beta \left( (1 - u)^{-1/\alpha} - 1 \right)^{1/\theta}$
7.          $\mathcal{X}_i^* \leftarrow \mathcal{X}_i^* + Z^*$
8.     **end for**
9. **end for**
10. **return**  $\mathcal{X}_1^*, \dots, \mathcal{X}_m^*$

The mean and sd of  $\mathcal{X}$  can be found using the equations from (5.6). ♣

Below is an algorithm for modelling random variables with stochastic volatility.

---

**Algorithm 3** Gaussian with stochastic volatility

---

Input:  $\xi, \xi_\sigma$ , model for  $Z$

1. Draw  $Z^*$  and  $\sigma^* \leftarrow \xi_\sigma \sqrt{Z^*}$
  2. Generate  $U^* \sim \text{uniform}$ .
  3. Return  $X^* \leftarrow \xi + \sigma^* \Phi^{-1}(U^*)$
-

## 5.3 Modelling claim frequency

This section will cover the basic concepts of claim frequency in actuarial modelling. We will dive deeper into the idea of modelling claim frequency using the Poisson distribution. When using the Poisson distribution, the incidents happen (for the most part) rarely, and independently of each other. This is the reason why it is massively utilized in the insurance world.

### 5.3.1 Claim intensities

As mentioned above, incidents in the insurance world happen rarely and independently of each other. Let  $T$  be some period of time (this is often set to be 1 year in the insurance world). We can divide  $T$  into  $K$  pieces of equal length  $h = \frac{T}{K}$ . Naturally, if  $K$  is a large number, the probability of more than one incident in a short interval is exceptionally small. If the number of events per interval is either 0 or 1 and is denoted by  $I_k$  on the  $k$ th interval, the count for the entire period is

$$N = I_1 + \cdots + I_m = \sum_{k=1}^m I_k,$$

where  $N$  is the total number of events. Assume that events happen independently of each other and that  $p$  is equal for all  $k$ , where  $p$  is given by

$$p = P(I_k = 1).$$

This is then an ordinary Bernoulli series, and  $N$  is binomally distributed with probability density function

$$P(N = n) = \binom{K}{n} p^n (1 - p)^{K-n},$$

for  $n = 0, \dots, K$ .

Let  $p = \mu h = \mu \frac{T}{K}$ . By inserting this into the binomial distribution and letting  $K \rightarrow \infty$ , the binomial distribution converges to the Poisson distribution; that is,

$$P(N = n) \rightarrow \frac{\mu^n T^n}{n!} e^{-\mu T} \quad \text{as } K \rightarrow \infty.$$

In other words,  $N$  is Poisson distributed with parameter  $\lambda = \mu T$  when  $K \rightarrow \infty$ .

From this, it follows that the portfolio number of claims  $\mathcal{N}$  is Poisson distributed with parameter

$$\lambda = (\mu_1 + \cdots + \mu_J)T = J\bar{\mu}T, \quad \text{where } \bar{\mu} = \frac{(\mu_1 + \cdots + \mu_J)}{J}.$$

For further detail on this, consult Bølviken [5].

Using the fact that  $N$  is the claim number for policies and  $\mathcal{N}$  is the claim number for portfolios, we have that both  $N$  and  $\mathcal{N}$  are Poisson distributed with parameters  $\lambda = \mu T$  on policy level and  $\lambda = J\mu T$  on portfolio level.

The mean and standard deviation of Poisson models are given by

$$\mathbb{E}(N) = \lambda \text{ and } \text{sd}(N) = \sqrt{\lambda},$$

respectively. A useful property is the *convolution* property:

If  $N_1, \dots, N_J$  are independent and Poisson distributed with parameters  $\lambda_1, \dots, \lambda_J$ , then

$$\mathcal{N} = N_1 + \dots + N_J \sim \text{Poisson}(\lambda_1 + \dots + \lambda_J).$$

Claim intensities are determined from historical data, and is rather straight-forward:

Let  $n_1, \dots, n_m$  be claim numbers from  $m$  policies exposed to risk during  $T_1, \dots, T_m$ . The usual estimate of a common intensity  $\mu$  is given by

$$\hat{\mu} = \frac{n_1 + \dots + n_m}{T_1 + \dots + T_m}.$$

It follows that

$$\mathbb{E}(\hat{\mu}) = \mu \quad \text{and} \quad \text{sd}(\hat{\mu}) = \sqrt{\frac{\mu}{T_1 + \dots + T_m}}.$$

**Example 5.2.**

The idea of claim intensity can be illustrated with an easy example. Assume an insurance company has a portfolio of car insurance. Let the total number of claims be 7512, and the total risk exposure  $T_1 + \dots + T_m = 143200$ . This gives the common intensity

$$\hat{\mu} = \frac{7512}{143200} = 0.0524 = 5.24\%$$

and standard deviation

$$\text{sd}(\hat{\mu}) = \sqrt{\frac{0.0524}{143200}} = 0.0006 = 0.06\%.$$



**5.3.2 Random claim intensities**

Although the claim intensities introduced in the previous section are mathematically correct, one will often encounter *random claim intensities* in practice. In other words,  $\mu$  will usually vary over the portfolio. If  $\mu$  varies over the portfolio, we cannot use a common  $\mu$  as shown in Section 5.3.1.



**Example 5.3.** Random claim intensities.

Assume we are interested in estimating the claim intensity for automobiles. The drivers of the automobiles will, naturally, not have the same personal factors. Their ability to drive and caution may vary, and the insurance company must take this into account. This is done by making  $\mu$  a random variable, and draw  $\mu$  for each driver. ♣

At policy level, we have the following:

$$\mathbb{E}(N|\mu) = \text{var}(N|\mu) = \mu T, \quad (5.14)$$

which follows from the properties of the Poisson distribution. Let  $\mathbb{E}(\mu) = \xi$  and  $\text{sd}(\mu) = \sigma$ . Using the double rules from Section 6.3.2. in Bølviken [5], we know that

$$\mathbb{E}(N) = \mathbb{E}(\mu T) = \xi T \quad \text{and} \quad \text{var}(N) = \mathbb{E}(\mu T) + \text{var}(\mu T) = \xi T + \sigma^2 T^2.$$

Due to the fact that  $\mathbb{E}(N) < \text{var}(N)$ , it follows that  $N$  is no longer Poisson distributed. This leads to the following definition.

**Definition 5.7.** Overdispersion.

*In the Poisson distribution, the mean is equal to the variance. In practice, count observations often exhibit variability exceeding that predicted by the Poisson. This phenomenon is called overdispersion.*

How does one deal with the problem of overdispersion? A solution is presented in Section 5.3.3.

**5.3.3 The mixed Poisson process.**

When modelling random intensities, a good way to account for overdispersion is by a mixture model. The most common mixture model used in non-life insurance is called *the negative binomial model*. It assumes that  $\mu$  is given by  $\mu = \xi G$ , where  $G \sim \text{Gamma}(\alpha)$ .

**Definition 5.8.** Mixed Poisson process

*Let  $\tilde{N}$  be a standard homogeneous Poisson process and  $\mu$  be the mean value function of a Poisson process on  $[0, \infty)$ . Let  $\theta > 0$  a.s. be a random variable independent of  $\tilde{N}$ . Then the process*

$$N(t) = \tilde{N}(\theta\mu(t)), \quad t \geq 0, \quad (5.15)$$

*is said to be a mixed Poisson process with mixing variable  $\theta$ .*

A mixed Poisson process is a special case of a Cox process (see Definition 2.11) where the mean value function  $\mu$  is a general random process with non-decreasing sample paths, independent of the underlying homogeneous Poisson process  $\tilde{N}$ . The following theorem borrows information from Section 7.3.2. in Agresti [1].

**Theorem 5.1.** The negative binomial model.

Assume that given  $\lambda$ , the random variable  $y$  has a  $\text{Poisson}(\lambda)$  distribution, and  $\lambda$  has the Gamma distribution. Using the properties of the Gamma distribution, we know that  $\mathbb{E}(\lambda) = \mu$  and  $\text{var}(\lambda) = \mu^2/k$  for a shape parameter  $k > 0$ . Marginally, the gamma mixture of the Poisson distribution gives the negative binomial distribution for  $y$ . The negative binomial distribution has the following properties:

(i) Its probability mass function is given by

$$p(y; \mu, k) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{\mu}{\mu+k}\right)^y \left(\frac{k}{\mu+k}\right)^k, \quad y = 0, 1, 2, \dots \quad (5.16)$$

(ii) It has a natural parameter  $\log[\mu/(\mu+k)]$ .

(iii) If  $\gamma = 1/k$ , then  $\mathbb{E}(y) = \mu$  and  $\text{var}(y) = \mu + \gamma\mu^2$ .

*Proof.*

(i) Using the probability density function of the Gamma distribution and the probability mass function of the Poisson distribution, we have the following:

For  $y = 0, 1, 2, \dots$  the marginal probability mass function of  $Y$  is given by

$$\begin{aligned} p(y; \mu, k) &= \int_0^\infty P(Y = y|\lambda) f(\lambda; k, \mu) d\lambda \\ &= \int_0^\infty \frac{\lambda^y}{y!} e^{-\lambda} \frac{(k/\mu)^k}{\Gamma(k)\lambda^{k-1}} e^{-k\lambda/\mu} d\lambda \\ &= \frac{(k/\mu)^k}{\Gamma(k)y!} \int_0^\infty \lambda^{y+k-1} e^{-(\mu+k)\lambda/\mu} d\lambda \\ &= \frac{(k/\mu)^k}{\Gamma(k)\Gamma(y+1)} \int_0^\infty \left(\frac{\mu}{\mu+k}\right)^{y+k-1} e^{-u} \frac{\mu}{\mu+k} du \quad (\text{substitute } u = (\mu+k)\lambda/\mu) \\ &= \frac{(k/\mu)^k}{\Gamma(k)\Gamma(y+1)} \left(\frac{\mu}{\mu+k}\right)^{y+k} \int_0^\infty u^{y+k-1} e^{-u} du \\ &= \frac{(k/\mu)^k}{\Gamma(k)\Gamma(y+1)} \left(\frac{\mu}{\mu+k}\right)^{y+k} \Gamma(y+k) \\ &= \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{\mu}{\mu+k}\right)^y \left(\frac{k}{\mu+k}\right)^k. \end{aligned}$$

(ii) Recall from Section 2.6 that the natural parameter is given by  $\theta_i$  in the exponential dispersion family. Introduce the random variable  $Y^* = Y/k$ .

$Y^*$  has the probability mass function given by

$$p^*(y^*; \mu, k) = \frac{\Gamma(ky^* + k)}{\Gamma(k)\Gamma(ky^* + 1)} \left(\frac{\mu}{\mu+k}\right)^{ky^*} \left(\frac{k}{\mu+k}\right)^k; \quad y^* = 0, \frac{1}{k}, \frac{2}{k}, \dots \quad (5.17)$$

In the negative binomial model, we can rewrite its probability mass function

as

$$\begin{aligned} p^*(y^*; \mu, k) &= \exp \left\{ (ky^*) \log \left( \frac{\mu}{\mu+k} \right) + k \log \left( \frac{k}{\mu+k} \right) + \log \left( \frac{\Gamma(ky^* + k)}{\Gamma(k)\Gamma(ky^* + 1)} \right) \right\} \\ &= \exp \left\{ \left[ y^* \log \left( \frac{\mu}{\mu+k} \right) + \log \left( \frac{k}{\mu+k} \right) \right] / \frac{1}{k} + \log \left( \frac{\Gamma(ky^* + k)}{\Gamma(k)\Gamma(ky^* + 1)} \right) \right\}. \end{aligned}$$

This is of the form (5.16), with

$$\begin{aligned} \theta &= \log \left( \frac{\mu}{\mu+k} \right), \\ b(\theta) &= -\log \left( 1 - \frac{\mu}{\mu+k} \right) = -\log(1 - e^\theta), \\ a(\phi) &= \frac{1}{k}. \end{aligned}$$

Hence the natural parameter is given by  $\theta = \log \left( \frac{\mu}{\mu+k} \right)$ .

(iii) Using (2.39) and (2.40), we have that

$$\mathbb{E}(Y^*) = b'(\theta) = \frac{d}{d\theta} [-\log(1 - e^\theta)] = \frac{e^\theta}{1 - e^\theta} = \frac{\frac{\mu}{\mu+k}}{1 - \frac{\mu}{\mu+k}} = \frac{\mu}{k}$$

and

$$\begin{aligned} \text{var}(Y^*) &= a(\phi)b''(\theta) \\ &= \frac{1}{k} \frac{d^2}{d\theta^2} [-\log(1 - e^\theta)] \\ &= \frac{1}{k} \frac{e^\theta}{(1 - e^\theta)^2} \\ &= \frac{1}{k} \frac{\frac{\mu}{\mu+k}}{\left(1 - \frac{\mu}{\mu+k}\right)^2} \\ &= \frac{1}{k} \frac{\frac{\mu}{\mu+k}}{\left(\frac{k}{\mu+k}\right)^2} \\ &= \frac{1}{k^3} \mu(\mu+k). \end{aligned}$$

Since  $Y = Y^*k$ , we have

$$\mathbb{E}(Y) = k\mathbb{E}(Y^*) = k \frac{\mu}{k} = \mu,$$

and

$$\text{var}(Y) = k^2 \text{var}(Y^*) = \frac{1}{k} \mu(\mu+k) = \mu + \frac{\mu^2}{k}.$$

By letting  $\gamma = 1/k$ , we have

$$\mathbb{E}(Y) = \mu \quad \text{and} \quad \text{var}(Y) = \mu + \gamma\mu^2.$$

□

The negative binomial model is an example of a mixed Poisson process.

