Utrecht University

MASTER THESIS

MATHEMATICAL INSTITUUT

Properties and Applications of Hawkes Processes

Author
Peggy BERGMAN BSc
5834783

Supervisor
Dr. W.M. RUSZEL

Second reader
Dr. K. DAJANI

2nd July 2023

Abstract

Point processes in probability theory are special stochastic processes which model the occurrence of events (points) in some space and parameterized by their frequency. Hawkes processes are general point processes where the frequency can be self-exciting or mutually exciting. They occur in various areas such as geological sciences, epidemiology and financial mathematics. The focus of this thesis lies in discussing different types of Hawkes processes and their properties as well as presenting financial applications. The (linear) Hawkes process or self-exciting Hawkes process is a process with a single counting process such that that the occurrence of an event increases the probability of the occurrence of another event. In the case of the (linear) Hawkes process, we prove the Law of Large Numbers and the Central Limit Theorem. Moreover, we study the Hawkes likelihood function and the Hawkes log-likelihood function. Besides a self-exciting Hawkes process there exists a mutually exciting Hawkes process, which is a process with multiple counting processes that are depended on each other. Meaning that the occurrence of an event in one of these counting processes also lead to an increased probability of an event occurring in the other counting processes. We study the Hawkes likelihood function and the Hawkes log-likelihood function for the mutually exciting Hawkes process. The marked Hawkes process is a (linear) Hawkes process with added random variables called the random marks. We prove the Hawkes likelihood function and the Hawkes log-likelihood function as well as the Central Limit Theorem. Furthermore, we derive the dynamics for the Hawkes jump-diffusion model given by three stochastic differential equations and prove the Law of Large Numbers and Central Limit Theorem for this particular model.

Acknowledgements

I would like to thank my supervisor Dr. Wioletta Ruszel for suggesting Hawkes processes as a subject for my thesis. Further, I would like to give my gratitude for her guidance during the process. She was always willing to lend me a hand and encouraged me during this process. Secondly, I would like to thank my second reader Dr. Karma Dajani for her helpful feedback and suggestions.

Contents

Introduction

Point processes in probability theory are special stochastic processes which model the occurrence of events (points) in some space and parameterized by their frequency. For example, the Poisson point process with constant rate μ models random events, such as the occurrence of earthquakes. The Poisson point process can defined by considering the number of points (events) in a specified time interval. The higher the constant rate, also called the intensity, the higher the probability of points/events occurring in the specified time interval on average.

In 1971 Alan Hawkes introduced a new kind of mathematical process that is able to model self-exciting and even mutually exciting phenomena. This new process is called the Hawkes process, which is a generalization of the point process with a more complex intensity called the conditional intensity function. Therefore, the Hawkes process is actually a self-exciting or mutually exciting point process. A self-exciting point process models the sequence of arrivals over some period of time in such a way that each arrival or event increases the likelihood of another event happening. The Hawkes process has numerous applications, for instance in geological sciences, epidemiology and mathematical finance [13]. For example, Ait-Sahalia et al [1] used a mutually exciting jump-diffusion model, the Hawkes jump-diffusion model, to model the financial markets in different regions all around the globe. An event occurring on the financial market in one region impacts the financial market in the other regions as well for a certain time period in the sense that the probability of an event occurring in these other regions increases.

The so-called power law turns out to be a popular choice for the excitation function, when modeling geological phenomena such as earthquakes. The Omori formula is a geological model that uses this so-called power law to predict the rate of aftershocks in a certain region after an earthquake has occurred [19].

Unwin and et el. [31] proposed a new method using Hawkes processes for modeling infectious diseases outbreaks. In their article they specifically model the malaria transmission in a near-elimination setting using a generalization of the Hawkes process.

One of the goals of the thesis is that it is a self-contained overview of the Hawkes process and the properties of the Hawkes process. Furthermore, a reader without a sufficient background in stochastic processes, measure theory and financial mathematics will still be able to read the thesis. However, the main goal is defining, explaining and proving certain properties of the Hawkes process and the generalizations of the (linear) Hawkes process, such as the mutually exciting Hawkes process and the marked Hawkes process, as well as introducing financial applications of the Hawkes process. This means that we introduce the definition and the conditional intensity function of the (linear) Hawkes process along with the definitions for the mutually exciting Hawkes process and the marked Hawkes process. In case of the (linear) Hawkes process, we may think of it as a counting process or point process that models a sequence of events or arrivals over a non-negative time period. Each event of arrival during this time period will excite the process in such a way that the probability of the next event happening increases and then decreases according to the excitation function. We explain the different choices for the excitation function, through various numerical examples and compare those examples with the numerical example of the homogeneous Poisson process. We also provide examples for the mutually exciting Hawkes process and the marked Hawkes process in order to make the material more tangible, which is on its own quite dense. Besides explaining the (linear) Hawkes process and its generalizations, we also provide various proofs for the Law of Large Numbers, Central Limit Theorem and the Hawkes likelihood function for the different cases. Simply stated, the Law of Large Numbers provides a statement of the concentration of the process around the mean and the Central Limit Theorem provides information about the asymptotic distribution of a properly rescaled Hawkes process, which is Gaussian. The Hawkes likelihood function gives information on the joint probability of the realizations of that certain Hawkes process. It basically states the joint probability of the realizations as a function of the parameters of the Hawkes process and is then used to estimate the unknown parameter(s). We prove the expression of the Hawkes likelihood function and the Hawkes log-likelihood function in the case of the (linear) Hawkes process, the mutually exciting Hawkes process and the marked Hawkes process. Furthermore, we give some numerical examples of estimating certain parameters. Regarding applications of the Hawkes process in finance, we describe and explain the Black-Scholes model and the Merton jump-diffusion model and provide some numerical examples of those models. After that we extend the Merton model by replacing the Poisson jump process by a mutually exciting Hawkes process, this model is known as the Hawkes jump-diffusion model. Moreover, we prove the Law of Large Numbers and the Central Limit Theorem in the univariate case of the Hawkes jump-diffusion model. Furthermore, we end the section on Hawkes jump-diffusion with some simple examples where we give explicit expressions in order to make it more comprehensible.

The thesis is structured as follows:

Chapter 1 consists of the preliminaries necessary to understand the Hawkes process and the applications of the Hawkes process in finance sufficiently. The chapter contains measure theoretical notions as σ-algebra, filtration and measurability. Also, the modes of convergence, the standard Law of Large Numbers and the standard Central Limit Theorem will be explained. In the second section of this chapter, the counting process and the point process will be discussed even as the conditional intensity function. The third and last section is about the Poisson process.

Chapter 2 contains the necessary notions of stochastic calculus such as Itô-Doeblin formula, Brownian motion, martingales and quadratic variation.

Chapter 3 discusses the (linear) Hawkes process. The definition is given even as the corresponding conditional intensity function. The most common choices for the excitation function are explained and we will derive the Hawkes process likelihood function. Lastly, the Law of Large Numbers and the Central Limit Theorem for the linear Hawkes process are defined and the proof is given for both theorems.

Chapter 4 discusses the mutually exciting Hawkes process. The definition of the mutually exciting Hawkes process is given and furthermore the multivariate Hawkes process likelihood will be stated and proven.

Chapter 5 introduces the marked Hawkes processes. Besides the definition and the corresponding conditional intensity function, the marked Hazard function is also defined. In the second section the marked Hawkes process likelihood will be stated and also a proof will be given. In the third section the Central Limit Theorem for the marked case will be proven.

Chapter 6 discusses the Black-Scholes model. In this chapter the Black-Scholes stochastic differential equation will be derived given certain assumptions on the financial market.

Chapter 7 covers the Merton jump-diffusion model. The Merton jump-diffusion model is an extension of the Black-Scholes model in the sense that the model incorporates a jump component. The stochastic differential equation of the option price, in case of this model, will be derived in this chapter.

Chapter 8 introduces the Hawkes jump-diffusion model. The Hawkes jump-diffusion model extends the Merton jump-diffusion model by replacing the Poisson jump process by a mutually exciting Hawkes process. In this chapter the Hawkes jump-diffusion model will be defined and we will prove the Law of Large Numbers and Central Limit Theorem in case of the Hawkes jump-diffusion model.

Chapter 9 provides a numerical study of the Hawkes process. Different realizations of the Hawkes process will be simulated using Python. Furthermore, the parameters will also be estimated.

Part I

Background

Chapter 1

Preliminary

1.1 Measure Theory

This section will consist of the necessary definitions and theorems for understanding the material on Hawkes processes. The reader who is already familiar with the basic notions of measure theory may safely skip this part. For this section, we made use of the lecture notes of A.W. van der Vaart [32], the book of L. Wasserman [35] and the lecture notes of P.J.C. Spreij [30].

Definition 1.1.1 (σ-algebra)**.** Let Ω be some set. A collection $\mathcal{F} \subseteq 2^{\Omega}$ is called a σ-algebra if

(i) $\Omega \in \mathcal{F}$

(ii) If $E \in \mathcal{F}$, then $E^c \in \mathcal{F}$

(iii) If $E_1, E_2, ... \in \mathcal{F}$, then $\bigcup_{n=1}^{\infty} E_n \in \mathcal{F}$

Definition 1.1.2 (Measurable)**.** Let $(\Omega, \mathcal{F})$ be a measurable space and let $f : \Omega \to \mathbb{R}$. We say that f is measurable if $\forall B \in \mathcal{B}(\mathbb{R})$, with $\mathcal{B}$ the Borel σ-algebra, it holds that

$$f^{-1}[B] := \{\omega \in \Omega : f(\omega) \in B\} \in \Sigma.$$

Definition 1.1.3 ((Continuous) Filtration)**.** Let $(X_n)_{n \in \mathbb{N}_0}$ be a stochastic process on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and let $\mathbb{F} = (\mathcal{F}_n)_{n \in \mathbb{N}_0}$ be a family of σ-algebras such that $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq ...$ and $\mathcal{F}_n := \sigma(X_0, ..., X_n)$. Then we call $\mathbb{F}$ a filtration.
In case of continuous time t, we define the filtration as an increasing sequence of σ-algebras $(\mathcal{F}_t)_{t \geq 0}$ such that $\mathcal{F}_t := \sigma(X_s | s \leq t)$.

We will state the three modes of convergence.

Definition 1.1.4 (Convergence in distribution)**.** Let $(X_n)_{n \in \mathbb{N}}$ be a sequence of random variables defined on some probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then $(X_n)_{n \in \mathbb{N}}$ converges in distribution to the random variable X defined on the same probability space if

$$\mathbb{P}(X_n \leq x) \to \mathbb{P}(X \leq x) \quad \text{as} \quad n \to \infty$$

for every x at which the distribution function $x \to \mathbb{P}(X \leq x)$ is continuous.
Notation: $X_n \rightsquigarrow X$.

Let $d(x,y)$ be a distance on $\mathbb{R}^k$. For instance, the Euclidean distance defined as

$$d(x,y) := ||x-y|| := \left(\sum_{i=1}^{k}(x_i-y_i)^2\right)^{\frac{1}{2}}.$$

Definition 1.1.5 (Convergence in probability)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$, then $(X_n)_{n\in\mathbb{N}}$ converges in probability to the random variable X defined on the same probability space if for all $\epsilon > 0$

$$\mathbb{P}(d(X_n,X) > \epsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

Notation: $X_n \xrightarrow{\mathbb{P}} X$.

Definition 1.1.6 (Almost sure convergence)**.** Let $(X_n)_{n\in\mathbb{N}}$ be a sequence of random variables defined on some probability space $(\Omega,\mathcal{F},\mathbb{P})$, then $(X_n)_{n\in\mathbb{N}}$ converges almost surely to the random variable X defined on the same probability space if

$$\mathbb{P}(\lim_{n\to\infty} d(X_n,X) = 0) = 1.$$

Notation: $X_n \xrightarrow{a.s.} X$.

Lemma 1.1.1 (Joint convergence in probability)**.** Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, X and Y be random variables. If $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$, then $(X_n,Y_n) \xrightarrow{\mathbb{P}} (X,Y)$.

Proof. Let $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$. We will prove $\forall\epsilon > 0$ that

$$\mathbb{P}(d((X_n,Y_n),(X,Y)) > \epsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

Let $\epsilon > 0$ be given, then

$$\begin{aligned}\mathbb{P}(d((X_n,Y_n),(X,Y)) > \epsilon) &\le \mathbb{P}(d(X_n,X) > \frac{\epsilon}{2}) + \mathbb{P}(d(Y_n,Y) > \frac{\epsilon}{2}) \quad \text{by triangle inequality}\\ &\to 0 \quad \text{as} \quad n \to \infty,\end{aligned}$$

since $\forall\epsilon > 0$ it holds that $\mathbb{P}(d(X_n,X) > \epsilon) \to 0$ and $\mathbb{P}(d(Y_n,Y) > \epsilon) \to 0$ as $n \to \infty$.

■

Lemma 1.1.2 (Joint convergence in distribution)**.** Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, X and Y be random variables and let X_n and Y_n be independent for all n. If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ and X and Y are independent, then $(X_n,Y_n) \rightsquigarrow (X,Y)$.

Proof. Let $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$ and suppose that X_n and Y_n are independent. Then

$$\mathbb{P}(X_n \le x, Y_n \le y) = \mathbb{P}(X_n \le x)\mathbb{P}(Y_n \le y) \rightsquigarrow \mathbb{P}(X \le x)\mathbb{P}(Y \le y) = \mathbb{P}(X \le x, Y \le y).$$

■

Lemma 1.1.3 (Continuous mapping theorem)**.** Let $(X_n)_{n\in\mathbb{N}}$ be random variable.

1. If $X_n \xrightarrow{\mathbb{P}} X$ and g is a continuous function, then $g(X_n) \xrightarrow{\mathbb{P}} g(X)$
2. If $X_n \rightsquigarrow X$ and g is a continuous function, then $g(X_n) \rightsquigarrow g(X)$

Proof. We only prove the first assertion. Let $X_n \xrightarrow{\mathbb{P}} X$. Let $\epsilon > 0$ be given. Suppose that there exists a K large enough such that $\mathbb{P}(|X| > K) \leq \frac{\delta}{2}$ for some $\delta > 0$. The function g is uniformly continuous on the interval $[-K, K]$ and therefore there exists a $\delta' > 0$ such that for $|x| \leq K$ and $|x-y| < \delta'$, we have that $|g(x) - g(y)| < \epsilon$. Hence,

$$\begin{aligned}\mathbb{P}(|g(X_n) - g(X)| > \epsilon) &\leq \mathbb{P}(|X| > K) + \mathbb{P}(|X_n - X| > \delta') \\ &\leq \frac{\delta}{2} + \frac{\delta}{2} = \delta,\end{aligned}$$

for n large enough. ([25], proof of Theorem 1.24, p. 55)

■

Lemma 1.1.4 (Slutsky)**.** Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, X and Y be random variables. If $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$, then $X_n + Y_n \xrightarrow{\mathbb{P}} X + Y$.

Proof. Let $X_n \xrightarrow{\mathbb{P}} X$ and $Y_n \xrightarrow{\mathbb{P}} Y$. We will prove $\forall\epsilon > 0$ that

$$\mathbb{P}(d(X_n + Y_n - (X+Y)) > \epsilon) \to 0 \quad \text{as} \quad n \to \infty.$$

Let $\epsilon > 0$ be given, then

$$\begin{aligned}\mathbb{P}(d(X_n + Y_n - (X+Y)) > \epsilon) &= \mathbb{P}(d((X_n - X) + (Y_n - Y)) > \epsilon) \\ &\leq \mathbb{P}(d(X_n - X) > \frac{\epsilon}{2}) + \mathbb{P}(d(Y_n - Y) > \frac{\epsilon}{2}) \quad \text{by triangle inequality} \\ &\to 0 \quad \text{as} \quad n \to \infty,\end{aligned}$$

since $\forall\epsilon > 0$ it holds that $\mathbb{P}(d(X_n - X) > \epsilon) \to 0$ and $\mathbb{P}(d(Y_n - Y) > \epsilon) \to 0$ as $n \to \infty$.

■

Lemma 1.1.5 (Slutsky)**.** Let $(X_n)_{n\in\mathbb{N}}$, $(Y_n)_{n\in\mathbb{N}}$, X and Y be independent random variables.

1. If $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$, then $X_n + Y_n \rightsquigarrow X + Y$.
2. If $X_n \rightsquigarrow X$ and $Y_n \xrightarrow{\mathbb{P}} c$ with c a constant, then $Y_n X_n \rightsquigarrow cX$.

Proof. Let $X_n \rightsquigarrow X$ and $Y_n \rightsquigarrow Y$. By Lemma 1.1.2, we have that $(X_n, Y_n) \rightsquigarrow (X, Y)$. By the Continuous Mapping Theorem 1.1.3 with $g(x, y) = x + y$, we obtain the the first assertion. Now, let $X_n \rightsquigarrow X$ and $Y_n \xrightarrow{\mathbb{P}} c$ with c a constant. By Lemma 1.1.2, we have that $(X_n, Y_n) \rightsquigarrow (X, c)$. By the Continuous Mapping Theorem 1.1.3 with $g(x, y) = yx$, we obtain the the second assertion.

■

The next theorems will be stated without a proof, since one can view them as classical results.

Theorem 1.1.6 (Law of Large Numbers (LLN))**.** Let $\frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean of the first n observations of a sequence of i.i.d. random variables with finite expectation. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{\mathbb{P}} \mathbb{E}[X_1].$$

Theorem 1.1.7 (Strong Law of Large Numbers)**.** Let $\frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean of the first n observations of a sequence of i.i.d. random variables with finite expectation. Then

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{a.s.} \mathbb{E}[X_1].$$

Theorem 1.1.8 (Central Limit Theorem (CLT))**.** Let $\frac{1}{n}\sum_{i=1}^{n} X_i$ be the sample mean of the first n
observations of a sequence of i.i.d. random variables. If $\mathbb{E}[X_1^2] < \infty$, then

$$\sqrt{n}(\frac{1}{n}\sum_{i=1}^{n} X_i - \mathbb{E}[X_1]) \rightsquigarrow \mathcal{N}(0, Var(X_1)).$$

Lastly, we will mention a couple important inequalities.

Definition 1.1.7 (Markov's Inequality)**.** Let X be a random variable and suppose that $\mathbb{E}[X]$ exists. Then, for all constant $c > 0$ and a non-decreasing, non-negative function g, it holds that

$$\mathbb{P}(X \geq c) \leq \frac{\mathbb{E}[g(X)]}{g(c)}. \tag{1.1}$$

Definition 1.1.8 (Chebyshev´s Inequality)**.** Let X be a random variable and suppose that $Var(X)$ is finite and non-zero. Then, for all $\epsilon > 0$, it holds that

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq \epsilon) \leq \frac{Var(X)}{\epsilon^2}. \tag{1.2}$$

Remark. $|\cdot|$ stands for the absolute value.

1.2 Counting Process and Point Process

This section is largely based on the article of P.J. Laub [15] and on Chapter 2 of *The Elements of Hawkes Processes* [16].

We may think of a counting process as a cumulative count of the number of arrivals into a system up to the current time. The formal definition is given in Definition 1.2.1. Another way of characterizing the number of arrivals is provided by the point process in Definition 1.2.3. Here, we consider the sequence of random arrival times $\mathbf{T} = \{T_1, T_2, ...\}$ at which the counting process has jumped.

Definition 1.2.1 (Counting process)**.** A counting process is a stochastic process $(N_t)_{t \geq 0}$, which satisfies the following conditions:

- The process takes values in $\mathbb{N}_0$ such that $N_0 = 0$
- The process is almost surely (a.s.) finite
- The process is a right-continuous step function with increments of size $+1$

In the next section, we will provide an example of a counting process, namely the Poisson process.

Definition 1.2.2 (History)**.** We define the history of the arrivals up to time u as $(\mathcal{H}_u)_{u \geq 0}$. Note that $\mathcal{H}_u$ is a filtration, Definition 1.1.3.

Definition 1.2.3 (Simple point process)**.** A sequence of random arrival times $\mathbf{T} = \{T_1, T_2, ...\}$ is a simple point process if the following conditions are satisfied:

- $\mathbf{T}$ takes values in $[0, \infty)$
- The simple point process is an almost surely increasing sequence, $\mathbb{P}(0 \leq T_1 \leq T_2 \leq ...) = 1$
- The number of points in a bounded region is almost surely (a.s.) finite.

We may characterize a certain point process by specifying the conditional distribution function. The conditional cumulative distribution function (c.d.f.) of the next arrival time T_{n+1} given the history of arrivals up to time u, $\mathcal{H}_u$, is defined as

$$F(t|\mathcal{H}_u) := \int_u^t \mathbb{P}(T_{n+1} \in [s, s+ds]|\mathcal{H}_u)ds = \int_u^t f(s|\mathcal{H}_u)ds \tag{1.3}$$

with f the conditional probability density function (p.d.f.). The joint p.d.f. for a realization $\{t_1, t_2, ..., t_n\}$ is then given by

$$f(t_1, t_2, ..., t_n) = \prod_{i=1}^{n} f(t_i|\mathcal{H}_{t_{i-1}}). \tag{1.4}$$

By convention, we write $F(t|\mathcal{H}_u)$ as $F^*(t)$ and $f(t|\mathcal{H}_u)$ as $f^*(t)$ for $u < t$. With the help of the conditional distribution function of a point process, we may define the so-called Hazard function.

Definition 1.2.4 (Hazard function)**.** Consider a point process N_t with associated history $\mathcal{H}_t$. Furthermore, consider the conditional c.d.f. and p.d.f. defined in (1.3). Then, the Hazard function, which is random, is defined as

$$\lambda^*(t) = \frac{f^*(t)}{1 - F^*(t)}.$$

Due to the complexity of working with the conditional c.d.f. and p.d.f., we rather prefer the so-called conditional intensity function $\lambda(\cdot)$, which is the expected rate of arrivals conditioned on the history. Please note that both definitions are valid.

Definition 1.2.5 (Conditional intensity function)**.** Consider a counting process N_t with associated history $\mathcal{H}_t$. $\lambda(t)$ is called the conditional intensity function of the counting process if $\lambda(t)$ is non-negative and $\mathcal{H}_t$-measurable (Definition 1.1.2) with $\lambda(t)$ defined as

$$\lambda(t) = \lim_{k \downarrow 0} \frac{\mathbb{E}[N_{t+k} - N_t | \mathcal{H}_t]}{k}.$$

By integrating the conditional intensity function we obtain the so-called compensator.

Definition 1.2.6 (Compensator)**.** The compensator for a counting process is defined as the non-decreasing deterministic function

$$\Lambda(t) = \int_0^t \lambda(s) ds.$$

1.3 Poisson Process

This section is largely based on Section 2.2 of *The Elements of Hawkes Processes* [16]. We will give the definitions of a Poisson distribution, a homogeneous Poisson process and an inhomogeneous Poisson process.

Definition 1.3.1 (Poisson distribution)**.** A random variable X is Poisson distributed with rate $\mu > 0$, $X \sim Pois(\mu)$, if it has a probability mass function given by

$$\mathbb{P}(X = k) = \frac{\mu^k}{k!} \cdot e^{-\mu}, \quad \text{with} \quad k = 0, 1, 2, ...$$

Definition 1.3.2 (Homogeneous Poisson process)**.** A counting process $(N_t)_{t \geq 0}$ is a homogeneous Poisson process with rate $\mu > 0$ if the following conditions are satisfied:

- For any interval I, we have that $N_I \sim Pois(\mu |I|)$ with $|I|$ the length of the interval and N_I the number of arrivals in interval I. In other words, if the length of the interval I equals t, we have that $N_t \sim Pois(\mu t)$.
- For any n disjoint intervals $I_1, I_2, ..., I_n$, the random variables $N_{I_1}, N_{I_2}, ..., N_{I_n}$ are independent.

Then it follows that the probability of N_t being equal to n is given by

$$\mathbb{P}(N_t = n) = \frac{(\mu t)^n}{n!} \cdot e^{-\mu t}, \quad \text{with} \quad n = 0, 1, 2, ...$$

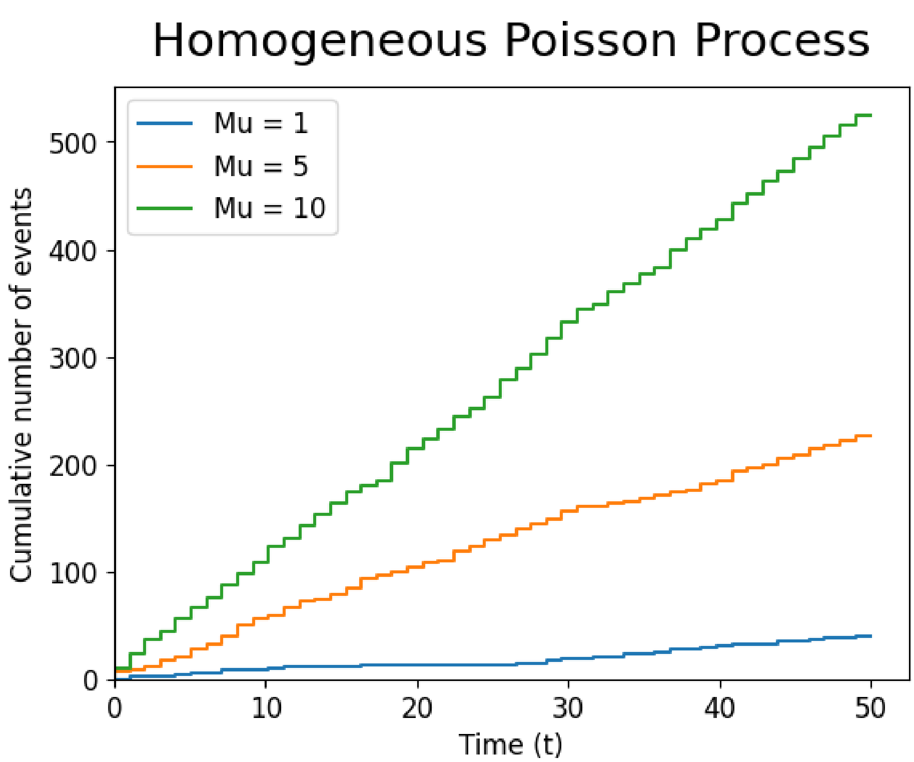

Figure 1.1: An example of Homogeneous Poisson process with $\mu = \{1, 5, 10\}$ and time up to 50. (see Appendix 11.1)

In figure 1.1, we see that for small μ there are rather long flat parts, which means that there are no arrivals of new events. Contrary we see that for larger μ there are more frequent steps, which shows that events are arriving quicker. Overall the graphs are roughly linear, which is to be expected due to the homogeneous nature.

Definition 1.3.3 (Inhomogeneous Poisson process)**.** A counting process $(N_t)_{t\geq 0}$ is a inhomogeneous Poisson process with rate $\mu(t) > 0$ if the following conditions are satisfied:

- For any interval $I = (a,b]$, we have that $N_I \sim Pois\left(\int_a^b \mu(s)ds\right)$ with $a,b \in \mathbb{R}$ and $a \leq b$.
- For any n disjoint intervals $I_1, I_2, ..., I_n$, the random variables $N_{I_1}, N_{I_2}, ..., N_{I_n}$ are independent.

The difference between a homogeneous Poisson process and that of an inhomogeneous Poisson process is that the rate μ is time-dependent in the inhomogeneous case and a constant in the homogeneous case.

Chapter 2

Stochastic Calculus for Finance

In this chapter, the necessary definitions for understanding stochastic calculus needed for the Black-Scholes model (Chapter 6), the Merton jump-diffusion model (Chapter 7) and the Hawkes jump-diffusion model (Chapter 8) will be discussed. This implies that we will give the definitions of Brownian motion, martingales, quadratic variation and the Itô-Doeblin formulas. Furthermore, we will give some examples corresponding to those definitions. The reader who is already familiar with these notions may safely skip this chapter. For this chapter, we made use of the book of S.E. Shreve [29] and the lecture notes of F. Boshuizen et al. [6].

Definition 2.0.1 (Brownian motion)**.** Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. A stochastic process $(W_t)_{t\geq 0}$ is a called a Brownian motion if the following holds:

(i) The process starts at 0: $W_0 = 0$

(ii) Normal increments: $W_t - W_s \sim \mathcal{N}(0, t-s)$ for all $0 \leq s < t$

(iii) Independent increments: $W_t - W_s \perp\!\!\!\perp (W_u : 0 \leq u < s)$ for all $0 \leq s < t$.

(iv) Any sample path $t \mapsto W_t(\omega)$ is continuous

Remark. We use the symbol $\perp\!\!\!\perp$ to indicate independence.

Example 2.0.1 (Brownian motion)**.** We show that $(-W_t)_{t\geq 0}$ is a Brownian motion. First, we have that $(W_t)_{t\geq 0}$ is a Brownian motion, so $W_0 = 0$ and thus $-W_0 = 0$. Secondly,

$$\mathbb{E}[-W_t - (-W_s)] = \mathbb{E}[-(W_t - W_s)] = -\mathbb{E}[W_t - W_s] = -0 = 0.$$

Furthermore,

$$Var(-(W_t - W_s)) = Var(W_t - W_s) = t - s,$$

hence $(-W_t)_{t\geq 0}$ has normal increments. Due to the fact that $(W_t)_{t\geq 0}$ is a Brownian motion, we know that $t \mapsto W_t(\omega)$ is continuous and so is $t \mapsto -W_t(\omega)$ is continuous. Hence, $(-W_t)_{t\geq 0}$ is a Brownian motion.

♦

Lemma 2.0.2 (Law of Large Numbers for Brownian motion)**.** Let $(W_t)_{t\geq 0}$ be a Brownian motion, then we have that $\lim_{t\to\infty} \frac{W_t}{t} = 0$ almost surely.

Proof. The proof can be found in ([18], Corollary 1.11, p. 14).

■

Definition 2.0.2 (Adapted)**.** A stochastic process $(X_t)_{t\geq 0}$ is adapted to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if $\sigma(X_t) \subset \mathcal{F}_t$.

Recall that a filtration $(\mathcal{F}_t)_{t\geq 0}$ in continuous time is an increasing sequence of σ-algebras, so $\mathcal{F}_s \subset \mathcal{F}_t$ for all $s < t$.

Definition 2.0.3 (Martingale)**.** A stochastic process $(X_t)_{t\geq 0}$ is a martingale with respect to the filtration $(\mathcal{F}_t)_{t\geq 0}$ if the following holds:

1. $(X_t)_{t\geq 0}$ is adapted to the filtration
2. $\mathbb{E}[|X_t|] < \infty$ for all t
3. Martingale property: $\mathbb{E}[X_t|\mathcal{F}_s] = X_s$ for $s < t$

Example 2.0.3 (Doob martingale)**.** Let Y be a random variable such that $\mathbb{E}[|Y|] < \infty$ and let $(\mathcal{F}_t)_{t\geq 0}$ be an arbitrary filtration, then we have that $X_t = \mathbb{E}[Y|\mathcal{F}_t]$ is a martingale with respect to the filtration. By using the tower property, we obtain the statement.

$$\mathbb{E}[X_t|\mathcal{F}_s] = \mathbb{E}[\mathbb{E}[Y|\mathcal{F}_t]|\mathcal{F}_s] = \mathbb{E}[Y|\mathcal{F}_s] = X_s \quad \text{for all} \quad s < t$$

♦

Example 2.0.4. We show that the Brownian motion $(W_t)_{t\geq 0}$ with respect to the natural filtration fulfills the martingale property, since the first two properties of a martingale are trivially satisfied. This means that we only need to show that $\mathbb{E}[W_t|\mathcal{F}_s] = W_s$ for $s < t$. So

$$\begin{aligned} \mathbb{E}[W_t|\mathcal{F}_s] &= \mathbb{E}[W_t - W_s + W_s|\mathcal{F}_s] \\ &= \mathbb{E}[W_t - W_s|\mathcal{F}_s] + \mathbb{E}[W_s|\mathcal{F}_s] \\ &= 0 + W_s = W_s \end{aligned}$$

Hence, $(W_t)_{t\geq 0}$ satisfies the martingale property.
Note: The natural filtration of a stochastic process $(X_t)_{t\geq 0}$ is $\mathcal{F}_t = \sigma(X_s|s \leq t)$.