### 5.3.4 Poisson regression

Insurance companies will often charge different fees to their customers, depending on which group their customers belong to. For instance, a driver who is between 18-21 years old is more likely to be a subject of an incident such as a car crash than a driver who is between 30-33 years old, who is normally more cautious in traffic.

This brings us to the topic of Poisson regression. As explained in Section 5.3.1, claim numbers are usually Poisson distributed. When constructing a model for estimating claim intensity, insurance companies will make sure to charge a higher fee for customers that are associated with higher risk. This is done in the form of Poisson regression. The idea is to attribute variations in  $\mu$  to variations in a set of observable variables  $x_1, \dots, x_j$ .

From Section 2.6.1 we know that the exponential dispersion family is on the form

$$f(y_i; \theta_i, \phi) = \exp \left\{ [y_i \theta_i - b(\theta_i)] / a(\phi) + c(y_i, \phi) \right\}. \quad (5.18)$$

It turns out the Poisson distribution is a member of the exponential dispersion family. We have

$$\begin{aligned} f(y_i; \mu_i) &= \frac{e^{-\mu_i} \mu_i^{y_i}}{y_i!} \\ &= \exp \left[ y_i \log \mu_i - \mu_i - \log(y_i!) \right] \\ &= \exp \left[ y_i \theta_i - \exp(\theta_i) - \log(y_i!) \right], \quad y_i = 0, 1, 2, \dots \end{aligned} \quad (5.19)$$

where the natural parameter  $\theta_i = \log \mu_i$ . This has exponential dispersion form (5.18) with  $b(\theta_i) = \exp(\theta_i)$ ,  $a(\phi) = 1$ , and  $c(y_i, \phi) = -\log(y_i!)$ . By (2.39) and (2.40),

$$\begin{aligned} \mathbb{E}(y_i) &= b'(\theta_i) = \exp(\theta_i) = \mu_i, \\ \text{var}(y_i) &= b''(\theta_i) = \exp(\theta_i) = \mu_i. \end{aligned}$$

Using the canonical link function, we end up with the final Poisson regression:

$$\theta_i = \log(\mu_i) = \sum_{j=1}^p \beta_j x_{ij} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ij} \quad (5.20)$$

which gives the intensity

$$\mu_i = \exp \left( \sum_{j=1}^p \beta_j x_{ij} \right). \quad (5.21)$$

#### Example 5.4.

Following the theme of automobiles, consider a case where we wish to estimate the intensity for a model with two explanatory variables. Assume the

intensity varies depending on the *age* and the *gender* of the customer. We have


$$\log(\mu_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}$$

where  $x_{i1}$  is the age of the owner of car  $i$  and

$$x_{i2} = \begin{cases} 0, & \text{if } j \text{ is male} \\ 1, & \text{if } j \text{ is female.} \end{cases}$$

The intensity  $\mu_i$  is then given by

$$\mu_i = e^{\beta_0} e^{\beta_1 x_{i1}} e^{\beta_2 x_{i2}}. \quad (5.22)$$

Here  $e^{\beta_0}$  is a baseline intensity, which is driven up or down by the explanatory variables. 

When developing our final model in Chapter 6, these explanatory variables are automatically included in our data. As a result of this, we do not have account for the problem of overdispersion when simulating our new model. Our model is also not limited to the negative binomial model, it can take any mixture model into account. Intuitively, one can imagine that the model that creates the explanatory variables is "baked" into our model.

## 5.4 Modelling claim size

Differently from estimating claim frequency, claim size estimation does not have a specific distribution to model from. Claim frequency relies on the Poisson distribution, whilst claim size is modelled by imposing a family of probability distributions and estimate their parameters from historical data. In this chapter we will introduce distributions that can be utilized in order to provide a good basis for estimating the size of the claims. We will focus on three main distribution families: Log-normal, Gamma and Pareto. These are the distributions that are most commonly used when modelling claim sizes.

### 5.4.1 Introduction

Our goal is to estimate the total claim amount

$$S(t) = \sum_{i=1}^{N(t)} X_i, \quad t \geq 0,$$

where the claim number process  $N(t), t \geq 0$  is independent of the i.i.d. claim size sequence  $(X_i)$ . This is of great importance to insurance companies in order to determine a premium which covers the losses represented by  $S(t)$ .

Having introduced claim frequency in Section 5.3, we will now introduce claim sizes. That is, we will consider realistic claim size distributions and their properties. We will discuss the notions of heavy- and light-tailed claim size distributions. By doing this, we will acquire solid methods for modelling large and small claims.

### 5.4.2 Parametric modelling

There are two approaches to modelling claim size. The first approach is known as *parametric* modelling, which is done through families of distribution such as the Gamma, log-normal or Pareto with parameters tuned to historical data. The second approach is *non-parametric* where each claim  $z_i$  of the past is assigned a probability  $1/n$  of reappearing in the future.

When modelling parametric claim size, distributions should always include  $\beta$  which is known as a parameter of scale. It is given by

$$Z = \beta Z_0, \quad (5.23)$$

where  $\beta \in \mathbb{R}^+$  and  $Z_0$  is a standardized random variable corresponding to  $\beta = 1$ .

As an example, we may look at the log-normal model. In the log-normal model, we have the random variable  $Z$  given by

$$Z = e^{\theta + \sigma \epsilon},$$

where  $\theta$  and  $\sigma$  are parameters and  $\epsilon \sim N(0, 1)$ . Using the parameter of scale,  $Z$  may be rephrased as

$$Z = \xi Z_0,$$

where  $Z_0 = e^{-\sigma^2/2 + \sigma \epsilon}$  and  $\xi = e^{\theta + \sigma^2/2}$ . Here  $\mathbb{E}(Z_0) = 1$ , and expectation and scale parameter is represented as  $\xi$ .

When finding a model that includes a scale parameter, we have by (5.23) that

$$P(Z \leq z) = P(Z_0 \leq z/\beta) \quad \text{or} \quad F(z|\beta) = F_0(z/\beta), \quad (5.24)$$

where  $F(z|\beta)$  and  $F_0(z)$  are the distribution functions of  $Z$  and  $Z_0$ . Denote  $f_0(z)$  as the derivative of  $F_0(z)$ . That is,  $f_0(z) = \frac{d}{dz}F_0(z)$ . Differentiating  $F(z|\beta)$  with respect to  $z$  yields

$$f(z|\beta) = \frac{1}{\beta} f_0(z/\beta), \quad z > 0,$$

where  $f(z|\beta)$  is the family of density functions.

The standard method of fitting a model that includes a scale parameter is through likelihood estimation. Assuming  $z_1, \dots, z_n$  are historical claims, the method is as follows:

1) Find the logarithm of the likelihood. The loglikelihood is given by

$$\mathcal{L}(\beta, f_0) = -n \log(\beta) + \sum_{i=1}^n \log(f_0(z_i/\beta)). \quad (5.25)$$

2) Maximize  $\mathcal{L}(\beta, f_0)$  with respect to  $\beta$  and other parameters. Let  $\mathbf{a}$  be a vector containing  $\beta$  and other parameters. The maximum is found by letting

$$\frac{\partial}{\partial \mathbf{a}} \mathcal{L}(\beta, f_0) = 0. \quad (5.26)$$

### 5.4.3 Non-parametric modelling

In non-parametric modelling each claim  $z_i$  of the past is assigned a probability  $1/n$  of reappearing in the future. These claims are connected to the random variable  $\hat{Z}$ . Expectation, standard deviation and percentiles are all closely related to the ordinary sample versions. Replace  $X_i^*$  with  $Z_i$  in (5.6) and we get the same results.

Upper percentiles are approximately the historical claims in descending order; that is,  $\hat{q}_\epsilon = z_{(en)}$ , where  $z_1 \geq \dots \geq z_n$ .

### 5.4.4 The log-normal family

The log-normal family of distributions is a common model for insurance losses. In this section we will introduce the most important properties of the log-normal family.

Similarly as in Section 5.4.2, let  $Z = \xi Z_0$ , where  $Z_0 = e^{-\sigma^2/2 + \sigma\epsilon}$ , where  $\epsilon \sim N(0, 1)$ . The mean, standard deviation and skewness are

$$\mathbb{E}(Z) = \xi, \quad \text{sd}(Z) = \xi\sqrt{e^{\sigma^2} - 1} \quad \text{and} \quad \text{skew}(Z) = (e^{\sigma^2} + 2)\sqrt{e^{\sigma^2} - 1}.$$

When estimating the parameters, notice that the log-normal family becomes normal when taking the logarithm, hence the name log-normal. Let  $Y = \log(Z)$ . We have

$$Y = \log(Z) = \log(\xi) - \frac{1}{2}\sigma^2 + \sigma\epsilon, \quad (5.27)$$

where

$$\mathbb{E}(Y) = \log(\xi) - \frac{1}{2}\sigma^2$$

is the mean and

$$\text{sd}(Y) = \sigma$$

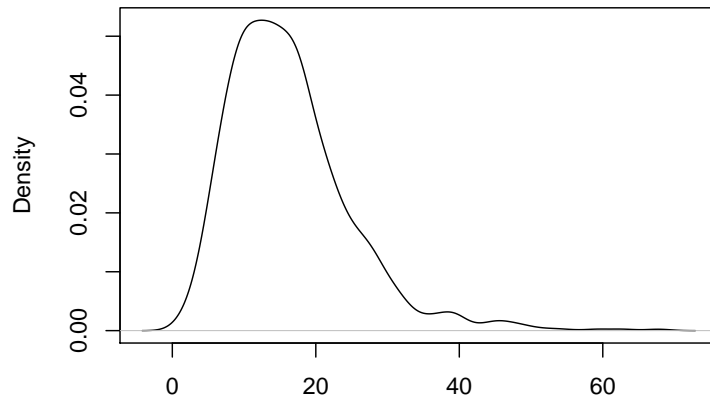
is the standard deviation.

If one tail is longer than the other, it means we have a skewed distribution. For a better understanding, let's look at an example.

**Example 5.5.** Skewness of the log-normal distribution.

If the distribution has a long left tail, it means that the distribution is left-skewed. Below is a graph showing the log-normal distribution with a long left tail.

### Density of log-normal claims



A right-skewed distribution is the opposite of a left-skewed distribution.

#### 5.4.5 The Gamma family

The second common model for insurance losses is the Gamma model. Define  $Z$  as in Section 5.3.3; that is,  $Z = \xi G$ , where  $G$  is Gamma distributed with parameter (shape)  $\alpha$ . The mean, standard deviation and skewness are

$$\mathbb{E}(Z) = \xi, \quad \text{sd}(Z) = \frac{\xi}{\sqrt{\alpha}} \quad \text{and} \quad \text{skew}(Z) = \frac{2}{\sqrt{\alpha}}. \quad (5.28)$$

The proposition below is a property borrowed from Chapter 9 in Bølviken [5].

**Proposition 5.1.** The convolution property of the Gamma distribution.

Assume  $G_1, \dots, G_n$  are independent with  $G_i \sim \text{Gamma}(\alpha_i)$  for  $i = 1, \dots, n$ . Then

$$\overline{G} \sim \text{Gamma}(\alpha_1 + \dots + \alpha_n) \quad \text{if} \quad \overline{G} = \frac{\alpha_1 G_1 + \dots + \alpha_n G_n}{\alpha_1 + \dots + \alpha_n}. \quad (5.29)$$

Obviously, when modelling claim size using the Gamma distribution, we are taking the parametric approach. The parameters  $\xi$  and  $\alpha$  in the Gamma model can be determined from historical data  $z_1, \dots, z_n$ . The sample mean  $\bar{z}$  and standard deviation  $s$  are then matched to the theoretical expressions, yielding  $\bar{z} = \hat{\xi}$ ,  $s = \hat{\xi}/\sqrt{\hat{\alpha}}$  with solution

$$\hat{\xi} = \bar{z} \quad \text{and} \quad \hat{\alpha} = \frac{\bar{z}^2}{s^2}.$$

When fitting the Gamma model, we follow the two steps given in Section 5.4.2:



1) The logarithm of the density function of the standard Gamma is

$$\log(f_0(z)) = \alpha \log(\alpha) - \log(\Gamma(\alpha)) + (\alpha - 1) \log(z) - \alpha z. \quad (5.30)$$

By (5.25), we have

$$\mathcal{L}(\xi, \alpha) = n\alpha \log\left(\frac{\alpha}{\xi}\right) - n \log \Gamma(\alpha) + (\alpha - 1) \sum_{i=1}^n \log(z_i) - \frac{\alpha}{\xi} \sum_{i=1}^n z_i.$$

2) When maximizing  $\mathcal{L}(\xi, \alpha)$ , we have

$$\frac{\partial \mathcal{L}}{\partial \xi} = -\frac{n\alpha}{\xi} + \frac{\alpha}{\xi^2} \sum_{i=1}^n z_i.$$

By simple algebra, we have that

$$\frac{\partial \mathcal{L}}{\partial \xi} = 0 \quad \text{when} \quad \xi = \frac{1}{n} \sum_{i=1}^n z_i = \bar{z}.$$

Thus  $\hat{\xi} = \bar{z}$  is the likelihood estimate and  $\mathcal{L}(\bar{z}, \alpha)$  can be tracked under variation of  $\alpha$  for the maximizing value  $\hat{\alpha}$ .

#### 5.4.6 The Pareto family

The final distribution we will examine is the Pareto distribution. The Pareto distribution has parameters  $\alpha$  and  $\beta$ , where  $\alpha, \beta > 0$ , and will be denoted by  $\text{Pareto}(\alpha, \beta)$ . It is very heavy tailed, which makes it a popular choice for large claims in property insurance.

The mean, standard deviation and skewness are given by

$$\mathbb{E}(Z) = \xi = \frac{\beta}{\alpha - 1}, \quad \text{sd}(Z) = \xi \sqrt{\left(\frac{\alpha}{\alpha - 2}\right)} \quad \text{and} \quad \text{skew}(Z) = 2\sqrt{\left(\frac{\alpha - 2}{\alpha}\right)} \frac{\alpha + 1}{\alpha - 3}.$$

This only holds if:  $\alpha > 1$  in  $\mathbb{E}(Z)$ ,  $\alpha > 2$  in  $\text{sd}(Z)$  and  $\alpha > 3$  in  $\text{skew}(Z)$ .

When fitting the Pareto model, we (again) follow the steps given in Section 5.4.2:

1) The log-likelihood of the Pareto distribution is given by

$$n \log\left(\frac{\alpha}{\beta}\right) - (1 + \alpha) \sum_{i=1}^n \log\left(1 + \frac{z_i}{\beta}\right).$$

However, claims may exceed a threshold  $b$ . That is, observed claims  $z_1, \dots, z_n$  may exceed the thresholds  $b_1, \dots, b_n$ . Assume that there are  $n_b$  claims that have exceeded certain thresholds  $b_1, \dots, b_{n_b}$ . We must now take these claims

into account, along with the ordinary claims. The full log-likelihood function (as given in Section 9.4.3 in Bølviken [5]) is then given by

$$\mathcal{L}(\alpha, \beta) = n \log\left(\frac{\alpha}{\beta}\right) - (1 + \alpha) \sum_{i=1}^n \log\left(1 + \frac{z_i}{\beta}\right) - \alpha \sum_{i=1}^{n_b} \log\left(1 + \frac{b_i}{\beta}\right).$$

2) We now maximize  $\mathcal{L}(\alpha, \beta)$  by differentiating with respect to  $\alpha$ .

$$\begin{aligned} \frac{\partial \mathcal{L}(\alpha, \beta)}{\partial \alpha} &= 0 \\ \Rightarrow n \log\left(\frac{\alpha}{\beta}\right) - (1 + \alpha) \sum_{i=1}^n \log\left(1 + \frac{z_i}{\beta}\right) - \alpha \sum_{i=1}^{n_b} \log\left(1 + \frac{b_i}{\beta}\right) &= 0. \end{aligned}$$

Let  $\hat{\alpha}_\beta$  denote the maximized value of  $\mathcal{L}(\alpha, \beta)$ . By simple algebra, solving for  $\alpha$  yields

$$\hat{\alpha}_\beta = \frac{n}{\sum_{i=1}^n \log\left(1 + \frac{z_i}{\beta}\right) + \sum_{i=1}^{n_b} \log\left(1 + \frac{b_i}{\beta}\right)}.$$

# Chapter 6

## Our new stochastic model for total claim amounts

We will begin this chapter by building our stochastic model in the framework of generalized Cox processes. Continuing, we will briefly discuss the distribution families from Chapter 5 that are appropriate for modelling claim size. Our final result will be a new stochastic model which allows for the modelling of regime switching effects of data arising, for example from natural disasters, regulatory changes or other "shocks" in insurance markets. As discussed in Chapter 4, we will apply Monte Carlo and non-linear filtering techniques in order to estimate the signal process as accurately as possible, which is modelled by a stochastic differential equation given by (4.1). With the introduction of our new model, we will examine the results given from the simulations, and conclude with a discussion on the advantage and drawbacks of using this model. The references used in this chapter are Applebaum [2], Bølviken, Duedahl, Proske [6], Øksendal [11] and Xiong [12].

### 6.1 Building our model in the framework of generalized Cox processes

In this section we will first give a general introduction to our new model. When the model has been introduced, we will continue with simulations of our model with specific functions in the next section. To conclude, we will discuss the results of the simulations from our new model. As mentioned in the introduction to this chapter, our model will allow for capturing both regime switching effects and mean reversion.

In order to estimate the unknown signal process  $X_t$ , which is of the same form as in the filtering framework of Chapter 4, we will apply non-linear filtering techniques for Lévy processes. In addition to this, the observation process  $Y_t$  will be described by a generalized Cox process.

Hence, our signal process is given by  $X_t$ , where  $X_t$  follows the dynamics

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t^X, \quad 0 \leq t \leq T, \quad (6.1)$$

where  $b, \sigma$  are Borel functions and  $B_t^X$ ,  $0 \leq t \leq T$  is a Brownian motion. The signal process  $X_t$  is not directly observable, hence we must turn to the observation process  $Y_t$  to acquire information about the signal  $X$ .

We choose a generalized Cox process in order to describe the dynamics of our observation process, as done in Chapter 4. The observation process  $Y_t$  is described by the dynamics:

$$dY_t = h(t, X_t)dt + dB_t^Y + \int_{\mathbb{R}_0^m} z N_\lambda(dt, dz), \quad 0 \leq t \leq T, \quad (6.2)$$

where  $h$  is a Borel function,  $B_t^Y$  is a Brownian motion and  $N_\lambda$  is the jump measure of a generalized Cox process with a predictable compensator  $\hat{\mu}$  given by

$$\hat{\mu}(dt, dz, \omega) = \lambda(t, X_t, z) dt \nu(dz) \quad (6.3)$$

for a Borel function  $\lambda$  and a Lévy measure  $\nu$  where

$$\int_{\mathbb{R}_0^d} \|z\| \nu(dz) < \infty. \quad (6.4)$$

When simulating our new model, we need to know at which time jump occurs. We define the function

$$\mu^*(t) := \int_0^t \lambda^*(s, X_s) ds, \quad 0 \leq t \leq T. \quad (6.5)$$

The "intensity process"  $\lambda^*(t, X_t), t \geq 0$  depends on the unknown signal process  $X_t$  given in (6.1) which captures regime switching effects of data.

Having defined  $\mu^*(t)$ , we can construct  $N(t)$ , which is the number of claims occurring at time  $t$ . Let  $N(t) = \tilde{N}(\mu^*(t))$  for a process  $\mu^*(t)$  given by (6.5). The process  $\mu^*(t)$  is independent of  $\tilde{N}(t), t \geq 0$ . Recall from Chapter 4 that the observation process  $Y_t$  can be decomposed (under a change of measure) in the following way:

$$dY_t = dB_t + dL_t, \quad (6.6)$$

where  $(Y_t)_{0 \leq t \leq T}$  is a Lévy process independent of  $(X_t)_{0 \leq t \leq T}$  and where

$$B_t = B_t^Y - \int_0^t (-h(s, X_s)) ds, \quad 0 \leq t \leq T \quad (6.7)$$

is the Gaussian part and

$$L_t = \int_0^t \int_{\mathbb{R}_0^m} z N(ds, dz) \quad (6.8)$$

is the jump component with respect to the Poisson random measure

$$N(ds, dz) := N_\lambda(ds, dz), \quad 0 \leq s \leq t \leq T,$$

with compensator  $ds\nu(dz)$ . For the simulations which will be performed later, recall that

$$\int_0^t \int_{\mathbb{R}_0^m} z N(ds, dz) = \sum_{0 \leq s \leq t} \Delta Y_s \mathbb{1}_{\{\Delta Y_s \neq 0\}}, \quad 0 \leq t \leq T. \quad (6.9)$$

How exactly do we allow for regime switching in our signal process  $(X_t)_{0 \leq t \leq T}$ ? As the dynamics of the signal process  $X$  is given by

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t^X, \quad 0 \leq t \leq T,$$

as in (6.1), we are able to capture regime switching by choosing a function for  $b$  and  $\sigma$ . When the primary objective is to allow for the capture of regime switching, the most interesting coefficient to look at is  $b$ . By choosing  $b$  to be a discontinuous vector field, we allow the model to capture regime switching effects arising from for example financial disasters of long-lasting nature.

If we were to choose  $b$  as a discontinuous vector field, it could for example be of the form

$$b(t, x) = \begin{cases} a_1 & \text{if } \|x\| \geq \rho, \\ a_2 & \text{else.} \end{cases}$$

Here  $a_1, a_2$  are vectors in  $\mathbb{R}^n$ , and they represent the different regime switching states the process  $X_t$  will assume, depending whether it exceeds a certain threshold  $\rho$  at time  $t$  or not. We could for example, in a modified mean-reverting Vasicek model, use the function

$$b(x) = \begin{cases} a(b_1 - x) & \text{if } x \geq \rho, \\ a(b_2 - x) & \text{else,} \end{cases}$$

for  $a, b_1, b_2 \geq 0$ , when  $\sigma(x) \equiv \sigma > 0$ . Here  $a$  is the mean reversion coefficient, and  $b_1, b_2$  are the stabilization levels in the long run. That is, given a certain amount of time, the process  $X_t$  will stabilize around either  $b_1$  or  $b_2$ . As an example, consider the case where  $x \geq \rho$ . Here, the drift of  $X_t$  will be negative at time  $t \leq T$  if  $x > b_1$ , and positive at time  $t \leq T$  if  $x < b_1$ .

The classical Vasicek model is obtained from the latter one, if  $b_1 = b_2$ ; that is, if

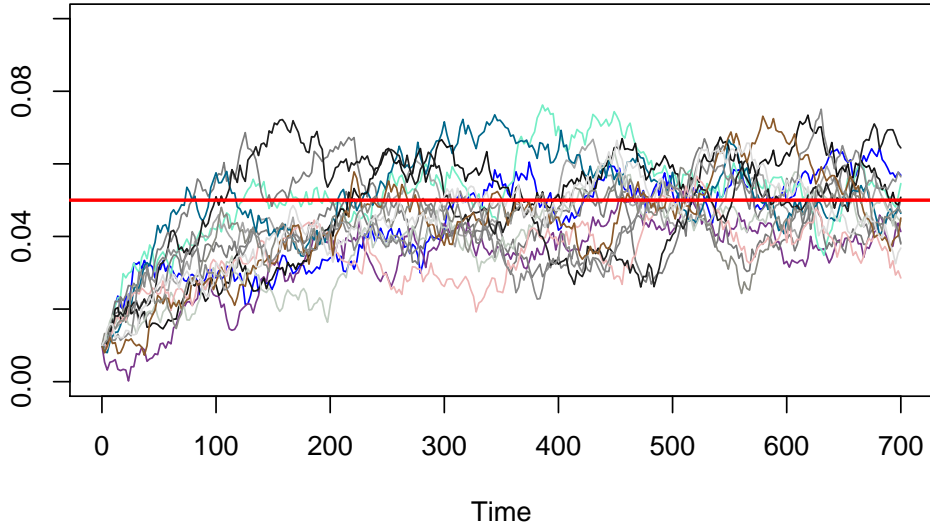
$$dX_t = a(b_1 - X_t)dt + \sigma dB_t, \quad 0 \leq t \leq T.$$

Its solution is given by

$$X_t = e^{-at}X_0 + b_1(1 - e^{-at}) + \sigma \int_0^t e^{-a(t-s)}dB_s,$$

which is derived in Appendix A.

## Simulations of Vasicek interest rate model



**Figure 2:** 15 simulated paths of the Vasicek interest rate model, with  $X_0 = 0.01$ ,  $b_1 = 0.05$ ,  $a = 0.01$ ,  $\sigma = 0.0015$ , maturity = 700.

An example of simulations of the Vasicek model can be found in Figure 2, where we can see the 15 paths stabilizing around the solid red line, which represents  $b_1$ .

Before introducing algorithms and specific choice of functions in order to simulate paths of the signal process and obtain the optimal filter, we will recap the most important results from Chapter 4. In other words, after using the known observation process to simulate the signal process, we will obtain an optimal filter. In order to obtain this optimal filter, we will rely on the change of measure which is described in detail in Chapter 4. Recall that the optimal filter is given by

$$\langle \pi_t, f \rangle = \mathbb{E}[f(X_t) | \mathcal{F}_t^Y] \quad \forall f \in C_b(\mathbb{R}^n), \quad (6.10)$$

where  $\pi_t : \Omega \times \mathcal{B}(\mathbb{R}^n) \rightarrow [0, \infty)$  is the regular conditional probability measure of the signal process  $X_t$  given the  $\sigma$ -algebra  $\mathcal{F}_t^Y$ , generated by  $\{Y_s, 0 \leq s \leq t\}$  and the null sets  $\mathcal{N}$ .