♦

Definition 2.0.4 (Quadratic variation)**.** Let $\Pi = \{t_0, t_1, ..., t_n\}$ be the partition of the interval $[0,T]$. Then the covariation of the stochastic processes $X = (X_t)_{t\geq 0}$ and $Y = (Y_t)_{t\geq 0}$ is defined as

$$[X,Y]_t := \lim_{||\Pi||\to 0} \sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})(Y_{t_k} - Y_{t_{k-1}}),$$

where we call the norm of the partition the mesh. Note that the convergence is in $L_2(\mathbb{R})$.

Remark. The quadratic variation for a single stochastic process follows by setting $Y = X$. Then we obtain the statement

$$[X,X]_t = [X]_t := \lim_{||\Pi||\to 0} \sum_{k=1}^{n} (X_{t_k} - X_{t_{k-1}})^2.$$

Definition 2.0.5 (Itô-Doeblin formula)**.** Let $f(t,x,y)$ be a function for which the partial derivatives $f_t(t,x,y)$, $f_x(t,x,y)$, $f_y(t,x,y)$, $f_{xx}(t,x,y)$, $f_{yy}(t,x,y)$ and $f_{xy}(t,x,y)$ are defined and continuous, i.e. $f \in C^2(\mathbb{R}\times\mathbb{R})$. Furthermore, let $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ be a stochastic processes. The three-dimensional Itô-Doeblin formula in differential form is then given by

$$\begin{aligned} df(t,X_t,Y_t) = f_t(t,X_t,Y_t)dt + f_x(t,X_t,Y_t)dX_t + f_y(t,X_t,Y_t)dY_t + \frac{1}{2}f_{xx}(t,X_t,Y_t)d[X]_t \\ + \frac{1}{2}f_{yy}(t,X_t,Y_t)d[Y]_t + f_{xy}(t,X_t,Y_t)d[X,Y]_t. \end{aligned}$$

Remark. The Itô-Doeblin formula for one-dimension and two-dimension follows from the three-dimensional Itô-Doeblin formula stated above.

Example 2.0.5 (Itô-Doeblin formula stock price process)**.** Let $(S_t)_{t\geq 0}$ be the stock price process assumed to follow a geometric Brownian motion. This means that S_t is defined as

$$S_t = S_0 e^{\mu t + \sigma W_t},$$

with S_0 the initial stock price, $(W_t)_{t\geq 0}$ a Brownian motion and $\mu \in \mathbb{R}$ and $\sigma > 0$ the drift and volatility, respectively. To obtain the stochastic differential equation corresponding to $(S_t)_{t\geq 0}$, we will apply the Itô-Doeblin formula to $f(x) = S_0 e^x$ with $X_t = \mu t + \sigma W_t$. Then

$$\begin{aligned} X_t &= \mu t + \sigma W_t \\ dX_t &= \mu dt + \sigma dW_t \\ d[X]_t &= \sigma^2 dt. \end{aligned}$$

Therefore, applying Itô-Doeblin yields

$$\begin{aligned} dS_t = df(X_t) &= f_x(X_t)dX_t + \frac{1}{2}f_{xx}(X_t)d[X]_t \\ &= S_0 e^{X_t}dX_t + \frac{1}{2}S_0 e^{X_t}d[X]_t \\ &= S_t dX_t + \frac{1}{2}S_t d[X]_t \\ &= S_t(\mu dt + \sigma dW_t) + \frac{1}{2}S_t\sigma^2 dt \\ &= \left(\mu + \frac{1}{2}\sigma^2\right)S_t dt + \sigma S_t dW_t. \end{aligned}$$

Note that $f_x(x) = f_{xx}(x) = S_0 e^x$.

Hence, the stock price process $(S_t)_{t\geq 0}$ satisfies the stochastic differential equation

$$dS_t = \left(\mu + \frac{1}{2}\sigma^2\right)S_t dt + \sigma S_t dW_t.$$

♦

Example 2.0.6 (Itô-Doeblin formula discounted stock price process)**.** Let $(\tilde{S}_t)_{t\geq 0}$ be the discounted stock price process defined as

$$\tilde{S}_t = e^{-rt}S_t = e^{-rt}S_0 e^{\mu t + \sigma W_t} = S_0 e^{(\mu - r)t + \sigma W_t}, \quad t \geq 0,$$

with S_t as in Example 2.0.5 and r is the interest rate. Hence, the drift term becomes $\mu - r \in \mathbb{R}$. We will apply the Itô-Doeblin formula to $f(x) = S_0 e^x$ with $X_t = (\mu - r)t + \sigma W_t$. Note that

$f_x(x) = f_{xx}(x) = S_0 e^x$. Then, we have

$$
\begin{aligned}
X_t &= (\mu - r)t + \sigma W_t \\
dX_t &= (\mu - r)dt + \sigma dW_t \\
d[X]_t &= \sigma^2 dt.
\end{aligned}
$$

Hence,

$$
\begin{aligned}
d\tilde{S}_t = df(X_t) &= f_x(X_t)dX_t + \frac{1}{2} f_{xx}(X_t)d[X]_t \\
&= S_0 e^{X_t} dX_t + \frac{1}{2} S_0 e^{X_t} d[X]_t \\
&= \tilde{S}_t dX_t + \frac{1}{2}\tilde{S}_t d[X]_t \\
&= \tilde{S}_t \left((\mu - r)dt + \sigma dW_t\right) + \frac{1}{2}\tilde{S}_t \sigma^2 dt \\
&= \left((\mu - r) + \frac{1}{2}\sigma^2\right)\tilde{S}_t dt + \sigma \tilde{S}_t dW_t.
\end{aligned}
$$

Thus, the discounted stock price process $(\tilde{S}_t)_{t \geq 0}$ satisfies the SDE

$$
d\tilde{S}_t = \left(\mu - r + \frac{1}{2}\sigma^2\right)\tilde{S}_t dt + \sigma \tilde{S}_t dW_t.
$$

♦

Part II

Hawkes Processes

Chapter 3

Hawkes Process

3.1 Linear Hawkes Process

The Hawkes process is a self-exciting point process and is named after Alan Hawkes, who introduced in his article *Spectra of some self-exciting and mutually-exciting point processes* [11] the notion of self-exciting and mutually exciting point processes in 1971. Since then, the Hawkes process is applied in numerous different fields such as seismology, ecology and finance. In this section, which is largely based on the article of P.J. Laub [15] and on Chapter 3 of *The Elements of Hawkes Processes* [16], we will introduce the linear Hawkes process. We added the derivation of the stochastic differential equation of the exponentially decaying intensity function and figures 3.1 and 3.4.

Definition 3.1.1 (Little "o" notation ([24], p. 314, Definition 5.2))**.** A function $f(\cdot)$ is said to be $o(k)$ if it holds that

$$\lim_{k\to 0}\frac{f(k)}{k}=0.$$

Definition 3.1.2 (Linear Hawkes process)**.** A counting process $(N_t)_{t\geq 0}$ with associated history $(\mathcal{H}_t)_{t\geq 0}$ that satisfies

$$\mathbb{P}(N_{t+k}-N_t=m \mid \mathcal{H}_t)=\begin{cases}\lambda(t)k+o(k), & m=1\\ o(k), & m>1\\ 1-\lambda(t)k+o(k), & m=0\end{cases}$$

is called a Hawkes process. The corresponding conditional intensity function (Definition 1.2.5) is defined as

$$\lambda(t)=\nu+\int_0^t h(t-s)dN_s.$$

We call $\nu>0$ the background intensity and the function $h:(0,\infty)\to[0,\infty)$ is referred to as the excitation function.

The conditional intensity function $\lambda(\cdot)$ may also be written as

$$\lambda(t)=\nu+\sum_{t_i<t}h(t-t_i),$$

where $\{t_1,t_2,...,t_n\}$ is an increasing sequence of random times up to time t at which events or arrivals occur. Therefore, we may think of a Hawkes process as a counting process or point process that models a sequence of events or arrivals over a non-negative time period. Each event or arrival during this time-period will excite the process in such a way that the probability of the next event happening increases and then decreases according to the excitation function. There are some well studied choices for the excitation function $h(\cdot)$, namely the exponentially decaying intensity and the power law function.

Definition 3.1.3 (Exponentially decaying intensity)**.** Let the excitation function $h(\cdot)$ be defined as

$$h(t) = \alpha e^{-\beta t},$$

with $\alpha, \beta > 0$ some constants. The exponentially decaying intensity function $\lambda(\cdot)$ is then given by

$$\lambda(t) = \nu + \int_0^t \alpha e^{-\beta(t-s)} dN_s = \nu + \sum_{t_i < t} \alpha e^{-\beta(t-t_i)}.$$

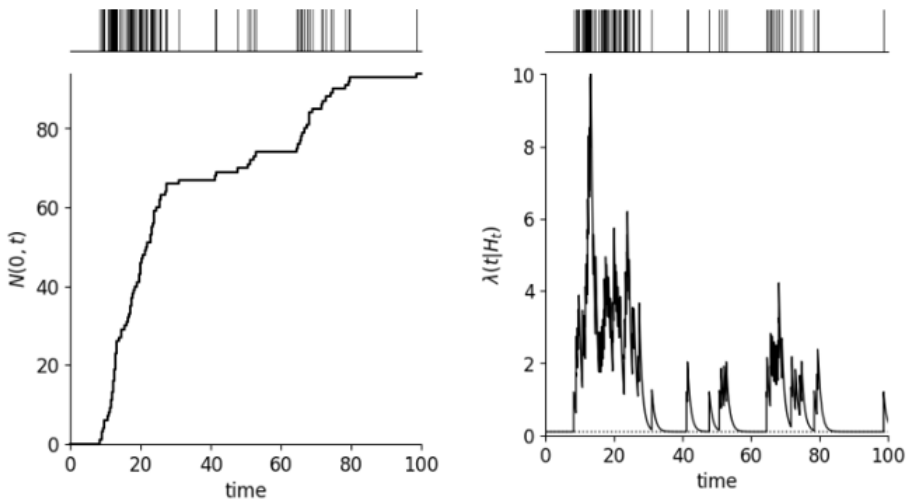

Figure 3.1: A realization of a Hawkes process N_t with exponentially decaying intensity function $\lambda(t)$ with parameters $\nu = 0.1$, $\alpha = 1.0$ and $\beta = 1.1$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

If we compare the realization of the Hawkes process with exponentially decaying intensity function given by

$$\lambda(t) = 0.1 + \sum_{t_i < t} 1.0 e^{-1.1(t-t_i)},$$

on the left in Figure 3.1, to the homogeneous Poisson process in Figure 1.1 then we can note a couple of differences. The graph of the homogeneous Poisson process is rather linear and especially for larger rates the intervals between two events are quite small. Whereas for the exponential case (Figure 3.1), we see that a lot of events occur in the beginning followed by some larger periods of time where no events happen at all. Overall the Hawkes process with exponentially decaying intensity does not really have a linear trend as opposed to the homogeneous Poisson process. We may also compare the Hawkes process with exponentially decaying intensity function with different parameters to each other. Let us compare the left side of Figure 3.1 to the left side of Figure 3.2. So, if parameter α decreases and β remains the same, then the time between two events occurring becomes significantly longer. This can been seen by the rather long flat parts in the graph of the left side of Figure 3.2. Now, if we compare the left side of Figure 3.1 to the left side of Figure 3.3, then we see that lots of activity will happen followed by periods of no activity at all. Thus, we may say that β suppresses activity.

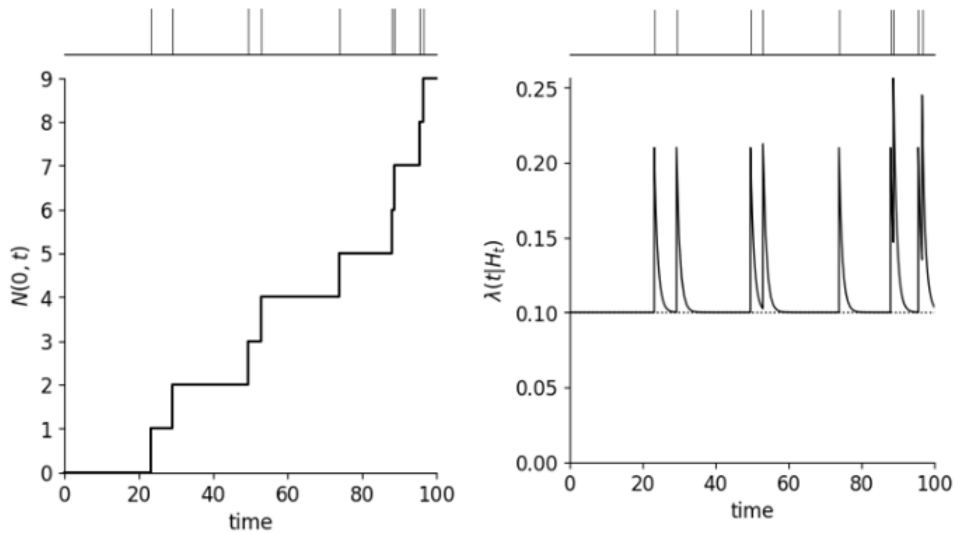

Figure 3.2: A realization of a Hawkes process N_t with exponentially decaying intensity function $\lambda(t)$ with parameters $\nu = 0.1$, $\alpha = 0.1$ and $\beta = 1.1$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

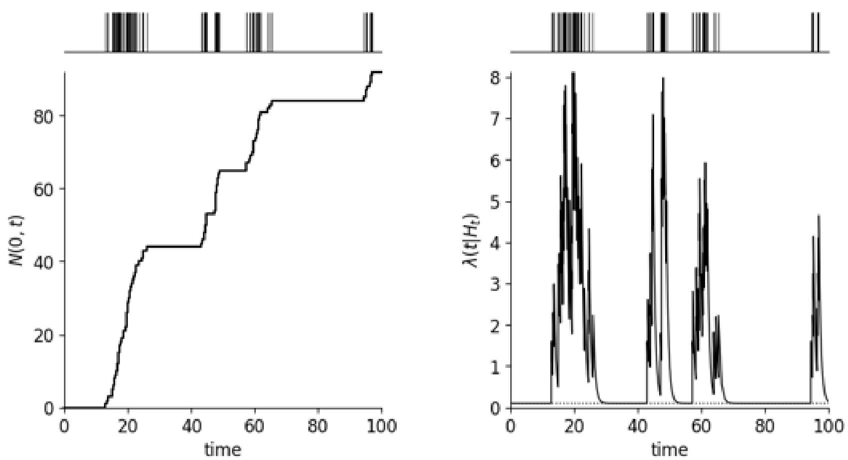

Figure 3.3: A realization of a Hawkes process N_t with exponentially decaying intensity function $\lambda(t)$ with parameters $\nu = 0.1$, $\alpha = 1.0$ and $\beta = 1.5$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

Definition 3.1.4 (Power law function)**.** Let the excitation function $h(\cdot)$ be defined as

$$h(t) = \frac{k}{(c+t)^p}$$

with $k, c, p > 0$ some constants. This specific excitation function $h(\cdot)$ is referred to as the power law function. The conditional intensity function $\lambda(\cdot)$ is then given by

$$\lambda(t) = \nu + \int_0^t \frac{k}{(c+(t-s))^p} dN_s = \nu + \sum_{t_i<t} \frac{k}{(c+(t-t_i))^p}.$$

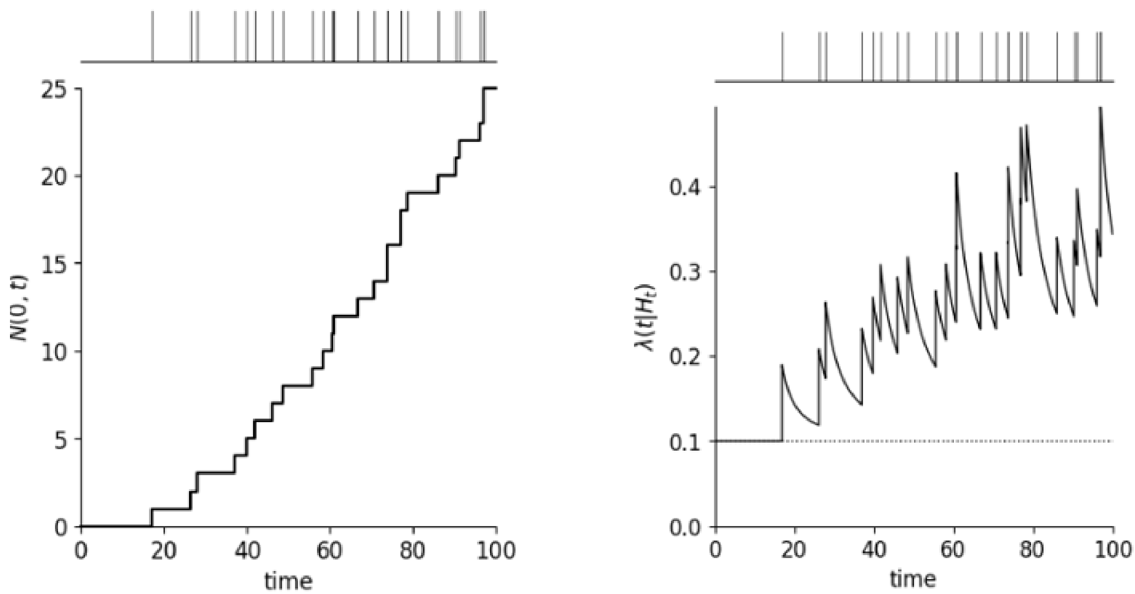

Figure 3.4: A realization of a Hawkes process N_t with corresponding conditional intensity function $\lambda(t)$ with power law function as the excitation function. The parameters are given by $\nu = 0.1$, $k = 1.0$, $c = 5.0$ and $p = 1.5$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

Let us now compare the realization of the Hawkes process with the power law as the excitation function, so we have the conditional intensity function given by

$$\lambda(t) = 0.1 + \sum_{t_i<t} \frac{1.0}{(5.0+(t-t_i))^{1.5}},$$

on the left in Figure 3.4, to the homogeneous Poisson process in Figure 1.1 then we can see some similarity. The graph of the homogeneous Poisson process is rather linear and especially for larger rates the intervals between two events are quite small. The graph shown on the left side in Figure 3.4 displays roughly similar behavior. However, if we change some parameters this behavior changes. We start with comparing the left side of Figure 3.4 to the left side of Figure 3.5. So, parameter k decreases and the other parameters remain the same. Then, the behavior remains roughly the same, but their are significantly less events occurring. Now, we increase parameter c and the others remain the same (Figure 3.6), then the behavior of events arriving becomes more erratic and the trend does not look linear anymore as opposed to Figure 3.4. Lastly, we compare the left side of Figure 3.4 to the left side of Figure 3.7. Here, the parameter p is increased and the others remain the same. Then we have more or less similar looking behavior, however the graph is less linear.

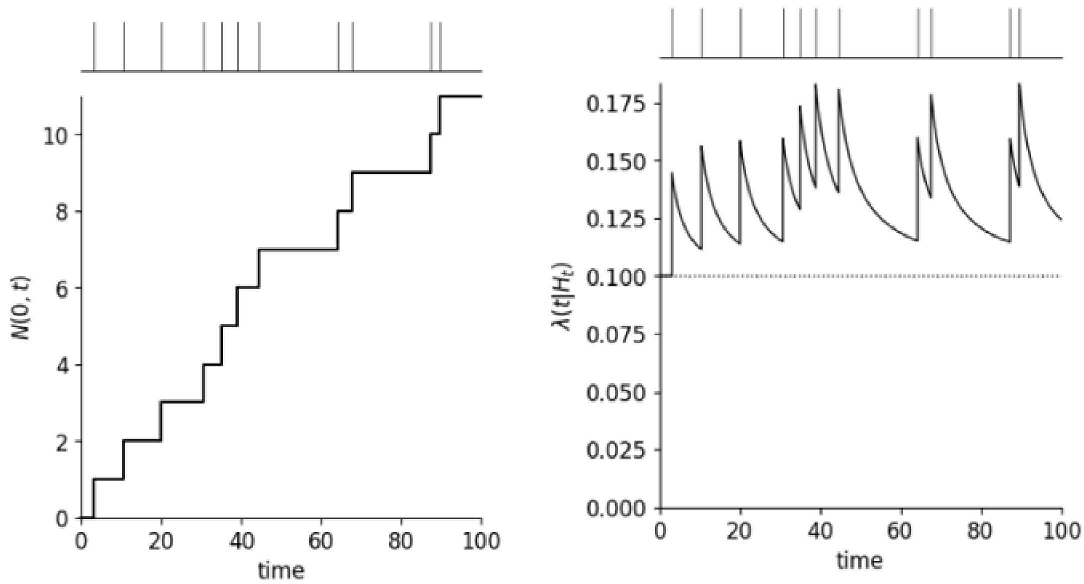

Figure 3.5: A realization of a Hawkes process N_t with corresponding conditional intensity function $\lambda(t)$ with power law function as the excitation function. The parameters are given by $\nu = 0.1$, $k = 0.5$, $c = 5.0$ and $p = 1.5$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

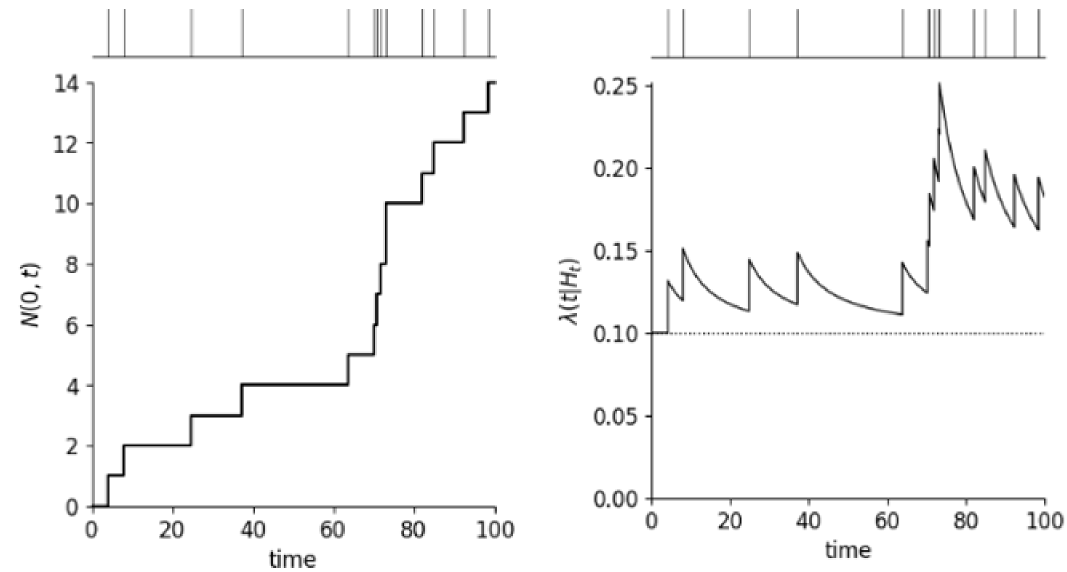

Figure 3.6: A realization of a Hawkes process N_t with corresponding conditional intensity function $\lambda(t)$ with power law function as the excitation function. The parameters are given by $\nu = 0.1$, $k = 1.0$, $c = 10.0$ and $p = 1.5$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

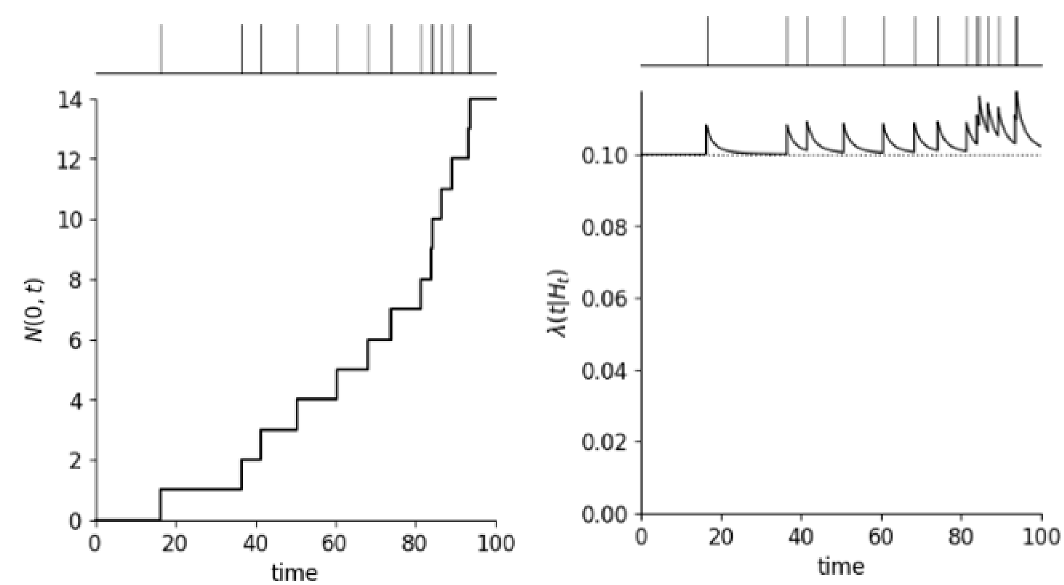

Figure 3.7: A realization of a Hawkes process N_t with corresponding conditional intensity function $\lambda(t)$ with power law function as the excitation function. The parameters are given by $\nu = 0.1$, $k = 1.0$, $c = 5.0$ and $p = 3.0$.
On the left: plot of time against the number of arrivals.
On the right: plot of time against the conditional intensity function

The linear Hawkes process is applicable in the seismology, in particular the power law. One is able to predict the rate of the aftershocks caused by an earthquake using this particular Hawkes process. As we will see, this model can become more advanced once one adds random marks in order to account for the magnitude of an earthquake. An example of this will be given in Chapter 5. The linear Hawkes process can also be applied in the epidemiology. In the article of M.-A. Rizoiu et al. [22] a link between the linear Hawkes process and an epidemic model with finite population size is proposed. Their conditional intensity function is given by

$$\tilde{\lambda}(t) = \left(1 - \frac{N_t}{N}\right)\left(\nu + \sum_{t_i<t} h(t-t_i)\right),$$

with $h(\cdot)$ the excitation function and $\nu > 0$ the background intensity just as in the linear Hawkes process and furthermore N_t a counting process and N the finite population size. Note that if $N \to \infty$, so an infinite population size, then we obtain the regular conditional intensity function for the linear Hawkes process.

Lemma 3.1.1. The exponentially decaying intensity function $\lambda(\cdot)$ satisfies the following stochastic differential equation

$$d\lambda(t) = \beta(\nu - \lambda(t))dt + \alpha dN_t, \quad t \geq 0. \tag{3.1}$$

Proof. The exponentially decaying intensity function $\lambda(\cdot)$ can be written as

$$\lambda(t) = \nu + \int_0^t \alpha e^{-\beta(t-s)} dN_s = \nu + \alpha e^{-\beta t}\int_0^t e^{\beta s} dN_s.$$

Taking the derivative on both sides yields

$$\begin{aligned}
d\lambda(t) &= \left(-\alpha\beta e^{-\beta t}\int_0^t e^{\beta s} dN_s\right) dt + \alpha e^{-\beta t}\cdot e^{\beta t} dN_t \\
&= \left(-\beta\int_0^t \alpha e^{-\beta(t-s)} dN_s\right) dt + \alpha dN_t \\
&= \left(-\beta\left(\int_0^t \alpha e^{-\beta(t-s)} dN_s + \nu - \nu\right)\right) dt + \alpha dN_t \\
&= \left(-\beta\left(\lambda(t) - \nu\right)\right) dt + \alpha dN_t \\
&= \beta\left(\nu - \lambda(t)\right) dt + \alpha dN_t.
\end{aligned}$$

Hence, the stochastic differential equation $d\lambda(t) = \beta(\nu - \lambda(t))dt + \alpha dN_t$ corresponds to exponentially decaying intensity function.

■

Lemma 3.1.1 will be used later to derive the univariate model for the Hawkes jump-diffusion model in Chapter 8.

Lastly, we will remark that the definition of the linear Hawkes process described above can be generalized in the form of a nonlinear Hawkes process.

Definition 3.1.5 (Non-linear Hawkes process)**.** A counting process $(N_t)_{t\geq 0}$ with conditional intensity function given by

$$\lambda(t) = \Psi\left(\int_0^t h(t-s)dN_s\right),$$

where $\Psi : \mathbb{R} \to [0,\infty)$ and excitation function $h : (0,\infty) \to [0,\infty)$, is called a nonlinear Hawkes process.

Then, choosing Ψ such that $\Psi(x) = \nu + x$ will reduce the counting process $(N_t)_{t\geq 0}$ to the linear Hawkes process.

3.2 Parameter Estimation

This section is largely based on the article of P.J. Laub [15] and on Chapter 5.1 of *The Elements of Hawkes Processes* [16]. In the proof of Theorem 3.2.1, we made some minor changes in the notation and added some steps in the calculations. Furthermore, we added the log-likelihood function.

Theorem 3.2.1 (Hawkes process likelihood)**.** Let $(N_t)_{t\geq 1}$ be a simple point process on $[0,T]$ for some finite positive T and let $\{t_1, t_2, ..., t_n\}$ denote the arrival times of the events on $[0,T]$. The likelihood function L of $(N_t)_{t\geq 1}$ is then of the form

$$L(t_1,...,t_n|\theta) = \left(\prod_{i=1}^{n} \lambda^*(t_i)\right) \cdot \exp\left\{-\int_0^T \lambda^*(u)du\right\}$$

and the log-likelihood l has the form

$$l(t_1,...,t_n|\theta) = \sum_{i=1}^{n} \lambda^*(t_i) - \int_0^T \lambda^*(u)du.$$

Remark. θ contains the background intensity ν and the parameters necessary for the excitation function h. In the case of the exponentially decaying intensity, we have that $\theta := (\nu, \alpha, \beta)$.

Proof. The likelihood function in general is given by

$$L(t_1,...,t_n|\theta) = \prod_{i=1}^{n} f^*(t_i)$$

with $f^*(t_i)$ the conditional probability density function as defined in (1.4). We want to express the likelihood function in terms of the hazard function λ^*, Definition 1.2.4.

$$\lambda^*(t) = \frac{f^*(t)}{1-F^*(t)} = \frac{\frac{d}{dt}F^*(t)}{1-F^*(t)} = \frac{d}{dt}\left(-\log(1-F^*(t))\right).$$

Integrating both sides over the interval (t_{i-1}, t_i) gives

$$\int_{t_{i-1}}^{t_i} \lambda^*(u)du = \int_{t_{i-1}}^{t_i} \frac{d}{du}\left(-\log(1-F^*(u))\right)du = -\log(1-F^*(t_i)) + \log(1-F^*(t_{i-1})).$$

The second term on the right-hand side will be equal to zero due to the assumption that $(N_t)_{t\geq 1}$ is a simple point process and therefore it cannot happen that multiple events occur at the same time, hence $F^*(t_{i-1}) = 0$. Thus,

$$\int_{t_{i-1}}^{t_i} \lambda^*(u)du = -\log(1-F^*(t_i)).$$

So,

$$1 - F^*(t_i) = \exp\left\{-\int_{t_{i-1}}^{t_i} \lambda^*(u)du\right\}$$
$$F^*(t_i) = 1 - \exp\left\{-\int_{t_{i-1}}^{t_i} \lambda^*(u)du\right\}$$

Now, we can define the conditional probability density function in terms of the hazard function, since

$$f^*(t_i) = \frac{d}{dt}F^*(t_i) = \lambda^*(t_i) \cdot \exp\left\{-\int_{t_{i-1}}^{t_i} \lambda^*(u)du\right\} \quad \text{by the funtamental theorem of calculus.}$$

Hence, the likelihood function becomes

$$\begin{aligned} L(t_1,...,t_n|\theta) &= \prod_{i=1}^{n} f^*(t_i) = \prod_{i=1}^{n}\left(\lambda^*(t_i)\cdot \exp\left\{-\int_{t_{i-1}}^{t_i} \lambda^*(u)du\right\}\right) \\ &= \left(\prod_{i=1}^{n}\lambda^*(t_i)\right)\cdot \exp\left\{-\int_{0}^{T} \lambda^*(u)du\right\}. \end{aligned}$$

The log-likelihood function is then given by

$$\begin{aligned} l(t_1,...,t_n|\theta) &= \log\left(\left(\prod_{i=1}^{n}\lambda^*(t_i)\right)\cdot \exp\left\{-\int_{0}^{T} \lambda^*(u)du\right\}\right) \\ &= \sum_{i=1}^{n}\lambda^*(t_i) - \int_{0}^{T}\lambda^*(u)du. \end{aligned}$$

■

In Chapter 9, we will give some numerical examples.