From Theorem 4.1, recall that the optimal filter  $\pi_t$  has the representation

$$\langle \pi_t, f \rangle = \frac{\langle \Psi_t, f \rangle}{\langle \Psi_t, 1 \rangle} \quad (6.11)$$

with

$$\langle \Psi_t, f \rangle = \mathbb{E}[Z_t f(X_t) | \mathcal{F}_t^Y] \quad (6.12)$$

for all  $f \in C_b(\mathbb{R}^n)$ , where

$$\begin{aligned}
Z_t &= \Lambda_t^{-1} \\
&= \exp \left\{ \sum_{i=1}^m \int_0^t h_i(s, X_s) dB_s^i - \frac{1}{2} \int_0^t \|h(s, X_s)\|^2 ds \right. \\
&\quad + \int_0^t \int_{\mathbb{R}_0^m} \log \lambda(s, X_s, z) N(ds, dz) \\
&\quad \left. + \int_0^t \int_{\mathbb{R}_0^m} (1 - \lambda(s, X_s, z)) ds \nu(dz) \right\}, \quad 0 \leq t \leq T,
\end{aligned} \tag{6.13}$$

under  $\mathbb{Q}$ .

As a final note before introducing the general approach to our simulations, it is necessary to know when a jump in the data can be interpreted as such. From Definition 3.6, we know that a jump (or jump size) at time  $t$  of a Lévy process  $Y_t$  ( $Y_t$  because we are interested in the observation process) is defined as

$$\Delta Y(t) = Y(t) - Y(t^-),$$

where  $Y(t^-)$  is the left limit of  $Y(t)$ . As we are observing the observation process directly, we define the jump as  $\Delta Y_i := Y_{i+1} - Y_i$ , where  $i = 1, 2, \dots, T$ , with  $T$  being the final time horizon. Now that the definition of a jump has been established, it remains to define *when* the jump occurs. This can be done in many different ways, but we choose to do this in the following way:

Let  $\overline{\Delta Y}$  denote the average distance in the interval  $[0, T]$ . That is, we define

$$\overline{\Delta Y} := \frac{1}{T} \sum_{i=1}^T |\Delta Y_i|. \tag{6.14}$$

We say a jump occurs if

$$|\Delta Y_i| > \overline{\Delta Y} + \epsilon^*, \tag{6.15}$$

where  $\epsilon^*$  is some chosen constant, usually a small number.

### 6.1.1 Choosing the distribution for claim sizes

Our goal is to model the total claim amount of insurance portfolios which is given by

$$S(t) = \sum_{i=1}^{N(t)} \xi_i^1, \tag{6.16}$$

where  $N(t) = \tilde{N}(\mu^*(t))$  for a process  $\mu^*(t)$  given by (6.5), and  $\xi_i$  is the claim size at time  $0 \leq t \leq T$ . Before constructing our new model, we must first choose a distribution for the individual claim sizes. Our alternatives,

---

<sup>1</sup>Going forward, we choose to denote claim size at time  $i$  as  $\xi_i$  instead of  $X_i$  as in Chapter 5, in order to avoid confusion with the signal process  $X_t, 0 \leq t \leq T$ .

as discussed in Chapter 5, are the log-normal, Gamma and Pareto families. As the Pareto family mainly focuses on very large claim sizes in addition to a small sample size, the preferred families will either be log-normal or Gamma. For our purposes, the choice between the log-normal distribution and Gamma distribution is rather arbitrary. Because of this, going forward, we will focus on the Gamma distribution.

### 6.1.2 A general approach to our model

In this section we will examine the general procedure to our simulations. By the end of this section, we are ready to use algorithms in order to obtain the signal process, the density process and the optimal filter.

Going forward, it is important to distinguish what we wish to estimate, and what is already known. In other words, we wish to estimate *future* claims occurring at time  $t$ , and not claims that already belongs to the past. We make this distinction by denoting the time interval we have data for as  $[0, T]$  and the time interval we wish to estimate data for as  $[0, T^*]$ .

The process for simulating future claim amounts can be constructed as the following:

**Step 1:** Simulate claims using the Gamma distribution (an example may be claims in car insurance) using a frequency drawn from the Poisson distribution. This data is regarded as the information given by  $\mathcal{F}_T^Y$ , that is, the observation process.

**Step 2:** Based on the information given by  $\mathcal{F}_T^Y$ , simulate paths of the signal process  $\hat{X}_T$ .

**Step 3:** Using the information acquired from the observation process and the simulated paths of the signal process, we simulate the density process, which is required for computing the optimal filter.

**Step 4:** Obtain the optimal filter by Monte Carlo. We do this by computing  $\langle \Psi_T, f_i \rangle, i = 1, \dots, m$  and  $\langle \Psi, 1 \rangle$ , given in (6.19) and (6.20), respectively.

**Step 5:** Based on the best estimate for  $\hat{X}_t^i, i = 1, \dots, m$ , we simulate future paths for the signal process. The future paths of the signal process will be denoted as  $\hat{X}_{T^*}$ .

**Step 6:** Simulate future paths of the observation process  $Y_t, 0 \leq t \leq T^*$ , based on the simulated paths of  $\mu^{*(j)}(t), 0 \leq t \leq T^*$ , where  $j = 1, 2, \dots, l$  is the number of simulations.

Hence, we obtain the best estimate for the signal process  $X_t$  by calculating the optimal filter  $\hat{X}_{T^*}$ . The optimal filter is given by the Kallianpur-Striebel-



formula, that is

$$\langle \pi_t, f \rangle = \frac{\langle \Psi_t, f \rangle}{\langle \Psi_t, 1 \rangle}, \quad (6.17)$$

and our goal is to compute  $\langle \pi_t, f \rangle = \hat{X}_t$  where  $0 \leq t \leq T$ .

By Proposition 4.1, we know that

$$\frac{1}{l} \sum_{j=1}^l Z_t^j f(X_t^j) \xrightarrow[l \rightarrow \infty]{} \langle \Psi_t, f \rangle = \mathbb{E}_{\mathbb{Q}}[Z_t f(X_t) | \mathcal{F}_t^Y] \quad \text{a.e.}, \quad (6.18)$$

which means we obtain the optimal filter  $\hat{X}_t$  by simulating  $l$  paths of the density process  $Z_t$  given by (6.13). The two functions needed to compute the optimal filter  $\hat{X}_t = \langle \pi_t, f \rangle$  are, by (6.17),

$$\langle \Psi_t, f \rangle \approx \frac{1}{l} \sum_{j=1}^l Z_t^j f(X_t^j) \quad (6.19)$$

and

$$\langle \Psi_t, 1 \rangle \approx \frac{1}{l} \sum_{j=1}^l Z_t^j, \quad (6.20)$$

for all  $0 \leq t \leq T$ .

In order to construct  $\langle \Psi_t, f \rangle$  and  $\langle \Psi_t, 1 \rangle$ , we need to decide on the following:

- The functions  $h$ ,  $b$  and  $\sigma$ , as needed in the dynamics of  $X_t$ .
- The function for  $\lambda$ , as it is given in the density process  $Z_t$ .
- Choosing a Lévy measure, as it is also given in the density process  $Z_t$ .
- Choosing a distribution for which the future points will be simulated by.
- Choosing a constant for  $\epsilon^*$  that will define when a jump occurs, defined by (6.15).
- Choosing the function  $f$ , as required in  $\langle \Psi_t, f \rangle$ .

The functions  $h$ ,  $b$  and  $\sigma$  have been covered in Section 6.1. The Lévy measure will, in our case, be the distribution of the jump sizes. Hence we define the Lévy measure as  $\nu(dz) := p_Z dz$ , where  $p_Z$  is the probability density function from a chosen probability distribution. As we are simulating claim size in non-life insurance, it would be natural to choose  $p_Z$  as the Gamma distribution, as discussed in Section 6.1.1.

We define the function  $\lambda$  in a way such that we avoid future problems with simulations. If  $\lambda$  depends on the spatial variable coming from jumps, we may encounter problems when trying to simulate the density process. Hence we

will define a new function which will be the main focus of our simulations. The new function is defined as

$$\lambda(t, x, z) := \lambda^*(t, x), \quad 0 \leq t \leq T, \quad (6.21)$$

where  $x \in \mathbb{R}^n$  and  $z \in \mathbb{R}_0^m$ . Finally, we choose the function  $f$  to be given as  $f_i(x) = x_i$ .

## 6.2 Simulating our model

We can construct a rather straightforward algorithm for simulating the path of the signal process  $X$  given by the dynamics

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t^X, \quad 0 \leq t \leq T.$$

Using the dynamics of  $X$  and the fact that  $B_t$  is normally distributed (see Algorithm 1 for the path of Brownian motion), we obtain the following algorithm:

---

**Algorithm 4** Simulating paths of the signal process  $X$

---

Input: Coefficients  $b$ ,  $\sigma$ , final point in our interval  $T$ ,  $n$  (number of intervals), initial values  $(x_i)_{i=1}^m$ .

- 1:  $\Delta t \leftarrow T/n$
  - 2:  $X_{t_0}^i \leftarrow x, \quad i = 1, \dots, m$
  - 3: draw  $\xi_i \sim N(0, 1), \quad i = 1, \dots, n$
  - 4: **for**  $i = 1, \dots, m$  **do**
  - 5:     **for**  $j = 1, \dots, n - 1$  **do**
  - 6:          $X_{t_{j+1}}^{(i)} \leftarrow X_{t_j}^{(i)} + b_i(X_{t_j})\Delta t + \sum_{k=1}^d \sigma_{ik}(X_{t_j})\xi \sqrt{\Delta t}$
  - 7:     **end for**
  - 8: **end for**
  - 9: Return  $X_{t_0}^1, X_{t_1}^2, \dots, X_{t_{n-1}}^m$
- 

In our case, because we choose the coefficient  $b$  as that of a mean reverting process, we will work with values for  $X$  in  $\mathbb{R}^3$ . We do this by defining the coefficients  $b$  and  $\sigma$  in a way that deliberately creates a mean reversion through the Vasicek model. That is, for all  $x \in \mathbb{R}^3$ , let

$$b_i(x) = \begin{cases} x_3(x_2 - x_1) & \text{if } i = 1, \\ 0 & \text{else.} \end{cases}$$

We also define the coefficient  $\sigma$  as

$$\sigma_{i,j}(x) = \begin{cases} 1 & \text{if } i = j = 1, \\ 0 & \text{else.} \end{cases}$$

Hence, for all  $0 \leq t \leq T$ , we get the following dynamics:

$$\begin{cases} dX_t^1 = dX_t^3(X_t^2 - X_t^1)dt + dB_t^{X,1}, \\ dX_t^2 = 0, \\ dX_t^3 = 0, \end{cases} \quad (6.22)$$

where the initial values  $X_0^1, X_0^2, X_0^3$  for the dynamics of  $(X_t^i)_{0 \leq t \leq T}, i = 1, 2, 3$  are to be generated from a distribution which is to be decided on beforehand. For our purposes, we can for example choose either the uniform distribution, or the Gamma distribution.

Having the procedure for the simulations of the signal process  $X$  in place, we continue with the procedure for the observation process  $Y$ . Recall that the dynamics of  $Y$  is given by

$$dY_t = h(t, X_t)dt + dB_t^Y + \int_{\mathbb{R}_0^m} zN_\lambda(dt, dz), \quad 0 \leq t \leq T. \quad (6.23)$$

When simulating our model, we choose to focus on capturing jumps. Hence we let  $h(t, x) = 0, 0 \leq t \leq T, x \in \mathbb{R}^m$ , and we also ignore the Brownian motion in  $Y$ . In other words, we choose to only look at  $Y$ , where  $Y$  follows the dynamics

$$dY_t = \int_{\mathbb{R}_0^m} zN_\lambda(dt, dz), \quad 0 \leq t \leq T, \quad (6.24)$$

where the predictable compensator  $\hat{\mu}$  is given as in (4.3). That is,

$$\begin{aligned} \hat{\mu}(dt, dz, \omega) &= \lambda(t, X_t, z)dt\nu(dz), \\ &\stackrel{(*)}{=} \lambda^*(t, X_t)dt\nu(dz) \quad 0 \leq t \leq T, z \in \mathbb{R}_0^m, \end{aligned}$$

where  $(*)$  uses  $\lambda(t, x, z)$  as defined in (6.21). It remains to choose a function  $\lambda^*(t, X_t)$ . We define this in the following way:

$$\begin{aligned} \lambda^*(t, X_t) &:= \epsilon \|X_t\|_1 \\ &= \epsilon \sum_{i=1}^3 |X_t^i|, \quad 0 \leq t \leq T, \end{aligned} \quad (6.25)$$

where  $\epsilon \in \mathbb{R}_{>0}$ . The Lévy measure was defined in Section 6.1.2 as  $\nu(dz) = p_Z dz$ . Because  $p_Z$  is the density of a probability function, we know that

$$\int_{\mathbb{R}_0^m} p_Z dz = 1. \quad (6.26)$$

Our observation process  $Y$  is now given by:

$$\begin{aligned} Y_t &= \int \int_{[0,t] \times \mathbb{R}_0^3} zN_\lambda(ds, dz) \\ &= \sum_{i=1}^{\tilde{N}(\mu^*(t))} \xi_i, \end{aligned} \quad (6.27)$$

where

- $\tilde{N}$  is a Poisson process with intensity equal to 1. That is,  $\tilde{N} = (\tilde{N}(t))_{0 \leq t \leq T}$ , with the property of stationary increments which states that

$$\tilde{N}\left(\mu^*(t_{j+1}) - \mu^*(t_j)\right) \sim \text{Poisson}\left(\int_{t_j}^{t_{j+1}} \lambda^*(s, X_s) ds\right), \quad j = 0, 1, 2, \dots, n-1. \quad (6.28)$$

Here  $\tilde{N}$  is independent of the signal process  $X$ .

- The intensity  $\mu^*(t)$  is given by

$$\mu^*(t) = \int_0^t \lambda^*(s, X_s) ds, \quad 0 \leq t \leq T. \quad (6.29)$$

- $(\xi_i)_{i \in \mathbb{N}}$  is a sequence of i.i.d. stochastic variables that follows the Gamma distribution.

In Algorithm 5, the integral

$$\mu^*(t) = \int_0^t \lambda^*(s, X_s) ds, \quad 0 \leq t \leq T,$$

is approximated by the classic trapezoidal rule.

---

**Algorithm 5** Simulating paths of the observation process  $Y$ .

---

Input: Paths of the signal process  $X$ ,  $n$  for the number of intervals, final time  $T$ , function  $\lambda^*(t, X_t)$ , jump size distribution Gamma with shape parameter  $\alpha_i$  and scale parameter  $\beta_i$  for  $i = 1, 2, 3$

```

1:  $\mu_{t_0}^* \leftarrow 0$ 
2:  $\Delta t \leftarrow T/n$ 
3:  $\lambda^*(t, X_t) \leftarrow \epsilon \sum_{i=1}^3 |X_t^i|$ 
4: for  $j = 0, \dots, n-1$  do
5:    $\mu_{t_{j+1}}^* = \mu_{t_j}^* + \Delta t/2 \left( \lambda^*(t_j, X_{t_j}) + \lambda^*(t_{j+1}, X_{t_{j+1}}) \right)$ 
6:   draw  $M \sim \text{Poisson}(\mu_{t_{j+1}}^* - \mu_{t_j}^*)$ 
7:   for  $i = 1, \dots, m$  do
8:     for  $k = 1, \dots, M$  do
9:       draw  $\xi_k^{j,i} \sim \text{Gamma}(\alpha_i, \beta_i)$ 
10:       $Y_{t_{j+1}}^i \leftarrow Y_{t_j}^i + \xi_k^{j,i}$ 
11:     end for
12:   end for
13: end for
19: end for
20: Return  $Y_{t_0}^1, Y_{t_1}^2, \dots, Y_{t_{n-1}}^m$ 

```

---

Finally, it remains to construct an algorithm for simulating the density process  $Z_t, 0 \leq t \leq T$ . As we choose to let  $h(t, x) = 0$  and ignore Brownian

motion, our density function is represented as the following:

$$\begin{aligned}
Z_t &= \exp \left\{ \int_0^t \int_{\mathbb{R}_0^3} \log \lambda^*(s, X_s) N(ds, dz) \right. \\
&\quad \left. + \int_0^t \int_{\mathbb{R}_0^3} (1 - \lambda^*(s, X_s)) ds \nu(dz) \right\} \\
&\stackrel{(*)}{=} \exp \left\{ \int_0^t \int_{\mathbb{R}_0^3} \log \lambda^*(s, X_s) N(ds, dz) \right. \\
&\quad \left. + \int_0^t \int_{\mathbb{R}_0^3} (1 - \lambda^*(s, X_s)) p_Z ds dz \right\} \\
&\stackrel{(**)}{=} \exp \left\{ \sum_{0 \leq s \leq t} \log \left( \epsilon \|X_s\|_1 \right) \mathbb{1}_{\{\Delta Y_s \neq 0\}} \right. \\
&\quad \left. + \int_0^t (1 - \epsilon \|X_s\|_1) ds \right\}, \quad 0 \leq t \leq T, \tag{6.30}
\end{aligned}$$

where (\*) uses the fact that  $\nu(dz) = p_Z dz$ , and (\*\*) uses  $\lambda^*(t, X_t)$  as defined in (6.21).

By (6.15), we know that a jump occurs if  $|\Delta Y_i| > \overline{\Delta Y} + \epsilon^*$ ,  $0 \leq i \leq T$ . When incorporating this into our model, the density function that will be used in our simulations is represented as the following:

$$\begin{aligned}
Z_t &= \exp \left\{ \underbrace{\sum_{0 \leq s \leq t} \log \left( \epsilon \|X_s\|_1 \right) \mathbb{1}_{\{|\Delta Y_s| > \overline{\Delta Y} + \epsilon^*\}}}_{=I_1} \right. \\
&\quad \left. + \underbrace{\int_0^t (1 - \epsilon \|X_s\|_1) ds}_{=I_2} \right\}, \quad 0 \leq t \leq T. \tag{6.31}
\end{aligned}$$

The final algorithm needed is the one simulating the density process  $Z$  given by (6.31), and is shown on the next page.

---

**Algorithm 6** Simulating paths of the density process  $Z_t, 0 \leq t \leq T$ 

---

Input: Paths of the signal process  $X$ ,  $n$  for the number of intervals, final time  $T$ , function  $\lambda^*(t, X_t)$ , observation process  $Y$ ,  $\epsilon^*$ , number of simulations  $nsim$ .

```
1:  $\Delta t \leftarrow T/n$ 
2:  $I_1 \leftarrow 0$ 
3:  $I_2 \leftarrow 0$ 
4:  $\lambda^*(t, X_t) \leftarrow \epsilon \sum_{i=1}^3 |X_t^i|$ 
5: for  $j = 0, \dots, n - 1$  do
6:    $\Delta Y_j \leftarrow Y_{j+1} - Y_j$ 
7: end for
8:  $\overline{\Delta Y} \leftarrow \frac{1}{T} \sum_{i=0}^{T-1} |\Delta Y_i|$  (As defined in (6.14))
9: for  $l = 1, \dots, nsim$  do
10:   for  $j = 0, \dots, T - 1$  do
11:     if  $|\Delta Y_j| > \overline{\Delta Y} + \epsilon^*$  then
12:        $I_{1j} \leftarrow \log(\lambda^*(j, X_j))$ 
13:     end if
14:   end for
15: end for
16:  $I_1 \leftarrow \sum_{j=0}^{T-1} I_{1j}$ 
17:  $S \leftarrow \sum_{j=1}^{n-1} \left(1 - \lambda^*(t_j, X_{t_j})\right)$ 
18:  $I_2 \leftarrow \Delta t \left[ S + \left\{ 1 - \frac{1}{2} \left( \lambda(0, X_0) + \lambda^*(t, X_{t_n}) \right) \right\} \right]$ 
19:  $Z_T \leftarrow \exp(I_1 + I_2)$ 
20: Return  $Z_T$ 
```

---

### 6.2.1 Numerical results

When simulating the total claim amounts, we conduct the following steps:

Step 1: **Create imaginary claims for a car insurance portfolio.**

We simulate these claims by assuming we have a portfolio consisting of the following:

- Let  $C = 10\,000$ , where  $C$  denotes the number of cars in our portfolio.
- We let the average claim frequency be equal to 2%. That is, we let  $\bar{\mu} = 2\%$ . This is denoted as the common claim frequency as given in Example 5.2. In our simulations, this claim frequency will be drawn for each day.
- Claim sizes are assumed to have an average of 10 000 NOK per claim.
- We look at the development of total claim amount for  $T = 500$  days.
- Claim sizes are given in 10 000 NOK.

As a result, we assume that out of 10 000 cars, 200 will make an insurance claim, and we expect each claim to be 10 000 NOK. This data is now considered as the information given by the observation process, which has the filtration  $\mathcal{F}_T^Y$ .

**Step 2: Simulate paths of the signal process.**

Simulate 15 paths of the process  $X^1$  given by (6.22), using Algorithm 4 and the generated initial values  $X_0^1, X_0^2, X_0^3$ . By doing this, we acquire 15 paths of the integral defined in (6.5).

**Step 3: Simulate the density process.**

Using Algorithm 6, the simulated signal process and the observation process, we simulate the density process  $(Z_t)_{0 \leq t \leq T}$ .

**Step 4: Obtain the optimal filter by Monte Carlo.**

Having simulated the paths of the process  $X^1$  from step 2, we compute  $\langle \Psi_t, f_i \rangle, i = 2, 3$  and  $\langle \Psi_t, 1 \rangle$  as given in (6.19) and (6.20), respectively. Once these have been computed, we obtain the optimal filters

$$\hat{X}_t^2 = \langle \pi_t, f_2 \rangle \text{ and } \hat{X}_t^3 = \langle \pi_t, f_3 \rangle, \quad 0 \leq t \leq T,$$

where  $(\langle \pi_t, f_i \rangle)_{i=2,3}$  is given by (6.17).

**Step 5: Simulate future paths of  $X^1$ .**

After obtaining the optimal filters from step 4, we simulate future paths of the process  $(X_t^1)_{0 \leq t \leq T^*}$ , which follows the dynamics given by (6.22):

$$dX_t^1 = d\hat{X}_t^3(\hat{X}_t^2 - X_t^1)dt + dB_t^{X,1}, \quad 0 \leq t \leq T.$$

**Step 6: Simulate future paths of the observation process.**

Knowing the future paths of our signal process  $X$ , we obtain 15 future paths for  $\mu^*(t)$ . That is, we obtain

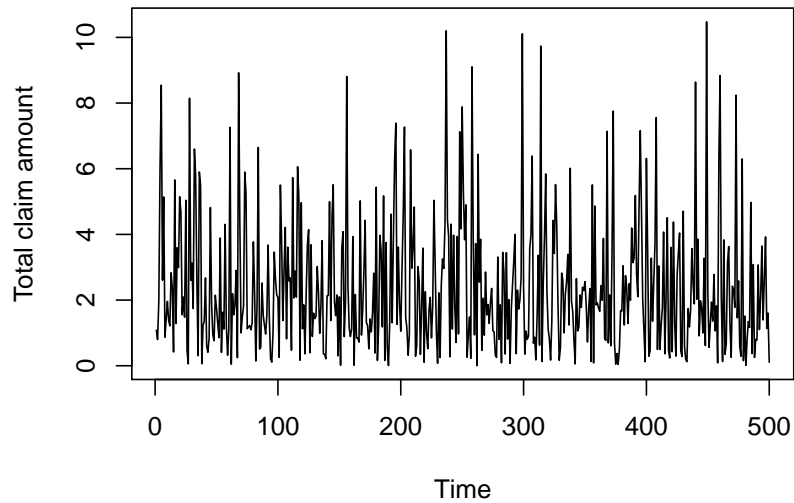
$$\mu^{*(j)}(t) = \int_0^t \lambda^*(s, X_s^{(j)})ds, \quad j = 1, 2, \dots, 15, \quad 0 \leq t \leq T^*.$$

Based on these simulations, we obtain the simulated future values for the observation process  $Y$  given by, for all  $0 \leq t \leq T^*$ ,

$$Y_t^{(j)} = \sum_{i=1}^{\tilde{N}(\mu^{*(j)}(t))} \xi_i, \quad j = 1, 2, \dots, 15, \quad \xi_i \sim \text{Gamma}(1, 1),$$

where  $\tilde{N}(\mu^*(t))$  is given in (6.28). Hence,  $(Y_t)_{0 \leq t \leq T^*}$  is regarded as the total claim amount at a future time  $t$ , where  $0 \leq t \leq T^*$ .

By step 1, we obtain the evolution of total claim amounts for the time interval  $[0, T]$ , shown in Figure 3.



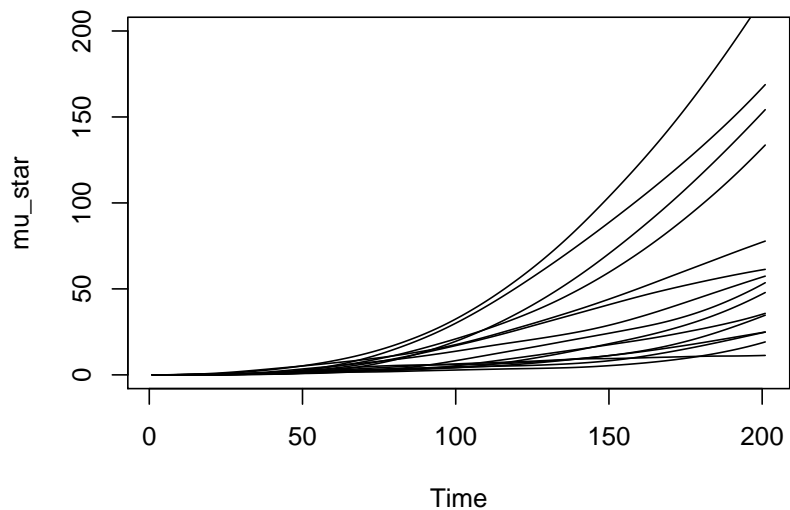
**Figure 3:** Evolution of total claim amount over 500 days.