3.3 Limit Theorems

In this section, which is largely based on the article of Y. Seol [27], we will prove the Law of Large Numbers and the Central Limit Theorem. Note that our description is more general, this is due to the fact that we made less assumptions. Specifically, we do not assume that $t\nu$ is a natural number. Furthermore, we will give a more detailed version of the proofs.

Let us first recall that the conditional intensity function for the linear Hawkes process is given by

$$\lambda(t) = \nu + \int_0^t h(t-s)dN_s,$$

where $\nu > 0$ the background intensity and $h(\cdot)$ the excitation function. In order to prove the Law of Large Numbers and the Central Limit Theorem it is convenient to characterize the linear Hawkes process by using the immigration-birth representation. The immigration-birth representation states that an immigrant arrives according to a standard homogeneous Poisson process with rate $\nu > 0$, Definition 1.3.2, and after that each immigrant may or may not have children according to a Galton-Watson tree (see [12] for details). We define the number of children of an immigrant as η and we assume that η is Poisson distributed with rate $||h||_{L^1} = \int_0^\infty h(t)dt$. Furthermore, the total number of immigrants and their descendants up to time t is denoted by N_t. In this section, we will assume that the immigrants arrive uniformly over time with rate $\nu > 0$ rather than Poisson with rate $\nu > 0$. This means that the immigrants arrive at times

$$\frac{i}{\nu}, \quad i = 1,2,3,....$$

We may write the number of immigrants and their descendants up to time t as a sum of the $i-$th immigrant that arrives at time $\frac{i}{\nu}$ and its descendants. So

$$N_t = \sum_{i=1}^{\lfloor \nu t \rfloor} X_i.$$

Here, the X_i are independent and denote the number of the $i-$th immigrant and its descendants. The $i-$th immigrant and its descendants arrive during the time interval $[\frac{i}{\nu}, \frac{\nu t}{\nu}]$ for $i = 1,2,3,...$, where each immigrant generates descendants according to the immigration-birth representation as described earlier. So, the probability that the number of immigrants and their descendants up to time t that arrives uniformly over time with rate $\nu > 0$ equals $k \in \mathbb{N}$ is given by

$$\mathbb{P}(N_t = k) = \frac{1}{\frac{\nu t}{\nu} - \frac{k}{\nu}} = \frac{\nu}{\nu t - k}.$$

Before we prove the LLN and the CLT, we will define the necessary assumptions.

Assumptions:

(i) The conditional intensity function is linear and increasing, so $\lambda(x) = \nu + x$ for some strictly positive ν.

(ii) It holds that the L^1-norm of the excitation function h is less than one, hence $||h||_{L^1} < 1$ with $||h||_{L^1} = \int_0^\infty h(t)dt < \infty$.

The first assumption assures that the Hawkes process has a nice immigration-birth representation, whereas the second assumption makes sure that the limits are well-defined and makes sure that the Hawkes process does not explode.

Theorem 3.3.1 (Law of Large Numbers for linear Hawkes process)**.** Assume that the provided assumptions are satisfied. Then we have

$$\frac{N_t}{t} \xrightarrow{\mathbb{P}} \frac{\nu}{1 - ||h||_{L^1}} \quad \text{as} \quad t \to \infty.$$

Proof. Let X_i be the number of the $i-$th immigrant and its descendants that arrive during the time interval $[\frac{i}{\nu}, \frac{\nu t}{\nu}]$. We have that the X_i are i.i.d.. Let Y_i be the number of the $i-$th immigrant and its descendants that arrive during the time interval $[\frac{i}{\nu}, \infty)$. Also, we have Y_i are i.i.d.. The moment generating function for Y_i is given by

$$M_Y(\theta) := \mathbb{E}[e^{\theta Y_i}].$$

One can use the large deviation principle, for instance ([33], p. 8, Theorem 2.1) can be modified to obtain the following statement. The moment generating function $M_Y(\theta) := \mathbb{E}[e^{\theta Y_i}]$ satisfies the equation

$$M_Y(\theta) = e^{\theta + (M_Y(\theta) - 1)||h||_{L^1}}$$

for $\theta \le \theta_c = ||h||_{L^1} - 1 - \log(||h||_{L^1})$. So,

$$\begin{aligned} M_Y'(\theta) &= \frac{d}{d\theta} e^{\theta + (M_Y(\theta)-1)||h||_{L^1}} \\ &= e^{\theta + (M_Y(\theta)-1)||h||_{L^1}} + e^{\theta + (M_Y(\theta)-1)||h||_{L^1}} \cdot M_Y'(\theta)||h||_{L^1}. \end{aligned}$$

Hence,

$$\begin{aligned} \mathbb{E}[Y_i] = M_Y'(0) &= e^0 + e^0 M_Y'(\theta)||h||_{L^1} \quad (\text{since} \quad M_Y(0) = 1) \\ &= 1 + \mathbb{E}[Y_i]||h||_{L^1} \end{aligned}$$

Thus,

$$\begin{aligned} \mathbb{E}[Y_i] &= 1 + \mathbb{E}[Y_i]||h||_{L^1} \\ (1 - ||h||_{L^1})\mathbb{E}[Y_i] &= 1 \\ \mathbb{E}[Y_i] &= \frac{1}{1 - ||h||_{L^1}}. \end{aligned}$$

We have that

$$\begin{aligned} \mathbb{E}[\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i] &= \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} \mathbb{E}[Y_i] = \frac{1}{t} \cdot \nu t \cdot \mathbb{E}[Y_1] \quad \text{since } Y_i \text{ are i.i.d.} \\ &= \nu \mathbb{E}[Y_1] = \frac{\nu}{1 - ||h||_{L^1}}. \end{aligned}$$

By the Strong Law of Large Numbers, we obtain

$$\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i \xrightarrow{a.s.} \nu \mathbb{E}[Y_1] = \frac{\nu}{1 - ||h||_{L^1}} \quad \text{as} \quad t \to \infty.$$

By definition, we have that $N_t = \sum_{i=1}^{\lfloor \nu t \rfloor} X_i$. So, we have

$$N_t = \sum_{i=1}^{\lfloor \nu t \rfloor} X_i \le \sum_{i=1}^{\lfloor \nu t \rfloor} Y_i \quad \text{a.s..}$$

Therefore,

$$\frac{N_t}{t} = \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} X_i \le \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i \xrightarrow{a.s.} \frac{\nu}{1 - ||h||_{L^1}} \quad \text{as} \quad t \to \infty.$$

Furthermore, recall that X_i is the number of descendants of the $i-$th immigrant, including the immigrant i, that arrives during the time interval $[\frac{i}{\nu}, \frac{\nu t}{\nu}]$. This means that we can write the expected value as

$$\mathbb{E}[X_i] = f\left(\frac{\nu t - i}{\nu}\right),$$

where $f(\cdot)$ satisfies the renewal equation for Hawkes processes (see [36]) with

$$f(t) = 1 + \int_0^t h(t-s) f(s) ds.$$

This can be derived from the fact that for all $\theta \leq \theta_c = ||h||_{L^1} - 1 - \log(||h||_{L^1})$, it holds that

$$M_S(t;\theta) \coloneqq \mathbb{E}[e^{\theta S_t}].$$

Note that S_t is the number of descendants of an immigrant, including the immigrant itself, that arrives on time interval $[0,t]$ and satisfies

$$M_S(t;\theta) = e^{\theta + \int_0^t (M_S(t;\theta)-1)h(t-s)ds}.$$

Then,

$$f(t) \coloneqq \mathbb{E}[S_t] = M_S'(t;0) = 1 + \int_0^t h(t-s)f(s)ds.$$

Moreover, we defined Y_i to be the number of descendants of the $i-$th immigrant, including the immigrant i, that arrives during the time interval $[\frac{i}{\nu}, \infty)$. So, we can write the expected value of Y_i in terms of the renewal equation as

$$\mathbb{E}[Y_i] = \lim_{t\to\infty} f(t) = \boldsymbol{f},$$

which is just a value. So, we obtain

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \sum_{i=1}^{\lfloor \nu t \rfloor} X_i\right] &= \sum_{i=1}^{\lfloor \nu t \rfloor} (\mathbb{E}[Y_i] - \mathbb{E}[X_i]) \quad \text{due to linearity of the expectation} \\ &= \sum_{i=1}^{\lfloor \nu t \rfloor} \left(\boldsymbol{f} - f\left(\frac{\nu t - i}{\nu}\right)\right).\end{aligned}$$

Hence,

$$\lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \sum_{i=1}^{\lfloor \nu t \rfloor} X_i\right] = 0.$$

Now, let $\epsilon > 0$ be given, then

$$\begin{aligned}\mathbb{P}\left(|\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} X_i - \frac{\nu}{1-||h||_{L^1}}| > \epsilon\right) &= \mathbb{P}\left(|\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} X_i - \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i + \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \frac{\nu}{1-||h||_{L^1}}| > \epsilon\right) \\ &\leq \mathbb{P}\left(|\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} X_i - \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i| > \frac{\epsilon}{2}\right) + \mathbb{P}\left(|\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \frac{\nu}{1-||h||_{L^1}}| > \frac{\epsilon}{2}\right) \quad \text{triangle inequality} \\ &= \mathbb{P}\left(\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} X_i > \frac{\epsilon}{2}\right) + \mathbb{P}\left(|\frac{1}{t}\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \frac{\nu}{1-||h||_{L^1}}| > \frac{\epsilon}{2}\right) \\ &\to 0 \quad \text{as} \quad t\to\infty,\end{aligned}$$

since the first component will go to zero due to the fact that

$$\lim_{t\to\infty} \frac{1}{t}\mathbb{E}\left[\sum_{i=1}^{\lfloor \nu t \rfloor} Y_i - \sum_{i=1}^{\lfloor \nu t \rfloor} X_i\right] = 0$$

and by the Weak Law of Large Numbers we have that the second component will go to zero.

Then, for all $\epsilon > 0$, it holds that

$$\begin{aligned}\mathbb{P}\left(|\frac{1}{t}\sum_{i=1}^{\lfloor \nu t\rfloor} X_i - \frac{\nu}{1-||h||_{L^1}}| > \epsilon\right) &= \mathbb{P}\left(|\sum_{i=1}^{\lfloor \nu t\rfloor} X_i - \frac{\nu t}{1-||h||_{L^1}}| > \epsilon t\right)\\ &\leq \frac{\sum_{i=1}^{\lfloor \nu t\rfloor} Var(X_i)}{\epsilon^2 t^2} \quad \text{by Chebyshev's inequality (1.1.8)}\\ &\leq \frac{\nu t Var(X_1)}{\epsilon^2 t^2} \quad X_i \text{ are i.i.d.}\\ &= \frac{\nu Var(X_1)}{\epsilon^2 t} \to 0 \quad \text{as} \quad t\to\infty.\end{aligned}$$

Hence,

$$\frac{N_t}{t} = \frac{1}{t}\sum_{i=1}^{\lfloor \nu t\rfloor} X_i \xrightarrow{a.s.} \frac{\nu}{1-||h||_{L^1}} \quad \text{as} \quad t\to\infty.$$

■

Theorem 3.3.2 (Central Limit Theorem for linear Hawkes process)**.** Assume that the provided assumptions are satisfied and

$$\lim_{t\to\infty}\frac{1}{\sqrt{t}}\int_0^t\int_u^\infty h(s)dsdu = 0.$$

Then we have

$$\sqrt{t}\left(\frac{N_t}{t} - \frac{\nu}{1-||h||_{L^1}}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu||h||_{L^1}}{(1-||h||_{L^1})^3}\right) \quad \text{as} \quad t\to\infty.$$

Proof. Let us recall that the moment generating function $M_Y(\theta) := \mathbb{E}[e^{\theta Y_i}]$ satisfies the equation

$$M_Y(\theta) = e^{\theta + (M_Y(\theta)-1)||h||_{L^1}}$$

for $\theta \leq \theta_c = ||h||_{L^1} - 1 - \log(||h||_{L^1})$. Furthermore, recall that

$$\begin{aligned}M'_Y(\theta) &= \frac{d}{d\theta}e^{\theta + (M_Y(\theta)-1)||h||_{L^1}}\\ &= e^{\theta + (M_Y(\theta)-1)||h||_{L^1}} + e^{\theta + (M_Y(\theta)-1)||h||_{L^1}} \cdot M'_Y(\theta)||h||_{L^1}\\ &= (1 + M'_Y(\theta)||h||_{L^1})M_Y(\theta).\end{aligned}$$

So the second derivative will be given by

$$\begin{aligned}M''_Y(\theta) &= \frac{d^2}{d\theta^2}e^{\theta + (M_Y(\theta)-1)||h||_{L^1}}\\ &= \frac{d}{d\theta}(1 + M'_Y(\theta)||h||_{L^1})M_Y(\theta)\\ &= M'_Y(\theta) + M'_Y(\theta)M'_Y(\theta)||h||_{L^1} + M_Y(\theta)M''_Y(\theta)||h||_{L^1}\\ &= (1 + M'_Y(\theta)||h||_{L^1})M'_Y(\theta) + M_Y(\theta)M''_Y(\theta)||h||_{L^1}.\end{aligned}$$

Thus,

$$\begin{aligned}\mathbb{E}[Y_i^2] = M''_Y(0) &= (1 + M'_Y(0)||h||_{L^1})M'_Y(0 + M_Y(0M''_Y(0)||h||_{L^1}\\ &= (1 + \mathbb{E}[Y_i]||h||_{L^1})\mathbb{E}[Y_i] + \mathbb{E}[Y_i^2]||h||_{L^1}\end{aligned}$$

Note that $M_Y(0)=1$ and $M'_Y(0)=\mathbb{E}[Y_i]$. It follows that

$$\begin{aligned}\mathbb{E}[Y_i^2] &= \frac{(1+\mathbb{E}[Y_i]||h||_{L^1})\mathbb{E}[Y_i]}{1-||h||_{L^1}} \\ &= \frac{\frac{1}{1-||h||_{L^1}} + \frac{||h||_{L^1}}{1-||h||_{L^1}}\cdot\frac{1}{1-||h||_{L^1}}}{1-||h||_{L^1}} \\ &= \frac{1}{(1-||h||_{L^1})^3}\end{aligned}$$

Hence,

$$Var(Y_i)=\mathbb{E}[Y_i^2]-(\mathbb{E}[Y_i])^2=\frac{1}{(1-||h||_{L^1})^3}-\left(\frac{1}{1-||h||_{L^1}}\right)^2=\frac{||h||_{L^1}}{(1-||h||_{L^1})^3}.$$

By the Central Limit Theorem, we have that

$$\sqrt{t}\left(\frac{1}{t}\sum_{i=1}^{\lfloor \nu t\rfloor}Y_i-\frac{\nu}{1-||h||_{L^1}}\right)\rightsquigarrow\mathcal{N}\left(0,\frac{\nu||h||_{L^1}}{(1-||h||_{L^1})^3}\right)\quad\text{as}\quad t\to\infty.$$

The statement follows if it holds that

$$\frac{1}{\sqrt{t}}\mathbb{E}\left[\sum_{i=1}^{\lfloor \nu t\rfloor}Y_i-\sum_{i=1}^{\lfloor \nu t\rfloor}X_i\right]\xrightarrow{\mathbb{P}}0\quad\text{as}\quad t\to\infty.$$