For the simulations,  $\epsilon^*$  was chosen rather arbitrarily, and was set equal to 4, which gave 28 jumps in our model. In addition,  $\epsilon$  was set equal to 0.001 in order to avoid simulation problems with values that are too extreme.

When performing steps 2-6, we obtain 15 paths for

$$\mu^{*(j)}(t), \quad j = 1, 2, \dots, 15, \quad 0 \leq t \leq T^*,$$

which is shown in Figure 4.



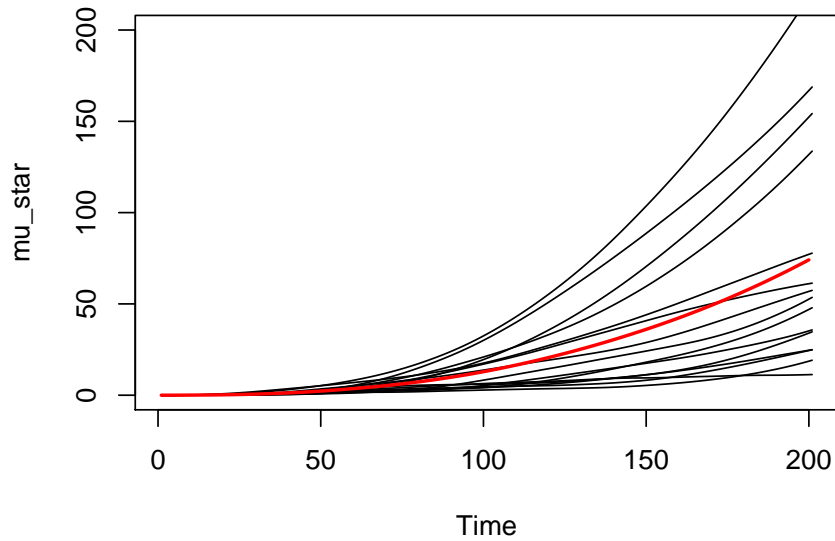
**Figure 4:** 15 paths for  $\mu^*(t)$ .



The 15 paths of  $\mu^*(t)$  given in Figure 4 are simulated based on the optimal filters

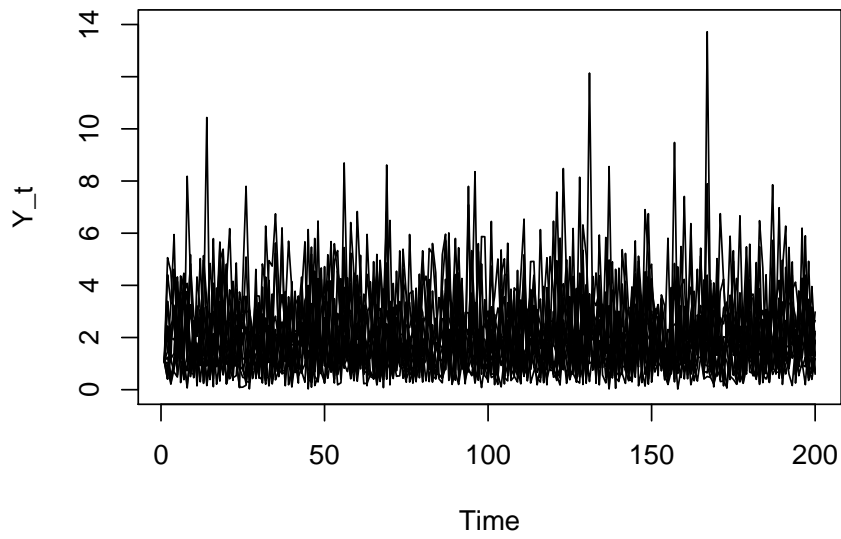
$$\begin{aligned}\hat{X}_t^2 &= 1.04275, \\ \hat{X}_t^3 &= 0.8792897.\end{aligned}$$

The average path for  $\mu^{*(j)}(t), 0 \leq t \leq T^*, j = 1, 2, \dots, 15$  is shown by a thick red line in Figure 5.



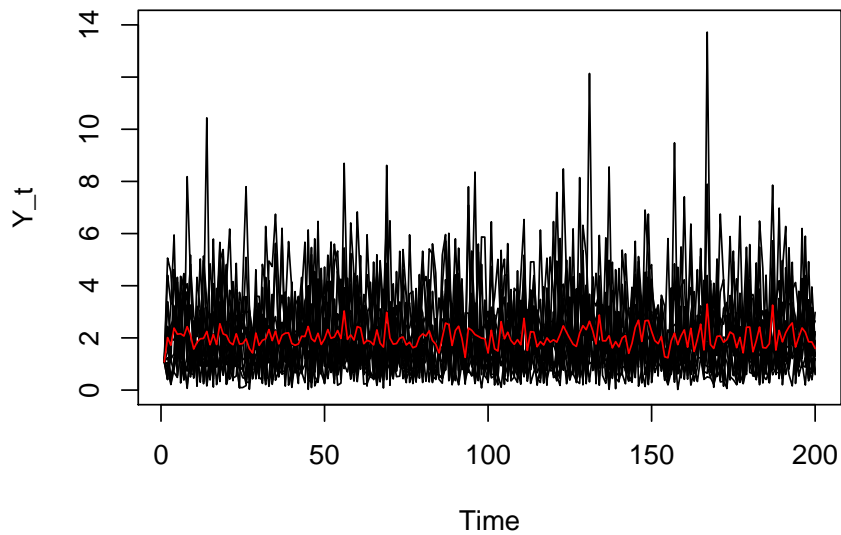
**Figure 5:** Average path for  $\mu^*(t)$ .

Continuing, we simulate 15 paths of the observation process  $Y$ , using the simulated paths for  $\mu^{*(j)}(t), 0 \leq t \leq T^*, j = 1, 2, \dots, 15$ , in Figure 6.



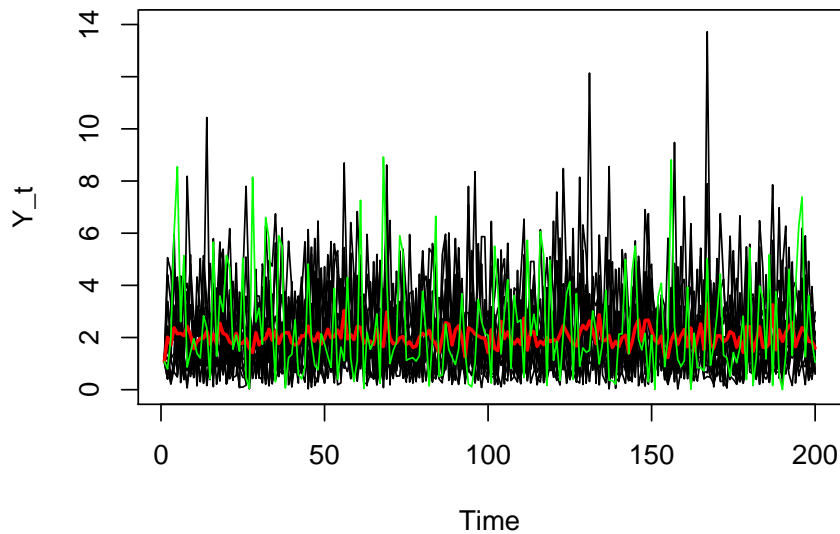
**Figure 6:** 15 simulated path for the observation process.

The average path for the 15 simulations of the observation process  $Y_t, 0 \leq t \leq T^*$  is shown as a thick red line in Figure 7.



**Figure 7:** Average path for 15 simulations of  $Y_t, 0 \leq t \leq T^*$ .

For comparison, we plot the first 200 days of claim sizes from step 1, 15 paths of the observation process  $Y_t$  for  $0 \leq t \leq T^*$  and the average of the observation process, in Figure 8.



**Figure 8:** The 15 simulated paths of  $Y$  for a future time  $[0, 200]$  are colored as black, the average for the 15 simulated paths is plotted with a thick red color, and the simulated claim sizes from step 1 is colored as green.

The simulations of our new model shows a fairly stable prediction of the total claim amounts from step 1, as shown in Figure 8.

### 6.3 Conclusion

In this thesis we have developed a new stochastic model for modelling total claim amounts in non-life insurance. We have introduced the necessary mathematical theory needed in connection with the development of our model. Most importantly, we introduced theory for stochastic analysis, Lévy processes and non-linear filtering. Based on the framework of generalized Cox processes, we were able to simulate a new model which can capture both short term (by jump processes) and long term (by a discontinuous vector field) shocks in the insurance market.

The simulations in Chapter 6 attempt to predict future values for total claim amounts, based on insurance data given in the past. However, our simulations are based on imaginary claims which are drawn from a Gamma distribution. As a result, it may be difficult to compare predicted values to the given data, considering both are drawn from the same distribution. For future work, it would be interesting to test our new model on real world insurance data, to see whether the model accurately predicts total claim amounts.

A model which examines stochastic transition rates has been discussed in Bølviken, Duedahl, Proske [6], which uses the concepts of non-linear filtering

techniques, but only applied in the realm of *life insurance*. By applying the same ideas developed in Bølviken, Duedahl, Proske [6], we were able to develop a new model in the realm of *non-life insurance*. Our new model is able to capture regime switching effects from data, for example the occurrence of natural disasters having a long-lasting impact, for example on the economy of a country. Not only is the model able to capture shocks in the market through regime switching and mean reversion as done in Chapter 6, it can also capture different phenomenons through the unknown parametrization process. This makes the model quite flexible for the user.

To summarize; our new model is flexible to the user, meaning the user can choose which phenomenons (types of jumps and switching between models with different types of jumps) they wish to capture. Even though we compare predicted values to the given data where they both are drawn from the same distribution, it is reasonable to conclude that our new model gives a stable prediction of given insurance data. It remains for future work to implement this model on real world insurance data and see if it gives good predictive values in the long run.

# Chapter 7

## Extensions

In this chapter we will look at extensions and future work that can be performed on the model constructed in Chapter 6. This may be beneficial in order to capture different scenarios occurring in the insurance market. Examples of such scenarios may be a financial crisis or the occurrence of war.

### 7.1 The signal process

The signal process  $X$  is given by the dynamics

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t^X, \quad 0 \leq t \leq T,$$

where  $b, \sigma$  are Borel functions and  $B_t^X$ ,  $0 \leq t \leq T$  is a Brownian motion. Our model in Chapter 6 takes the following definitions for  $b$  and  $\sigma$ :

$$b_i(x) = \begin{cases} x_3(x_2 - x_1) & \text{if } i = 1, \\ 0 & \text{else.} \end{cases}$$
$$\sigma_{i,j}(x) = \begin{cases} 1 & \text{if } i = j = 1, \\ 0 & \text{else.} \end{cases} \quad (7.1)$$

We will now examine different choices for the coefficients  $b$  and  $\sigma$ .

#### 7.1.1 Choosing a different coefficient for $b$

The coefficient  $b$  can be changed from a mean reversion model to something that captures different shocks from the insurance market. As an example: for all  $0 \leq t \leq T$ , choose the following dynamics for  $X$ :

$$\begin{cases} dX_t^1 = \left( X_t^3 \mathbb{1}_{(-\infty, X_t^5]}(X_t^1) + X_t^4 \mathbb{1}_{(X_t^5, \infty)}(X_t^1) \right) (X_t^2 - X_t^1) dt + dB_t^{X^1,1}, \\ dX_t^2 = 0, \\ dX_t^3 = 0, \\ dX_t^4 = 0, \\ dX_t^5 = 0, \end{cases}$$

where  $X_t^5$  is the threshold process to be estimated, and the initial values  $X_0^1, X_0^2, X_0^3, X_0^4, X_0^5$  for the dynamics of  $(X_t^i)_{0 \leq t \leq T}$ ,  $i = 1, 2, 3, 4, 5$  are to be generated from a distribution which is to be decided on beforehand.

#### 7.1.2 Choosing a different coefficient for $\sigma$

It would be interesting to allow the coefficient  $\sigma$  to capture regime switching. In other words, it would be interesting to capture regime switching through the coefficient  $\sigma$  by choosing  $\sigma_{i,j}$  to be equal to something different from (7.1).

If however  $\sigma$  is discontinuous, a strong solution with respect to the stochastic differential equation for the signal process does not in general exist. In this case, there is no systematic theory for the construction of such solutions in the current literature.

## 7.2 The observation process

The observation process in Chapter 6 is given by

$$dY_t = h(t, X_t)dt + dB_t^Y + \int_{\mathbb{R}_0^m} zN_\lambda(dt, dz), \quad 0 \leq t \leq T. \quad (7.2)$$

In our simulations, we chose to let  $h(t, x) = 0, 0 \leq t \leq T, x \in \mathbb{R}^m$ , and we also chose to ignore the Brownian motion in  $Y$ . For future work, it would be interesting to let  $h(t, x) \neq 0$  and not ignore Brownian motion, to examine how this would affect the simulations constructed in Chapter 6.

## 7.3 Claim size distribution

Naturally, for future work, it would be interesting to have real world insurance data instead of simulated claims as demonstrated in Chapter 6. Besides looking at real world insurance claims, it would be interesting to let the claim sizes in (6.27) to follow either the log-normal or Pareto distribution. Keep in mind that we simulated claims from an imaginary car insurance portfolio. If one were to for example apply this model on a portfolio of property insurance, it could be natural to let the claims follow the Pareto distribution.

## 7.4 Calculating the reserve

For future work, it would be interesting to estimate the reserve for the simulated total claim amounts. That is, we calculate the reserve by looking at the given total claim amount for a future time interval  $0 \leq t \leq T^*$ :

$$Y_t^{(j)} = \sum_{i=1}^{\tilde{N}(\mu^{*(j)}(t))} \xi_i, \quad j = 1, 2, \dots, 15, \quad \xi_i \sim \text{Gamma}(1, 1).$$

Using the simulations of the observation process given above, let

$$\begin{aligned} \mathcal{X}^{(j)} &= Y_t^{(j)} \\ &= \sum_{i=1}^{\tilde{N}(\mu^{*(j)}(t))} \xi_i, \quad \xi_i \sim \text{Gamma}(1, 1), \end{aligned}$$

for the simulations  $j = 1, 2, \dots, 15$ . The reserve  $q_\epsilon^{(j)}$  can then be calculated by letting

$$\Pr(\mathcal{X}^{(j)} > q_\epsilon^{(j)}) = \epsilon,$$

where  $\epsilon$  is some small, positive number. The average reserve can be calculated by

$$\bar{q}_\epsilon = \frac{1}{15} \sum_{j=1}^{15} q_\epsilon^{(j)}.$$

# Appendix A - Calculations

## 8.1 A.1 - Chapter 3

### A.1.1 - Proof of proposition 3.3

(i) We want to prove that  $\left(\frac{e^{iuX_t}}{\mathbb{E}[e^{iuX_t}]} \right)_{t \geq 0}$  is a martingale for all  $u \in \mathbb{R}$ . Denote the function  $g$  as the cumulant of  $L_1$ , that is,  $g(u) = \log \mathbb{E}[e^{iuX_t}]$ . Define the process  $M = (M_t)_{0 \leq t \leq T}$  as

$$M_t = \frac{e^{iuX_t}}{e^{tg(u)}}.$$

By Theorem 3.2 in Cont, Tankov [7], we have that  $M_t < \infty$  for all  $0 \leq t \leq T$ . Furthermore, we obtain the following:

For  $0 \leq s \leq t$ , we can re-write  $M_t$  as

$$M_t = \frac{e^{iuX_s}}{e^{sg(u)}} \frac{e^{iu(X_t - X_s)}}{e^{(t-s)g(u)}} = M_s \frac{e^{iu(X_t - X_s)}}{e^{(t-s)g(u)}}$$

Using the property of independent increments, we can conclude

$$\begin{aligned} \mathbb{E}[M_t | \mathcal{F}_s] &= M_s \mathbb{E}\left[\frac{e^{iu(X_t - X_s)}}{e^{(t-s)g(u)}} \middle| \mathcal{F}_s\right] = M_s e^{(t-s)g(u)} e^{-(t-s)g(u)} \\ &= M_s. \end{aligned}$$

Hence  $M_t$  is a martingale.

(ii) We want to prove that  $\left(\frac{e^{uX_t}}{\mathbb{E}[e^{uX_t}]} \right)_{t \geq 0}$  is a martingale, where  $u \in \mathbb{R}$ ,  $\mathbb{E}[e^{uX_t}] < \infty \forall t \geq 0$ . Similarly to (i), denote the function  $g$  as the cumulant of  $L_1$ , that is,  $g(u) = \log \mathbb{E}[e^{uX_t}]$ . Define the process  $M = (M_t)_{0 \leq t \leq T}$  as

$$M_t = \frac{e^{uX_t}}{e^{tg(u)}}.$$

We obtain the following:

For  $0 \leq s \leq t$ , we can re-write  $M_t$  as

$$M_t = \frac{e^{uX_s}}{e^{sg(u)}} \frac{e^{u(X_t - X_s)}}{e^{(t-s)g(u)}} = M_s \frac{e^{u(X_t - X_s)}}{e^{(t-s)g(u)}}$$

Using the property of independent increments, we can conclude

$$\begin{aligned} \mathbb{E}[M_t | \mathcal{F}_s] &= M_s \mathbb{E}\left[\frac{e^{u(X_t - X_s)}}{e^{(t-s)g(u)}} \middle| \mathcal{F}_s\right] = M_s e^{(t-s)g(u)} e^{-(t-s)g(u)} \\ &= M_s. \end{aligned}$$

Hence  $M_t$  is a martingale.

(iii) We want to prove that if  $\mathbb{E}[X_t] < \infty \forall t \geq 0$  then the process  $M_t =$



$X_t - \mathbb{E}[X_t]$  is a martingale.

We have for all  $0 \leq s \leq t \leq T$ :

$$\begin{aligned} \mathbb{E}[M_t | \mathcal{F}_s] &= \mathbb{E}[X_t | \mathcal{F}_s] - \mathbb{E}[X_t] \\ &\quad (\text{Use the simple manipulation } X_t = X_t + X_s - X_s) \\ &= \mathbb{E}[X_t - X_s | \mathcal{F}_s] + X_s - \mathbb{E}[X_t] \\ &= \mathbb{E}[X_t] - \mathbb{E}[X_s] + X_s - \mathbb{E}[X_t] \\ &= X_s - \mathbb{E}[X_s] \end{aligned}$$

Hence  $M_t$  is a martingale.

(iv) We want to prove that the process  $(M_t)^2 - \mathbb{E}[(M_t)^2]$  is a martingale. We will use the result from (iii); that is,  $M_t$  is a martingale with the property of independent increments. For all  $0 \leq s \leq t \leq T$ , we obtain:

$$\begin{aligned} \mathbb{E}[(M_t)^2 - \mathbb{E}[(M_t)^2] | \mathcal{F}_s] &= \mathbb{E}[(M_t - M_s + M_s)^2 | \mathcal{F}_s] - \mathbb{E}[(M_t)^2] \\ &= \mathbb{E}[(M_t - M_s)^2 | \mathcal{F}_s] + 2\mathbb{E}[M_s(M_t - M_s) | \mathcal{F}_s] \\ &\quad + \mathbb{E}[(M_s)^2 | \mathcal{F}_s] - \mathbb{E}[(M_t)^2] \\ &= \mathbb{E}[(M_t - M_s)^2] + 2M_s\mathbb{E}[M_t - M_s] \\ &\quad + (M_s)^2 - \mathbb{E}[(M_t)^2] \\ &= \mathbb{E}[(M_t)^2] - 2\mathbb{E}[M_t M_s] + \mathbb{E}[(M_s)^2] - \mathbb{E}[(M_t)^2] \\ &\quad + 2M_s[\mathbb{E}[\mathbb{E}[M_t | \mathcal{F}_s]] - \mathbb{E}[M_s]] + (M_s)^2 \\ &= (M_s)^2 - 2\mathbb{E}[M_s \mathbb{E}[M_t | \mathcal{F}_s]] \mathbb{E}[(M_s)^2] \\ &= (M_s)^2 - \mathbb{E}[(M_s)^2] \end{aligned}$$

Hence for all  $0 \leq t \leq T$ ,  $(M_t)^2 - \mathbb{E}[(M_t)^2]$  is a martingale. ■

## 8.2 A.2 - Chapter 6

### The Vasicek model.

The dynamics of the signal process  $X$  is given by

$$dX_t = b(X_t)dt + \sigma dB_t, \quad 0 \leq t \leq T.$$

When we wish to capture mean reversion, we look at the Vasicek model in chapter 6. Hence the dynamics of  $X$  is now given by

$$dX_t = a(b_1 - X_t)dt + \sigma dB_t, \quad (8.1)$$

where  $a$  is the mean reversion coefficient and  $b_1$  is the stabilization level in the long run. The equation (8.1) can be solved using Itô's formula given in Theorem 2.5: Define  $f(t, x) = e^{at}x$ . This gives

$$\begin{aligned} (i) \quad & \frac{\partial f}{\partial t} = ae^{at}x \\ (ii) \quad & \frac{\partial f}{\partial x} = e^{at} \\ (iii) \quad & \frac{\partial^2 f}{\partial x^2} = 0. \end{aligned}$$

Applying Itô's formula yields

$$\begin{aligned}df(t, X_t) &= e^{at} dX_t = aX_t e^{at} dt + e^{at} dX_t \\&= aX_t e^{at} dt + e^{at} a(b_1 - X_t) dt + e^{at} \sigma dB_t \\&= ab_1 e^{at} dt + e^{at} \sigma dB_t.\end{aligned}$$

On integral form, this becomes

$$f(t, X_t) = e^{at} X_t = X_0 + ab_1 \int_0^t e^{as} ds + \sigma \int_0^t e^{as} dB_s.$$

Solving for  $X_t$  yields

$$\begin{aligned}X_t &= e^{-at} X_0 + ab_1 \int_0^t e^{-a(t-s)} ds + \sigma \int_0^t e^{-a(t-s)} dB_s \\&= e^{-at} X_0 + ab_1 \frac{1}{a} (1 - e^{-at}) + \sigma \int_0^t e^{-a(t-s)} dB_s \\&= e^{-at} X_0 + b_1 (1 - e^{-at}) + \sigma \int_0^t e^{-a(t-s)} dB_s.\end{aligned}$$

## Appendix B - Distributions

### 9.1 B.1 - The normal distribution

A random variable is said to be normally distributed if it has the probability density function

$$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-(x - \mu)^2}{2\sigma^2}\right),$$

where  $\mu$  is the mean and  $\sigma^2$  is the variance.

If  $X$  is a random variable taking values in  $\mathbb{R}$  and  $X$  is normally distributed, we write  $X \sim N(\mu, \sigma^2)$ .

### 9.2 B.2 - Poisson distribution

A random variable  $X$ ,  $X \in \mathbb{N} \cup \{0\}$  is Poisson distributed if it has the probability mass function given by

$$P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad k = 0, 1, \dots,$$

where  $\lambda$  is the expectation and the variance. If  $X$  is Poisson distributed, we write  $X \sim \text{Poisson}(\lambda)$ .

### 9.3 B.3 - Gamma distribution

Let  $X$ ,  $X \in \mathbb{R}$  be a random variable.  $X$  is Gamma distributed if it has the probability density function given by

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-x/\beta}, \quad x > 0,$$

with  $\mathbb{E}(X) = \alpha\beta$  and  $\text{var}(X) = \alpha\beta^2$ .

### 9.4 B.4 - Pareto distribution

Random variables are said to be Pareto distributed if they have the density function

$$f(x) = \frac{\alpha/\beta}{(1 + x/\beta)^{1+\alpha}}, \quad x > 0,$$

with  $\mathbb{E}(X) = \frac{\beta}{\alpha-1}$  and  $\text{sd}(X) = \mathbb{E}(X) \sqrt{\frac{\alpha}{\alpha-2}}$ .