We have that

$$\begin{aligned}\frac{1}{\sqrt{t}}\mathbb{E}\left[\sum_{i=1}^{\lfloor \nu t\rfloor}Y_i-\sum_{i=1}^{\lfloor \nu t\rfloor}X_i\right] &= \frac{1}{\sqrt{t}}\sum_{i=1}^{\lfloor \nu t\rfloor}\mathbb{E}[Y_i-X_i] \\ &= \frac{1}{\sqrt{t}}\sum_{i=1}^{\lfloor \nu t\rfloor}\left(\boldsymbol{f}-f\left(\frac{\nu t-i}{\nu}\right)\right) \\ &= \frac{1}{\sqrt{t}}\sum_{i=1}^{\lfloor \nu t\rfloor-1}\left(\boldsymbol{f}-f\left(\frac{i}{\nu}\right)\right),\end{aligned}$$

since $\mathbb{E}[Y_i]=\lim_{t\to\infty}f(t)=\boldsymbol{f}$, $\mathbb{E}[X_i]=f\left(\frac{\nu t-i}{\nu}\right)$ and $f(t)=1+\int_0^t h(t-s)f(s)ds$ the renewal equation. Note that

$$\boldsymbol{f}=\mathbb{E}[Y_i]=\frac{1}{1-||h||_{L^1}}.$$

We have that,

$$\boldsymbol{f}-f(t)=\frac{\int_t^\infty h(s)ds}{1-||h||_{L^1}}+\int_0^t h(t-s)(\boldsymbol{f}-f(s))ds,$$

since

$$\begin{aligned}
\frac{\int_t^\infty h(s)ds}{1-||h||_{L^1}} + \int_0^t h(t-s)(\boldsymbol{f} - f(s))ds &= \boldsymbol{f}\int_t^\infty h(s)ds + \boldsymbol{f}\int_0^t h(t-s)ds - \int_0^t h(t-s)f(s)ds \\
&= \boldsymbol{f}\left(\int_t^\infty h(s)ds + \int_0^t h(t-s)ds\right) - \int_0^t h(t-s)f(s)ds \\
&= \boldsymbol{f}\left(\int_t^\infty h(s)ds + \int_0^t h(t-s)ds\right) - (f(t)-1) \\
&= \boldsymbol{f}\left(\int_t^\infty h(s)ds + \int_0^t h(t-s)ds\right) + 1 - f(t) \\
&= \boldsymbol{f}\left(\int_0^\infty h(s)ds - \int_0^t h(s)ds + \int_0^t h(t-s)ds\right) + 1 - f(t) \\
&= \boldsymbol{f}\left(||h||_{L^1} - \int_0^t h(s)ds + \int_0^t h(t-s)ds\right) + 1 - f(t) \\
&= \boldsymbol{f}||h||_{L^1} + 1 - f(t) \\
&= \boldsymbol{f} - f(t).
\end{aligned}$$

Then,

$$\boldsymbol{f} - f(s) = \frac{\int_s^\infty h(u)du}{1-||h||_{L^1}} + \int_0^s h(s-u)(\boldsymbol{f} - f(u))du.$$

Integrating both sides yields

$$\begin{aligned}
\int_0^{\nu t} (\boldsymbol{f} - f(s))\,ds &= \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{1-||h||_{L^1}} + \int_0^s h(s-u)(\boldsymbol{f} - f(u))du\right)ds \\
&= \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{1-||h||_{L^1}}\right)ds + \int_0^{\nu t}\left(\int_0^s h(s-u)(\boldsymbol{f} - f(u))du\right)ds \\
&= \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{1-||h||_{L^1}}\right)ds + \int_0^{\nu t}\left(\int_u^{\nu t} h(s-u)ds\right)(\boldsymbol{f} - f(u))\,du \\
&\le \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{1-||h||_{L^1}}\right)ds + ||h||_{L^1}\int_0^{\nu t}(\boldsymbol{f} - f(u))\,du
\end{aligned}$$

This implies that

$$\begin{aligned}
\int_0^{\nu t} (\boldsymbol{f} - f(s))\,ds &\le \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{1-||h||_{L^1}}\right)ds + ||h||_{L^1}\int_0^{\nu t}(\boldsymbol{f} - f(s))\,ds, \\
(1-||h||_{L^1})\int_0^{\nu t} (\boldsymbol{f} - f(s))\,ds &\le \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{1-||h||_{L^1}}\right)ds.
\end{aligned}$$

So,

$$\int_0^{\nu t} (\boldsymbol{f} - f(s))\,ds \le \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{(1-||h||_{L^1})^2}\right)ds.$$

Hence, we can approximate

$$\int_0^{\nu t} (\boldsymbol{f} - f(s))\,ds \quad \text{by} \quad \int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{(1-||h||_{L^1})^2}\right)ds \quad \text{as} \quad t\to\infty.$$

Thus,

$$\lim_{t\to\infty}\frac{1}{\sqrt{t}}\int_0^{\nu t}(\boldsymbol{f}-f(s))\,ds=\lim_{t\to\infty}\frac{1}{\sqrt{t}}\int_0^{\nu t}\left(\frac{\int_s^\infty h(u)du}{(1-||h||_{L^1})^2}\right)ds=0.$$

Lastly, by monotonicity of $\boldsymbol{f}-f(t)$ as a function of t, we obtain

$$\lim_{t\to\infty}\frac{1}{\sqrt{t}}\sum_{i=1}^{\lfloor \nu t\rfloor -1}\left(\boldsymbol{f}-f\left(\frac{i}{\nu}\right)\right)=0.$$

This gives the statement.

■

Chapter 4

Mutually Exciting Hawkes Process

4.1 Mutually Exciting Hawkes Process

In this section we will define the mutually exciting Hawkes process. Before we do, let us recall the self-exciting Hawkes process, also known as the (linear) Hawkes process. The (linear) Hawkes process is defined as a counting process $(N_t)_{t\geq 0}$ with associated history $(\mathcal{H}_t)_{t\geq 0}$ that satisfies

$$\mathbb{P}(N_{t+k} - N_t = m \mid \mathcal{H}_t) = \begin{cases} \lambda(t)k + o(k), & m = 1 \\ o(k), & m > 1 \\ 1 - \lambda(t)k + o(k), & m = 0 \end{cases}.$$

The corresponding conditional intensity function is given by

$$\lambda(t) = \nu + \int_0^t h(t-s)dN_s,$$

where we call $\nu > 0$ the background intensity and the function $h : (0,\infty) \to [0,\infty)$ the excitation function. A collection of one-dimensional (linear) Hawkes process that self-excites and excites each other is called a mutually exciting Hawkes process. The formal definition is given below.

Definition 4.1.1 (Mutually exciting Hawkes process)**.** A collection of n counting processes $(N_{i,t})_{t\geq 0, i\in\{1,\dots,n\}}$ is called a mutually exciting Hawkes process if the conditional intensity functions are defined as

$$\lambda_i(t) = \nu_i + \sum_{j=1}^{n} \int_0^t h_{i,j}(t-s)dN_{j,s}, \quad i = 1,\dots,n.$$

We call $\nu_i > 0$ the background intensities and the functions $h_{i,j} : (0,\infty) \to [0,\infty)$ are the excitation functions.

As we have seen in Chapter 3, the (linear) Hawkes process is a self-exciting process, so it has one counting process. The mutually exciting process has as its name already suggests multiple counting processes with each a conditional intensity function that depends on each other. Simply stated, as one process spikes we expect that the others will do so as well. For example, let us consider the following situation. One counting process models the number of car accidents in a specified region and the second counting process models the number of bike accidents in the same specified region. Then we expect that if there are a high number of car accidents that the probability that a bike accident will happen will be larger. So, the two random events depend on each other. This is a very simplified explanation of a mutually exciting Hawkes process.

The mutually exciting Hawkes processes are, for instance, also usable in finance. In fact, we will encounter the mutually exciting Hawkes process in Chapter 8, which is about the Hawkes jump-diffusion model. Simply stated the Hawkes jump-diffusion model is able to adequately capture the behavior of financial contagion.

Example 4.1.1. Let us consider a 3-dimensional Hawkes process. This means that we will consider a collection of three counting processes $\{(N_t)_{t\geq 0}\}_{i=1}^{3}$ with corresponding conditional intensity functions

$$\lambda_i(t) = \nu_i + \sum_{j=1}^{3} \int_0^t h_{i,j}(t-s)dN_{j,s}, \quad i \in \{1,2,3\}.$$

For this example, we consider the exponentially conditional intensity function, so the conditional intensity function with the exponential kernel

$$h_{i,j}(t) = \alpha_{ij} e^{-\beta_{ij} t}.$$

The parameters are given by

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} = \begin{pmatrix} 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 \\ 0.2 & 0.2 & 0.2 \end{pmatrix}$$

and

$$\begin{pmatrix} \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \\ \beta_{31} & \beta_{32} & \beta_{33} \end{pmatrix} = \begin{pmatrix} 3 & 3 & 3 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{pmatrix}.$$

The background intensity $\nu_i = 0.5$ for all $i \in \{1,2,3\}$.

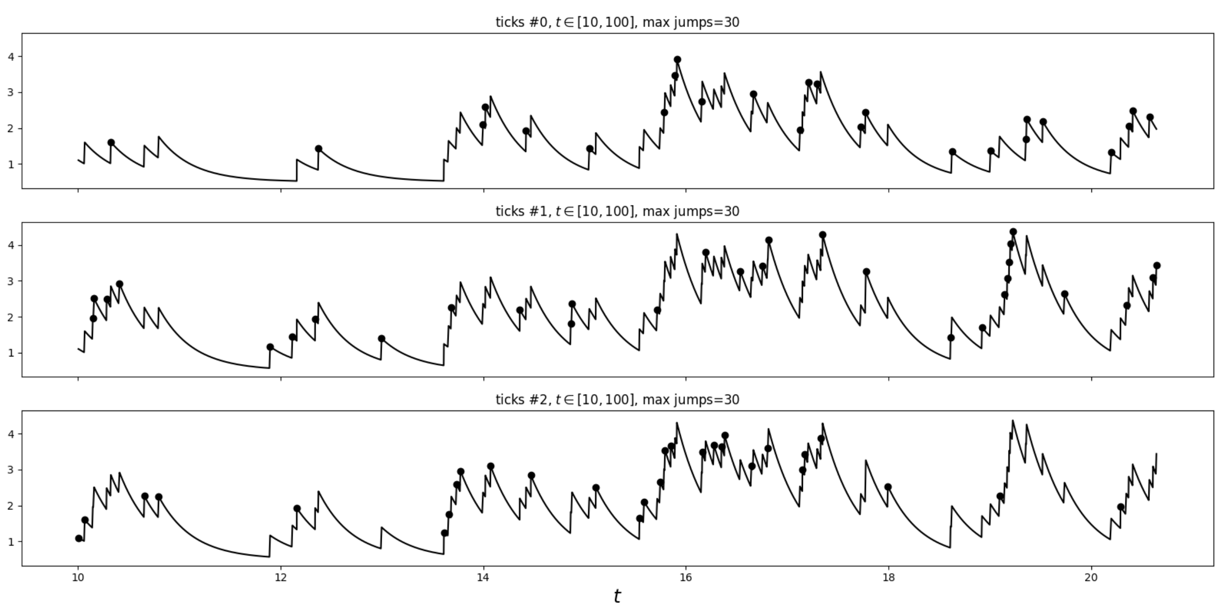

Figure 4.1: Realization of a 3-dimensional Hawkes process. The plot at the top is the conditional intensity function $\lambda_1(t)$, in the middle we have $\lambda_2(t)$ and lastly $\lambda_3(t)$.
Reference: E. Bacry, M. Bompaire, S. Gaiffas, M. Morel, S. Poulsen (see Appendix 11.2)

♦

4.2 Parameter Estimation

This section is based on Chapter 5.3 of *The Elements of Hawkes Processes* [16] with exception of the proof of Theorem 4.2.1. Also, note that the notation is changed.

Theorem 4.2.1 (Multivariate Hawkes process likelihood)**.** Let $\mathbf{N}(\cdot) = (N_1(\cdot),...,N_n(\cdot))$ be a collection simple point processes on $[0,T]$ for some finite positive T and let $\{(t_1,d_1),(t_2,d_2),...,(t_p,d_p)\}$ denote a sequence of arrival times on $[0,T]$ with $\mathbf{N}(T) = \sum_{k=1}^{n} N_k(T)$. The index d_i assigns time t_i to the component d_i of the simple point process. The likelihood function L of $\mathbf{N}(\cdot)$ is then of the form

$$L((t_1,d_1),(t_2,d_2),...,(t_p,d_p)|\theta) = \left(\prod_{i=1}^{p} \lambda^*_{d_i}(t_i)\right) \cdot \exp\left\{-\sum_{k=1}^{n}\int_0^T \lambda^*_k(u)du\right\}$$

and the log-likelihood l has the form

$$l((t_1,d_1),(t_2,d_2),...,(t_p,d_p)|\theta) = \sum_{i=1}^{p} \lambda^*_{d_i}(t_i) - \sum_{k=1}^{n}\int_0^T \lambda^*_k(u)du.$$

Remark. θ contains the background intensity ν and the parameters necessary for the excitation function h. In the case of the exponentially decaying intensities, we have that $\theta \coloneqq (\nu_i, \alpha_{ij}, \beta_{ij})$ for $i,j = 1,2,...,n$.

Proof. The likelihood function in general is given by

$$L((t_1,d_1),(t_2,d_2),...,(t_p,d_p)|\theta) = \prod_{i=1}^{p} f^*_{d_i}(t_i).$$

Using the same kind of argument as in the proof of Theorem 3.2.1, we may write

$$f^*_{d_i}(t_i) = \lambda^*_{d_i}(t_i) \cdot \exp\left\{-\sum_{k=1}^{n}\int_{t_{i-1}}^{t_i} \lambda^*_k(u)du\right\}.$$

Hence, the likelihood function becomes

$$\begin{aligned} L((t_1,d_1),(t_2,d_2),...,(t_p,d_p)|\theta) &= \prod_{i=1}^{p} f^*_{d_i}(t_i) \\ &= \prod_{i=1}^{p}\left(\lambda^*_{d_i}(t_i) \cdot \exp\left\{-\sum_{k=1}^{n}\int_{t_{i-1}}^{t_i} \lambda^*_k(u)du\right\}\right) \\ &= \left(\prod_{i=1}^{p} \lambda^*_{d_i}(t_i)\right) \cdot \exp\left\{-\sum_{k=1}^{n}\int_0^T \lambda^*_k(u)du\right\}. \end{aligned}$$

The log-likelihood is then given by

$$\begin{aligned} l((t_1,d_1),(t_2,d_2),...,(t_p,d_p)|\theta) &= \log\left(\left(\prod_{i=1}^{p} \lambda^*_{d_i}(t_i)\right) \cdot \exp\left\{-\sum_{k=1}^{n}\int_0^T \lambda^*_k(u)du\right\}\right) \\ &= \sum_{i=1}^{p} \lambda^*_{d_i}(t_i) - \sum_{k=1}^{n}\int_0^T \lambda^*_k(u)du. \end{aligned}$$

■

In Chapter 9 one may find some numerical studies regarding the parameters and likelihood estimation for a self-exciting Hawkes process with an exponentially decaying intensity and the power law excitation function.

Chapter 5

Marked Hawkes Process

5.1 Marked Hawkes Process

This section is largely based on the article of J.G. Rasmussen [21] with the exception of example 5.1.1.

Definition 5.1.1 ((Linear) Marked Hawkes process)**.** A counting process $(N_t)_{t\geq 0}$ is called a Marked Hawkes process if it has the following conditional intensity function

$$\lambda(t) = \nu + \sum_{t_i < t} h(t - t_i, \xi_i).$$

We call $\nu > 0$ the background intensity, $(t_i)_{i\geq 1}$ are the arrival times of the points, $(\xi_i)_{i\geq 1}$ are i.i.d. random marks and the function $h : (0,\infty) \to [0,\infty)$ is referred to as the excitation function. Note that ξ_i is independent of the previous arrival times t_j with $j \leq i$.

The conditional density and conditional distribution of mark ξ associated with time t can be defined. We denote with $F^*(\xi, t) \coloneqq F(\xi, t \mid \mathcal{H}_s)$, $s < t$, the conditional distribution function, where $\mathcal{H}(\cdot)$ contains the information of both times and marks of past events. The conditional density function is denoted by $f^*(\xi \mid t) \coloneqq f(\xi \mid t, \mathcal{H}_s)$, $s < t$. Using these characterization, we may now define the Hazard function for the marked case.