## Appendix C - R-code

### 10.1 C.1 - Chapter 2

Simulating a stock price with brownian motion:

```
1 n = 1000
2 y = rnorm(n)
3 x = y
4
5 for(i in 1:n) {
6 x[i] = 1/sqrt(n)*(sum(y[1:i])*sqrt(i))
7 }
8
9 plot(x,type="l", xlab = "Time", ylab = "Profit", main =
    "Evolution of a stock price with Brownian motion")
```

### 10.2 C.2 - Chapter 5

Simulating the density function of claims that are log-normally distributed:

```
1 m = 1000
2 lambda = 10
3 xi = 0
4 sigma = 1
5 N = rpois(m,lambda)
6 X = 1:m*0
7 for (i in 1:m) {
8 Z = rlnorm(N[i],xi,sigma)
9 X[i] = sum(Z)
10 }
11 plot(density(X), main = "Density of log-normal claims",
    xlab = "")
```

### 10.3 C.3 - Chapter 6

Simulating the Vasicek model:

```
1 #-----
2 # model parameters
3 #-----
4 X0 = 0.01
5 b1 = 0.05
6 a = 0.01
7 sigma = 0.0015
8 T0 = 700
9
10 #-----
11 # simulation parameters
12 #-----
13
14 N = 300
```

```

15 nsim = 15
16 maturity = T0
17
18
19 dt = maturity/(N+1)
20 timeline = seq(0,maturity, dt)
21
22 f = matrix(X0,(N+2),nsim)
23
24 vasicek_rate <- function(r,a,theta,sigma,dt){
25   expadt = exp(-a*dt)
26   volatility = (sigma^2)*(1-expadt^2)/(2*a)
27   result = r*expadt+theta*(1-expadt)+sqrt(volatility)*
      rnorm(1)
28   return(result)
29 }
30
31 for(i in 2:(N+2)){
32   for(j in 1:nsim){
33     f[i,j] = vasicek_rate(f[i-1,j],a,b1,sigma,dt)
34   }
35 }
36
37
38 #-----
39 # plot of interest rate simulations
40 #-----
41
42 plot(timeline,f[,1], ylim=range(0,0.1), type="l", col="
      blue", ylab="", xlab = "Time", main = "Simulations
      of Vasicek interest rate model")
43 for(j in 2:nsim){
44   lines(timeline,f[,j], col=colors()[floor(runif(1,1,657)
      )])
45 }
46 abline(h = b1, col = "red", lwd=2)
47
48 mean_f = vector(mode = "numeric", length = T0)
49 for(j in 1:T0) {
50   mean_f[j] = mean(f[j,])
51 }
52 plot(mean_f, type="l", ylim=range(-0.1,0.5), main = "
      Average of 15 simulations of the Vasicek interest
      model",ylab = "", xlab = "Time")
53 abline(h = b1, col = "red")
54 lines(mean_f, col="black")

```

Simulating our new stochastic model:

```

1 #-----
2 #Simulating claim size data for 500 days
3 #-----
4 days = 500
5 ClaimSize = vector(mode="numeric", length = days)

```

```

6 temp = vector(mode="numeric", length = 1)
7
8 expectedClaimSize = 1
9 expectedClaimFrequency = 2
10
11 for(i in 1:days) {
12   for(j in 1:1) {
13     N = rpois(1, expectedClaimFrequency)
14     temp[j] = 0
15     for(k in 1:N) {
16       Z = rgamma(1, expectedClaimSize)
17       temp[j] = temp[j] + Z
18     }
19   }
20   ClaimSize[i] = mean(temp)
21 }
22
23 plot(ClaimSize, type = "l", xlab = "Time", ylab = "Total
      claim amount")
24
25 #-----
26 #Creating our observation process
27 #-----
28 Y = c()
29 Y = ClaimSize
30 nsim = 15 #Number of simulations
31 T = days #Final time in our interval
32 N = 500 #Number of steps in the interval [0,T]
33 dt = T/N #Step size
34 delta_Y = c()
35 delta_Y[1] = Y[1] #Simply the first value from the
      observation process
36
37 for(i in 1:(length(Y)-1)) {
38   delta_Y[i+1] = Y[i+1]-Y[i]
39 } #Creating delta_Y vector
40
41 #-----
42 #Simulate the coefficients in the Vasicek model
43 #-----
44
45 L = 100 #Number of times we simulate nsim
46 X1 = array(0, c(N, nsim, L))
47 X2 = array(0, c(1, nsim, L))
48 X3 = array(0, c(1, nsim, L))
49
50 #Choose the initial values according to the Gamma
      distribution and inserting into the signal process
51
52 #Inserting the initial values into the signal process
53 for(i in 1:L) {
54   for(j in 1:nsim) {
55     X1[1, j, i] = rgamma(1, 1)
56     X2[1, j, i] = rgamma(1, 1)

```

```

57 X3[1,j,i] = rgamma(1,1)
58 }
59 }
60
61 #-----
62 #Simulating the Brownian motion part
63 #-----
64 B_t = array(0,c(N,nsim,L))
65 for(l in 1:L) {
66   for(i in 1:nsim) {
67     dB_t = rnorm(T,0,1)*sqrt(dt) #Standard Brownian motion
        with mean 0 and variance 1
68     for(j in 1:(T-1)) {
69       B_t[j+1,i,l] = B_t[j,i,l] + dB_t[j]
70     }
71   }
72 }
73
74 #-----
75 #Simulate nsim paths of X1, L times
76 #-----
77
78 for(l in 1:L) {
79   for(i in 1:nsim) {
80     for(j in 1:(T-1)) {
81       X1[j+1,i,l] = X1[j,i,l] + X3[1,i,l]*(X2[1,i,l]-X1[1,i,l]
        ) *dt + B_t[j,i,l]*sqrt(dt)
82     }
83   }
84 }
85
86 #-----
87 #Lambda function
88 #-----
89 epsilon = 0.001
90 lambda <- function(epsilon,t,X_1,X_2,X_3) {
91   epsilon*(abs(X_1)+abs(X_2)+abs(X_3))
92 }
93
94 #-----
95 #Defining f(x)
96 #-----
97 f <- function(x) {
98   x
99 }
100
101 #-----
102 #Simulations of the density process Z_t, as given by
        algorithm 6
103 #-----
104

```

```

105 #Defining the jump given by I1 in (6.31)
106
107 epsilon_star = expectedClaimFrequency*2 #Chosen in
      order to give give a stable prediction
108 mean_delta_Y = mean(abs(delta_Y))
109 threshold = mean_delta_Y + epsilon_star #We have a jump
      if abs(delta_Y[i]) > threshold
110 jump_counter = 0 #Count the number of jumps occurring
111 I1temp = matrix(0,(length(t)),nsim)
112
113 #Count the number of jumps
114 for(j in 1:length(t)) {
115   if(abs(delta_Y[j]) > threshold) {
116     jump_counter = jump_counter+1
117   }
118 }
119
120 #Compute I1 given by (6.31)
121 I1 = array(0,c(1,nsim,L))
122 for(l in 1:L) {
123   for(i in 1:nsim) {
124     for(j in 1:(T-1)) {
125       if (abs(delta_Y[j]) > threshold) {
126         I1[1,i,l] = I1[1,i,l] + log(lambda(epsilon,l,X1[j+1,i,l],
          X2[1,i,l],X3[1,i,l]))
127       }
128     }
129   }
130 }
131
132 #Compute I2 given by (6.31) using the trapezoidal rule
133 I2 = array(0,c(1,nsim,L))
134 for(l in 1:L) {
135   for(i in 1:nsim) {
136     I2[1,i,l] = dt*((1-0.5*lambda(epsilon,l,X1[1,i,l],X2[1,
      i,l],X3[1,i,l]))-0.5*lambda(epsilon,l,X1[T,i,l],X2
        [1,i,l],X3[1,i,l])) + dt*sum(1-lambda(epsilon,l,X1[
          i,l],X2[1,i,l],X3[1,i,l]))
137   }
138 }
139
140 #Obtain Z_t
141 Z_t = array(0,c(1,nsim,L))
142 for(l in 1:L){
143   for(i in 1:nsim) {
144     Z_t[1,i,l] = exp(I1[1,i,l]+I2[1,i,l])
145   }
146 }
147
148 #

```

---

```

149 #Compute the optimal filters for X1, X2 and X3 by
      (6.19) and (6.20)

```



```

150 #


---


151
152 pi_2 = matrix(0,1,L)
153 pi_3 = matrix(0,1,L)
154
155 for(l in 1:L) {
156   Psi1_t = sum(Z_t[1,,l]) #Here f(x) = 1
157   Psi2_t = sum(Z_t[1,,l]*X2[1,,l])
158   Psi3_t = sum(Z_t[1,,l]*X3[1,,l])
159
160 #Compute X2_hat and X3_hat, as in step 4 in section
161     6.2.1
162   pi_2[1,l] = Psi2_t/Psi1_t
163   pi_3[1,l] = Psi3_t/Psi1_t
164 }
165 X2_hat = sum(pi_2)/length(pi_2)
166 X3_hat = sum(pi_3)/length(pi_3)
167
168 lines(pi_2[,], type = "l", xlab = "Simulation nr
169     1,2,...,15", ylab = "pi_2")
170 lines(pi_3[,], type = "l", xlab = "Simulation nr
171     1,2,...,15", ylab = "pi_3")
172
173 #-----
174 #Simulate nsim paths of X1
175 #-----
176 T_star = 200 # We simulate 200 days into the future
177 N_new = 200
178 dt = T_star/N_new
179 X1_new = matrix(0,N_new+1,nsim)
180
181 #Insert initial values
182 for(i in 1:nsim) {
183   X1_new[1,i] = rgamma(1,1)
184 }
185
186 #Simulating Brownian motion as before
187 B_t_new = matrix(0,N_new,nsim)
188 for(i in 1:nsim) {
189   dB_t_new = rnorm(N_new-1,0,1)*sqrt(dt) #Standard
190     Brownian motion with mean 0 and variance 1
191   for(j in 1:(N_new-1)) {
192     B_t_new[j+1,i] = B_t_new[j,i]+dB_t_new[j]
193   }
194 }
195 for(i in 1:nsim) {
196   for(j in 1:N_new) {
197     X1_new[j+1,i] = X1_new[j,i] + X3_hat*(X2_hat-X1_new[1,i]

```

```

    ])*dt+B_t_new[j , i]*sqrt(dt)
198 }
199 }
200 X1_0 = rgamma(1,1) #Initial value for the signal
      process X1
201
202 #-----
203 #Simulate the observation process Y for [0,T_star]
204 #-----
205
206 #Compute mu_star as in (6.28)
207 mu_star = matrix(0,T_star+1,nsim)
208 for(i in 1:nsim) {
209   for(j in 1:T_star) {
210     mu_star[j+1,i] = mu_star[j , i] + (dt/2)*(lambda(epsilon ,
      i ,X1_new[j , i] ,X2_hat ,X3_hat)+ lambda(epsilon , i ,X1_
      new[j+1,i] ,X2_hat ,X3_hat))
211   }
212 }
213 plot(mu_star[1 ,] , type ="l")
214 for(i in 1:nsim) {
215   lines(mu_star[i ,] , ylab = "mu_star" , xlab = "Time")
216 }
217 plot(mu_star[,1] , type="l" , xlab = "Time" , ylab = "mu_
      star")
218 for(i in 2:nsim) {
219   lines(mu_star[,i] , type ="l" , xlab = "Time" , ylab = "mu
      _star")
220 }
221 mean_mu_star = matrix(0,T_star , nsim)
222 for(i in 1:nsim) {
223   for(j in 1:T_star) {
224     mean_mu_star[j] = mean(mu_star[j ,])
225   }
226 }
227
228 lines(mean_mu_star[,1] , col = "red" , lwd=1)
229
230 Y_t = matrix(0,T_star , nsim)
231
232 plot(Y_t , type="l")
233
234 Y_t = matrix(0,T_star , nsim)
235 mean_Y_t = vector(mode = "numeric" , length = T_star)
236 mean_Y_t = matrix(0,T_star , nsim)
237
238 #Y_t[1 ,] = ClaimSize[1]
239 for(i in 1:T_star) {
240   for(j in 1:nsim) {
241     M = rpois(1 ,(mu_star[i ,j]-mu_star[i ,j]))
242     Y_t[1 ,j] = ClaimSize[1]
243     for(k in 1:M) {
244       Z = rgamma(1 , expectedClaimSize)

```

```

245 Y_t[i,j] = Y_t[i,j] + Z
246 }
247 }
248 }
249
250 #Plotting the average of Y_t
251 for(i in 1:nsim) {
252   for(j in 1:T_star) {
253     mean_Y_t[j] = mean(Y_t[j,])
254   }
255 }
256 plot(mean_Y_t)
257
258 plot(Y_t[,1], type = "l", xlab = "Time", ylab = "Y_t",
      ylim = c(0,14))
259
260
261 for(i in 2:nsim) {
262   lines(Y_t[,i], type = "l", xlab = "Time", ylab = "Y_t",
        ylim = c(0,14))
263 }
264
265 lines(mean_Y_t[,1], col = "red", lwd = 2)
266 ClaimSizeFirst200Days = c()
267 for(i in 1:200) {
268   ClaimSizeFirst200Days[i] = ClaimSize[i]
269 }
270 lines(ClaimSizeFirst200Days, type = "l", col = "blue",
      ylim = c(0,14))

```

## Bibliography

- [1] A. Agresti: *Foundations of Linear and Generalized Linear Models*. Wiley, 2015.
- [2] D. Applebaum: *Lévy Processes and Stochastic Calculus*. Cambridge University Press, 2nd edition, 2009.
- [3] V. E. Bening and V. Yu. Korolev: “On Approximations to Generalized Cox Processes”. In: *Probability and Mathematical Statistics* 18, Fasc 2 (1998), pp. 247–270.
- [4] Fred E. Benth: *Option Theory with Stochastic Analysis: An Introduction to Mathematical Finance*. Springer, 2004.
- [5] E. Bølviken: *Computation and Modelling in Insurance and Finance*. Cambridge University Press, 2014.
- [6] E. Bølviken, S. Duedahl, and F. Proske: *Modeling and estimation of stochastic transition rates based on generalized Cox processes with regime switching*. University of Oslo, 2015.
- [7] R. Cont and P. Tankov: *Financial Modelling with Jump Processes*. Chapman & Hall, 2008.
- [8] Jay L. Devore and Kenneth N. Berk: *Modern Mathematical Statistics with Applications*. Springer, 2007.
- [9] N. John and Neil A. Weiss: *A Course In Real Analysis, 2nd edition*. Academic press, 2012.
- [10] T. Mikosch: *Non-Life Insurance Mathematics*. Springer, 2nd edition, 2009.
- [11] B. Øksendal: *Stochastic Differential Equations: An Introduction with Applications*. Springer, 6th edition, 2004.
- [12] J. Xiong: *An Introduction to Stochastic Filtering Theory*. Oxford University Press, 2008.