Definition 5.1.2 (Marked Hazard function)**.** The Hazard function for the marked case is defined as

$$\lambda^*(t,\xi) = \lambda^*(t) f^*(\xi|t)$$

with $\lambda^*(t) = \frac{f^*(t)}{1-F^*(t)}$ the Hazard function for the unmarked case, Definition 1.2.4.

We may rewrite $\lambda^*(t,\xi)$ as

$$\lambda^*(t,\xi) = \frac{f^*(t)}{1-F^*(t)} \cdot f^*(\xi|t) = \frac{f^*(t,\xi)}{1-F^*(t)} \tag{5.1}$$

with $f^*(t,\xi)$ the joint conditional density function of the time and the mark.

As an example of the marked Hawkes process we will explain the ETAS model. ETAS stands for Epidemic Type Aftershock Sequence and is used to model earthquake times and magnitudes.

Example 5.1.1. Let $m_i \in [0,\infty]$ denote the magnitude of an earthquake occurring at time t_i. Notice that the m_i is the random mark ξ_i in our model. The ground state also known as the Hazard function for the unmarked case of the ETAS model is given by

$$\lambda^*(t) = \nu + \sum_{t_i<t} \alpha e^{\beta m_i} e^{-\gamma(t-t_i)},$$

with $\nu, \alpha, \beta, \gamma > 0$ some constants. Note that this ground state looks similar to the exponentially decaying intensity function

$$\lambda(t) = \nu + \sum_{t_i<t} \alpha e^{-\beta(t-t_i)}$$

for $\alpha, \beta > 0$ some constants.

The conditional density function $f^*(m|t)$ is given by

$$f^*(m|t) = \mu e^{\mu m},$$

the probability density function of an exponential distribution with parameter $\mu > 0$.

Then the marked Hazard function is defined as

$$\lambda^*(t,m) = \lambda^*(t) f^*(m|t) = \left(\nu + \sum_{t_i<t} \alpha e^{\beta m_i} e^{-\gamma(t-t_i)}\right) \cdot \mu e^{\mu m}.$$

The ETAS model is capable of modeling the aftershocks that might occur after an earthquake, since each event that occurs will increase the probability of another event occurring by a certain factor. In this case, the event will be an aftershock and the factor will be equal to $\alpha e^{\beta m_i}$. Notice that α and β are some positive constants, so the factor will depend on the random magnitude m_i. That means that the factor will be larger for earthquakes with an higher magnitude.

Figure 5.1 shows the occurrence of earthquakes in 2019 in Ridgecrest, California. Furthermore, it shows the fitted ETAS model as well as the Omori formula, which we refer to as the power law function

$$h(t) = \frac{k}{(c+t)^p}.$$

Note that the probability that an event, in this case an aftershock, occurs depends on the magnitude of the earthquake. A higher magnitude generally implies a higher probability of another event occurring.

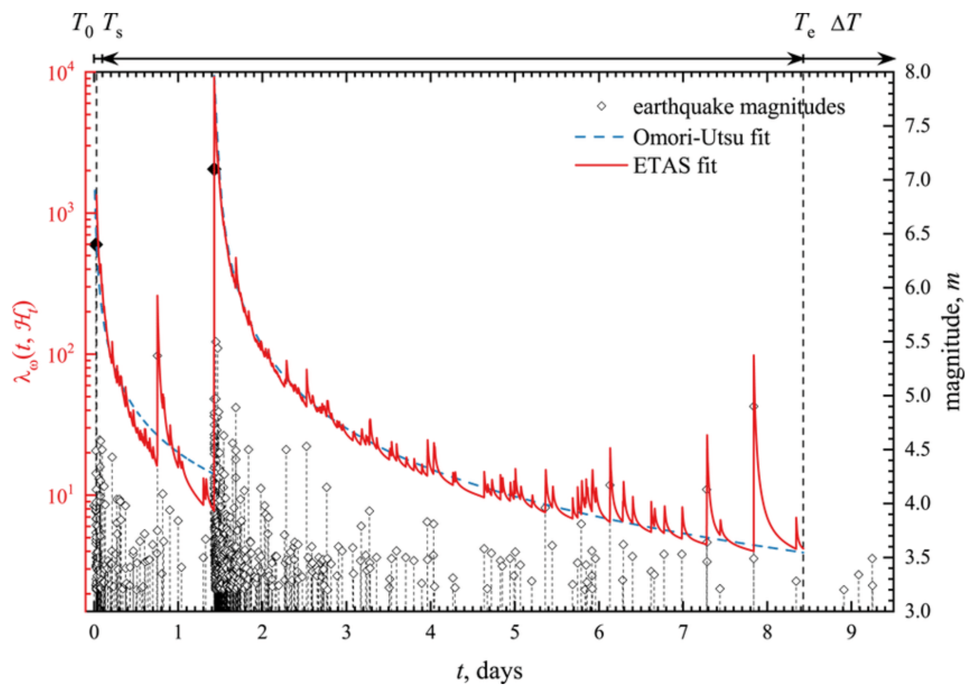

Figure 5.1: Realization of a marked Hawkes process applied to seismology.
Reference: ([28], Figure 5, p. 11)

♦

5.2 Parameter Estimation

This section is largely based on the article of J.G. Rasmussen [21]. Note that the proof of Theorem 5.2.1 differs a bit from the article, namely we made use of Section 3.2.

Theorem 5.2.1 (Marked Hawkes process likelihood)**.** Let $N(\cdot)$ be a simple point process on $[0,T]$ for some finite positive T and let $\{(t_1,\xi_1),(t_2,\xi_2),...,(t_n,\xi_n)\}$ denote a realization on $[0,T]\times\mathbb{M}$, where $\mathbb{M}$ stands for the marked space (usually $\mathbb{R}$ or $\mathbb{N}$). The likelihood function L is then of the form

$$L((t_1,\xi_1),(t_2,\xi_2),...,(t_n,\xi_n)|\theta) = \left(\prod_{i=1}^{n} \lambda^*(t_i,\xi_i)\right) \cdot \exp\left\{-\int_0^T \lambda^*(u)du\right\}$$

and the log-likelihood l has the form

$$l((t_1,\xi_1),(t_2,\xi_2),...,(t_n,\xi_n)|\theta) = \sum_{i=1}^{n} \lambda^*(t_i,\xi_i) - \int_0^T \lambda^*(u)du.$$

Remark. θ contains the background intensity ν and the parameters necessary for the excitation function h.

Proof. The likelihood function in general is given by

$$L((t_1,\xi_1),(t_2,\xi_2),...,(t_n,\xi_n)|\theta) = \prod_{i=1}^{n} f^*(t_i,\xi_i) = \prod_{i=1}^{n} f^*(t_i)f(\xi_i|t_i).$$

Using the proof of Theorem 3.2.1, we may write

$$f^*(t_i) = \lambda^*(t_i) \cdot \exp\left\{-\int_{t_{i-1}}^{t_i} \lambda^*(u)du\right\}.$$

Hence, the likelihood function becomes

$$\begin{aligned}
L((t_1,\xi_1),(t_2,\xi_2),...,(t_n,\xi_n)|\theta) &= \prod_{i=1}^{n} f^*(t_i) f(\xi_i|t_i) \\
&= \prod_{i=1}^{n}\left(\left(\lambda^*(t_i)\cdot\exp\left\{-\int_{t_{i-1}}^{t_i}\lambda^*(u)du\right\}\right)\cdot f(\xi_i|t_i)\right) \\
&= \left(\prod_{i=1}^{n}\lambda^*(t_i)f(\xi_i|t_i)\right)\cdot\exp\left\{-\int_0^T \lambda^*(u)du\right\} \\
&= \left(\prod_{i=1}^{n}\lambda^*(t_i,\xi_i)\right)\cdot\exp\left\{-\int_0^T \lambda^*(u)du\right\},
\end{aligned}$$

since $\lambda^*(t_i,\xi_i) = \lambda^*(t_i)f^*(\xi_i|t_i)$.

The log-likelihood function is then given by

$$\begin{aligned}
l((t_1,\xi_1),(t_2,\xi_2),...,(t_n,\xi_n)|\theta) &= \log\left(\left(\prod_{i=1}^{n}\lambda^*(t_i,\xi_i)\right)\cdot\exp\left\{-\int_0^T\lambda^*(u)du\right\}\right) \\
&= \sum_{i=1}^{n}\lambda^*(t_i,\xi_i) - \int_0^T \lambda^*(u)du.
\end{aligned}$$

■

In Chapter 9, we will provide some numerical examples regarding the parameters and likelihood estimation for a self-exciting Hawkes process with an exponentially decaying intensity and the power law excitation function.

5.3 Limit Theorems

This section is largely based on the article of L. Zhu [37]. Note that our notation differs from the article and that we added more details to the proof. We will recall the conditional intensity for marked Hawkes processes and after that we will define the assumptions needed for proving the Central Limit Theorem. Let us first start by noting the differences between the limit theorems for the linear Hawkes process and marked Hawkes process. Due to the extra randomness in the marked case we will expect some differences in the Law of Large Numbers and the Central Limit Theorem, the latter we will prove. In the linear Hawkes process we saw that the Law of Large Numbers was given by

$$\frac{N_t}{t} \xrightarrow{\mathbb{P}} \frac{\nu}{1-||h||_{L^1}} \quad \text{as} \quad t\to\infty$$

and the Central Limit Theorem by

$$\sqrt{t}\left(\frac{N_t}{t} - \frac{\nu}{1-||h||_{L^1}}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu||h||_{L^1}}{(1-||h||_{L^1})^3}\right) \quad \text{as} \quad t\to\infty.$$

We will see that the Law of Large Numbers and Central Limit Theorem are given by

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{\nu}{1-\mathbb{E}[H(\xi)]} \quad \text{as} \quad t \to \infty$$

and

$$\sqrt{t}\left(\frac{N_t}{t} - \frac{\nu}{1-\mathbb{E}[H(\xi)]}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu(1+Var(H(\xi)))}{(1-\mathbb{E}[H(\xi)])^3}\right) \quad \text{as} \quad t \to \infty,$$

respectively. The proof of the Central Limit Theorem for the marked Hawkes process will also be entirely different compared to the linear Hawkes process due to the added randomness of the marks.

The conditional intensity function for the marked Hawkes process is given by

$$\lambda(t) = \nu + \sum_{t_i<t} h(t-t_i, \xi_i).$$

We call $\nu > 0$ the background intensity, $(t_i)_{i\geq 1}$ are the arrival times of the points, $(\xi_i)_{i\geq 1}$ are i.i.d. random marks and the function h is referred to as the excitation function. Note that ξ_i is independent of the previous arrival times t_j with $j \leq i$. We assume that the random marks have a common distribution $F(\xi)$ and density $f(\xi)$ on a measurable space $\mathbb{S}$.

Assumptions:

(i) We assume that $h(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{S} \to \mathbb{R}^+$ is integrable.

(ii) We define $H(\xi) \coloneqq \int_0^\infty h(t,\xi)dt$ for all $\xi \in \mathbb{S}$. Then, we assume that $\int_{\mathbb{S}} H(\xi)f(\xi)d\xi < 1$.

The assumptions assure that there exists a linear marked Hawkes process and that the limits are well-defined.

Lemma 5.3.1. For all $t > 1$, we have that $\mathbb{E}[\lambda(t)] \leq \frac{\nu}{1-\mathbb{E}[H(\xi)]}$ is uniformly over t.

Proof.

$$\begin{aligned}
\mathbb{E}[\lambda(t)] &= \mathbb{E}[\nu + \int_{-\infty}^{t} h(t-s,\xi)dN_s d\xi] \\
&= \nu + \mathbb{E}[\int_{-\infty}^{t} h(t-s,\xi)dN_s d\xi] \\
&\approx \nu + \mathbb{E}[\int_{-\infty}^{t} h(t-s,\xi)\lambda(s)ds d\xi] \qquad \text{since} \qquad dN_s \approx \lambda(s)ds \\
&= \nu + \mathbb{E}[\int_{0}^{\infty} h(s,\xi)\lambda(t-s)ds d\xi]
\end{aligned}$$

Since $h(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{S} \to \mathbb{R}^+$ non-increasing and $\nu > 0$, we have that $\lambda(t-s) \leq \lambda(t)$. Therefore,

$$\mathbb{E}[\lambda(t)] \leq \nu + \mathbb{E}[\lambda(t)]\mathbb{E}[H(\xi)].$$

Thus,

$$\mathbb{E}[\lambda(t)]\,(1-\mathbb{E}[H(\xi)]) \leq \nu.$$

Hence, for all $t > 1$, it holds that $\mathbb{E}[\lambda(t)] \leq \frac{\nu}{1-\mathbb{E}[H(\xi)]}$.

■

Theorem 5.3.2 (Central Limit Theorem for marked Hawkes process)**.** Assume that the provided assumptions are satisfied and assume that $\lim_{t\to\infty}\sqrt{t}\int_t^\infty \mathbb{E}[h(s,\xi)]ds = 0$. Then,

$$\sqrt{t}\left(\frac{N_t}{t} - \frac{\nu}{1-\mathbb{E}[H(\xi)]}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu(1+Var(H(\xi)))}{(1-\mathbb{E}[H(\xi)])^3}\right) \quad \text{as} \quad t\to\infty.$$

Proof. Let us recall that the conditional intensity function for the marked Hawkes process is given by

$$\lambda(t) = \nu + \sum_{t_i<t} h(t-t_i,\xi_i).$$

If we integrate both sides from zero to t, we obtain the following

$$\begin{aligned}\Lambda(t) &:= \int_0^t \lambda(s)ds = \int_0^t \nu ds + \sum_{t_i<t}\int_0^t h(s-t_i,\xi_i)ds \\ &= \nu t + \sum_{t_i<t}\int_0^t h(s-t_i,\xi_i)ds \\ &= \nu t + \sum_{t_i<t} H(\xi_i) - \sum_{t_i<t}\int_t^\infty h(s-t_i,\xi_i)ds \\ &= \nu t + \sum_{t_i<t} H(\xi_i) - E_t,\end{aligned}$$

since $H(\xi) := \int_0^\infty h(t,\xi)dt$. We define $E_t := \sum_{t_i<t}\int_t^\infty h(s-t_i,\xi_i)ds$ as the error term.

Furthermore,

$$\begin{aligned}\frac{N_t - \int_0^t \lambda(s)ds}{\sqrt{t}} &= \frac{N_t - \nu t - \sum_{t_i<t} H(\xi_i) + E_t}{\sqrt{t}} \\ &= \frac{N_t - \nu t - \sum_{t_i<t} H(\xi_i)}{\sqrt{t}} + \frac{E_t}{\sqrt{t}} \\ &= \frac{N_t - N_t\mathbb{E}[H(\xi)] + N_t\mathbb{E}[H(\xi)] - \nu t - \sum_{t_i<t} H(\xi_i)}{\sqrt{t}} + \frac{E_t}{\sqrt{t}} \\ &= \frac{(1-\mathbb{E}[H(\xi)])N_t - \nu t}{\sqrt{t}} + \frac{N_t\mathbb{E}[H(\xi)] - \sum_{t_i<t} H(\xi_i)}{\sqrt{t}} + \frac{E_t}{\sqrt{t}} \\ &= (1-\mathbb{E}[H(\xi)])\cdot\frac{N_t - \frac{\nu t}{(1-\mathbb{E}[H(\xi)])}}{\sqrt{t}} + \frac{N_t\mathbb{E}[H(\xi)] - \sum_{t_i<t} H(\xi_i)}{\sqrt{t}} + \frac{E_t}{\sqrt{t}} \\ &= (1-\mathbb{E}[H(\xi)])\cdot\frac{N_t - \mu t}{\sqrt{t}} + \frac{N_t\mathbb{E}[H(\xi)] - \sum_{t_i<t} H(\xi_i)}{\sqrt{t}} + \frac{E_t}{\sqrt{t}},\end{aligned}$$

with $\mu := \frac{\nu}{1-\mathbb{E}[H(\xi)]}$.
Hence, rearranging gives us

$$(1-\mathbb{E}[H(\xi)])\cdot\frac{N_t - \mu t}{\sqrt{t}} = \frac{N_t - \int_0^t \lambda(s)ds}{\sqrt{t}} - \frac{N_t\mathbb{E}[H(\xi)] - \sum_{t_i<t} H(\xi_i)}{\sqrt{t}} - \frac{E_t}{\sqrt{t}}.$$

Dividing both sides by $1-\mathbb{E}[H(\xi)]$, which is unequal to zero due to the assumptions, we obtain

$$\begin{aligned}\frac{N_t-\mu t}{\sqrt{t}} &= \frac{1}{1-\mathbb{E}[H(\xi)]}\left(\frac{N_t-\int_0^t\lambda(s)ds}{\sqrt{t}}-\frac{N_t\mathbb{E}[H(\xi)]-\sum_{t_i<t}H(\xi_i)}{\sqrt{t}}-\frac{E_t}{\sqrt{t}}\right)\\ &= \frac{1}{1-\mathbb{E}[H(\xi)]}\left(\frac{N_t-\int_0^t\lambda(s)ds}{\sqrt{t}}+\frac{\sum_{t_i<t}(H(\xi_i)-\mathbb{E}[H(\xi)])}{\sqrt{t}}-\frac{E_t}{\sqrt{t}}\right),\end{aligned}$$

since N_t is a simple point process, where $N_t := N(0,t]$ is defined as the number of points in the interval $(0,t]$ and that can be denoted with the given random sum.
We will now show that $\frac{E_t}{\sqrt{t}} \xrightarrow{\mathbb{P}} 0$ as $t\to\infty$.

We have that $\mathbb{E}[\lambda(t)] \leq \frac{\nu}{1-\mathbb{E}[H(\xi)]}$ uniformly over t by Lemma 5.3.1.

Then:

$$\begin{aligned}\mathbb{E}[E_t] &= \mathbb{E}[\sum_{t_i<t}\int_t^\infty h(s-t_i,\xi_i)ds]\\ &= \int_0^t\int_{\mathbb{S}}\left(\int_{t-s}^\infty h(u,\xi)du\right)f(\xi)\mathbb{E}[\lambda(s)]d\xi ds\\ &\leq \frac{\nu}{1-\mathbb{E}[H(\xi)]}\int_0^t\int_{\mathbb{S}}\left(\int_{t-s}^\infty h(u,\xi)du\right)f(\xi)d\xi ds\\ &= \frac{\nu}{1-\mathbb{E}[H(\xi)]}\int_0^t\mathbb{E}[\int_s^\infty h(u,\xi)du]ds.\end{aligned}$$

Using L'Hôpital's rule, we obtain

$$\begin{aligned}\lim_{t\to\infty}\frac{1}{\sqrt{t}}\int_0^t\mathbb{E}[\int_s^\infty h(u,\xi)du]ds &= \lim_{t\to\infty}\frac{1}{\frac{1}{\frac{1}{2}\sqrt{t}}}\mathbb{E}[\int_s^\infty h(u,\xi)du]\\ &= \lim_{t\to\infty}2\sqrt{t}\int_s^\infty\mathbb{E}[h(u,\xi)]du \quad \text{by Fubini Theorem}\\ &= 2\lim_{t\to\infty}\sqrt{t}\int_s^\infty\mathbb{E}[h(u,\xi)]du = 0, \quad \text{by assumption.}\end{aligned}$$

By Markov's inequality, we have that for all constant c it holds

$$\mathbb{P}\left(\frac{E_t}{\sqrt{t}}\geq c\right) = \mathbb{P}(E_t\geq c\sqrt{t}) \leq \frac{\mathbb{E}[E_t]}{c\sqrt{t}} \to 0 \quad \text{as} \quad t\to\infty.$$

Hence, $\frac{E_t}{\sqrt{t}} \xrightarrow{\mathbb{P}} 0$ as $t\to\infty$.

Let us define $M_1(t) := N_t - \int_0^t\lambda(s)ds$ and $M_2(t) := \sum_{t_i<t}(H(\xi_i)-\mathbb{E}[H(\xi)])$. We show that M_1 and M_2 are martingales with respect to the filtration $(\mathcal{H}_t)_{t\geq 0}$. We start with showing that

$$\mathbb{E}[M_1(t)-M_1(u)|\mathcal{H}_u] = 0 \quad \text{for} \quad u\leq t.$$

$$\begin{aligned}\mathbb{E}[M_1(t)-M_1(u)|\mathcal{H}_u] &= \mathbb{E}[N_t - \int_0^t \lambda(s)ds - (N_u - \int_0^u \lambda(s)ds)|\mathcal{H}_u] \\ &= \mathbb{E}[N_t - N_u - (\int_0^t \lambda(s)ds - \int_0^u \lambda(s)ds)|\mathcal{H}_u] \\ &= \mathbb{E}[N_t - N_u - \int_u^t \lambda(s)ds|\mathcal{H}_u] \\ &= \mathbb{E}[N_t - N_u|\mathcal{H}_u] - \mathbb{E}[\int_u^t \lambda(s)ds|\mathcal{H}_u] \\ &= \mathbb{E}[N(u,t]|\mathcal{H}_u] - \mathbb{E}[\int_u^t \lambda(s)ds|\mathcal{H}_u] = 0,\end{aligned}$$

by definition. Hence, M_1 is a martingale with respect to the filtration $(\mathcal{H}_t)_{t\geq 0}$. We now show that

$$\mathbb{E}[M_2(t)-M_2(u)|\mathcal{H}_u] = 0 \quad \text{for} \quad u \leq t.$$

So,

$$\begin{aligned}\mathbb{E}[M_2(t)-M_2(u)|\mathcal{H}_u] &= \mathbb{E}[\sum_{t_i<t}(H(\xi_i)-\mathbb{E}[H(\xi)]) - \sum_{t_i<u}(H(\xi_i)-\mathbb{E}[H(\xi)])|\mathcal{H}_u] \\ &= \mathbb{E}[\sum_{u<t_i<t}(H(\xi_i)-\mathbb{E}[H(\xi)])|\mathcal{H}_u] \\ &= \mathbb{E}[\sum_{u<t_i<t}(H(\xi_i)-\mathbb{E}[H(\xi)])] = 0,\end{aligned}$$

by definition. Thus, M_2 is a martingale with respect to the filtration $(\mathcal{H}_t)_{t\geq 0}$ as well. Moreover, since $\int_0^t \lambda(s)ds$ has a finite variation, the quadratic variation of $M_1(t)+M_2(t)$ is the same as the quadratic variation of $N_t + M_2(t)$. So

$$\begin{aligned}N_t + M_2(t) &= N_t + \sum_{t_i<t}(H(\xi_i)-\mathbb{E}[H(\xi)]) \\ &= \sum_{t_i<t}(1+H(\xi_i)-\mathbb{E}[H(\xi)])\end{aligned}$$

Thus the quadratic variation is given by

$$[N+M_2]_t = \sum_{t_i<t}(1+H(\xi_i)-\mathbb{E}[H(\xi)])^2.$$

By the standard Law of Large Numbers 1.1.6, we have that

$$\begin{aligned}\frac{1}{t}\sum_{t_i<t}(1+H(\xi_i)-\mathbb{E}[H(\xi)])^2 &= \frac{N_t}{t}\cdot\frac{1}{N_t}\sum_{t_i<t}(1+H(\xi_i)-\mathbb{E}[H(\xi)])^2 \\ &\xrightarrow{a.s.} \frac{\nu}{1-\mathbb{E}[H(\xi)]}\cdot\mathbb{E}[(1+H(\xi)-\mathbb{E}[H(\xi)])^2] \quad \text{as} \quad t\to\infty.\end{aligned}$$

Notice that

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{\nu}{1-\mathbb{E}[H(\xi)]} \quad \text{as} \quad t\to\infty$$

by the Law of Large Numbers.

We may now rewrite $\mathbb{E}[(1+H(\xi)-\mathbb{E}[H(\xi)])^2]$, so

$$\begin{aligned}\mathbb{E}[(1+H(\xi)-\mathbb{E}[H(\xi)])^2] &= \mathbb{E}[(1+(H(\xi)-\mathbb{E}[H(\xi)]))^2] \\ &= \mathbb{E}[1^2]+\mathbb{E}[2(H(\xi)-\mathbb{E}[H(\xi)])]+\mathbb{E}[(H(\xi)-\mathbb{E}[H(\xi)])^2] \\ &= 1+2\mathbb{E}[H(\xi)-\mathbb{E}[H(\xi)]]+\mathbb{E}[(H(\xi)-\mathbb{E}[H(\xi)])^2] \\ &= 1+Var(H(\xi)).\end{aligned}$$

Thus,

$$\frac{1}{t}\sum_{t_i<t}(1+H(\xi_i)-\mathbb{E}[H(\xi)])^2 \xrightarrow{a.s.} \frac{\nu(1+Var(H(\xi)))}{1-\mathbb{E}[H(\xi)]} \quad \text{as} \quad t\to\infty.$$

Then by the standard martingale Central Limit Theorem 5.3.3, we have that

$$\sqrt{t}\left(\frac{N_t}{t}-\frac{\nu}{1-\mathbb{E}[H(\xi)]}\right) \rightsquigarrow \mathcal{N}\left(0,\frac{\nu(1+Var(H(\xi)))}{(1-\mathbb{E}[H(\xi)])^3}\right) \quad \text{as} \quad t\to\infty.$$

■

Theorem 5.3.3 (Martingale Central Limit Theorem ([23], p. 172, Remark 14.3.3))**.** Let $(X_i)_{i=0}^n$ be a martingale with respect to the filtration $\mathcal{F}_n=\sigma(X_0,X_1,...,X_n)$. Set $\sigma_0^2=Var(X_0)$ and for $n\geq 1$ set $\sigma_n^2=Var(X_n|\mathcal{F}_{n-1})=\mathbb{E}[X_n^2-X_{n-1}^2|\mathcal{F}_{n-1}]$. Then, by induction, we have that

$$\mathbb{E}[X_n]=\mathbb{E}[X_0] \quad \text{and} \quad Var(X_n)=\sum_{k=0}^{n}\mathbb{E}[\sigma_k^2].$$

If we take $v_t=\min\{n\geq 0|\sum_{k=0}^n\sigma_k^2\geq t\}$, then $\frac{X_{v_t}}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0,1) \quad \text{as} \quad t\to\infty$.

Part III

Applications of Hawkes Processes in Finance

Chapter 6

Black-Scholes Model

In this chapter we will introduce the Black-Scholes model, which was introduced by Fischer Black and Myron Scholes in the nineteen-seventies in there article *The Pricing of Options and Corporate Liabilities* [5]. In order to derive the valuation formula, which is the formula for the value of an option in terms of the price of a stock, Black and Scholes made certain assumptions on the market [5].

Assumptions:

(i) The short-term interest rate is known and constant through time.

(ii) The stock price process follows a geometric Brownian motion with constant drift and volatility.

(iii) The stock does not pay dividends.

(iv) The option is European. This means that the option can only be exercised on the maturity time T.

(v) There are no transaction costs in buying or selling the stock or the option.

(vi) One may borrow any fraction of the price of a security to buy it or to hold it at the short-term interest rate.

(vii) Short selling will not be penalized.

Given these assumptions, the value of the option will only depend on the price of the stock, the time and known constants. Furthermore, the derivation of the Black-Scholes equation relies on the no-arbitrage argument, which was introduced by Robert Merton. No-arbitrage means that there is no riskless way to make money. Due to the involvement of Merton, the Black-Scholes equation is also known as the Black-Scholes-Merton equation. To derive the Black-Scholes equation we will make use of the book *Stochastic Calculus for Finance II - Continuous-Time Models* Chapter 4.5 written by S.E. Shreve [29].

In the Black-Scholes model, the stock price process $(S_t)_{t\geq 0}$ of a risky asset is assumed to be a geometric Brownian motion. This means that the stochastic differential equation that satisfies S_t is given by

$$dS_t = \mu S_t dt + \sigma S_t dW_t \tag{6.1}$$

with $(W_t)_{t\geq 0}$ a Brownian motion, $\mu \in \mathbb{R}$ the drift and $\sigma > 0$ the volatility. We may interpret the drift as a trend and the volatility as a variation or spread.

Lemma 6.0.1. The stock price process $(S_t)_{t\geq 0}$ of a risky asset that satisfies the stochastic differential equation

$$dS_t = \mu S_t dt + \sigma S_t dW_t$$

is given by

$$S_t = S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}.$$

Proof. In order to solve the stochastic differential equation we may apply Itô-Doeblin formula, but first we rewrite the SDE.

$$dS_t = \mu S_t dt + \sigma S_t dW_t = (\mu dt + \sigma dW_t) S_t$$

Now, we divide both sides by S_t and note that the left-hand side looks like the derivative of the logarithm. Hence, we will apply Itô-Doeblin formula to $f(x) = \log(x)$. Note that $f_x(x) = \frac{1}{x}$ and $f_{xx}(x) = -\frac{1}{x^2}$. So,

$$\begin{aligned} d(\log(S_t)) &= df(S_t) = f_x(S_t) dS_t + \frac{1}{2} f_{xx}(S_t) d[S]_t \\ &= \frac{1}{S_t} dS_t - \frac{1}{2} \cdot \frac{1}{S_t^2} d[S]_t \\ &= \frac{1}{S_t} (\mu S_t dt + \sigma S_t dW_t) - \frac{1}{2} \cdot \frac{1}{S_t^2} \cdot \sigma^2 S_t^2 dt \\ &= \mu dt + \sigma dW_t - \frac{\sigma^2}{2} dt \\ &= \left(\mu - \frac{\sigma^2}{2}\right) dt + \sigma dW_t. \end{aligned}$$

Integration on both sides yields

$$\begin{aligned} \log(S_t) - \log(S_0) &= \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t. \\ \log\left(\frac{S_t}{S_0}\right) &= \left(\mu - \frac{\sigma^2}{2}\right) t + \sigma W_t. \end{aligned}$$

Taking the exponent on both sides and multiply with S_0 gives the solution to the stochastic differential equation, hence

$$S_t = S_0 \cdot e^{\left(\mu - \frac{\sigma^2}{2}\right)t + \sigma W_t}.$$

■

Lemma 6.0.2. The stochastic differential equation of the discounted stock price process $\tilde{S}_t := e^{-rt} S_t$ with r the interest rate $(r > -1)$ is given by

$$d\tilde{S}_t = d\left(e^{-rt} S_t\right) = (\mu - r)\tilde{S}_t dt + \sigma \tilde{S}_t dW_t.$$

Proof. The differential of the discounted stock price process $\tilde{S}_t := e^{-rt}S_t$ with r the interest rate ($r > -1$) can be derived by applying Itô-Doeblin formula to $f(t,x) = e^{-rt}x$.
Note that $f_t(t,x) = -re^{-rt}x$, $f_x(t,x) = e^{-rt}$ and $f_{xx}(t,x) = 0$. Hence,

$$\begin{aligned}
d\tilde{S}_t = df(t,S_t) &= f_t(t,S_t)dt + f_x(t,S_t)dS_t + \frac{1}{2}f_{xx}(t,S_t)d[S]_t \\
&= f_t(t,S_t)dt + f_x(t,S_t)dS_t \\
&= -re^{-rt}S_tdt + e^{-rt}dS_t \\
&= -re^{-rt}S_tdt + e^{-rt}(\mu S_tdt + \sigma S_tdW_t) \\
&= (\mu - r)e^{-rt}S_tdt + \sigma e^{-rt}S_tdW_t \\
&= (\mu - r)\tilde{S}_tdt + \sigma\tilde{S}_tdW_t
\end{aligned}$$

Hence, the SDE for the discounted stock price process is

$$d\tilde{S}_t = d\left(e^{-rt}S_t\right) = (\mu - r)\tilde{S}_tdt + \sigma\tilde{S}_tdW_t.$$

■

With these stochastic differential equations in mind, we can start deriving the Black-Scholes equation. The main idea behind obtaining the Black-Scholes equation is derive the delta-hedging rule. This means that at each time t prior to the maturity time, the numbers of shares held by the hedge of an short option position is the partial derivative with respect to the stock price S_t of the option value at that certain time t. In order to derive the delta-hedging rule, we first need to derive the SDEs for the value of the portfolio and the value of an option. We will start with the evolution of the portfolio value.

Let X_t be the value of a portfolio at time t consisting of an investment in the stock market plus the money market. At time t, assume that an amount of Δ_t units, Δ_t a random variable, is invested in the stock at a price S_t. This means that the total value invested in the stock market equals Δ_tS_t. Assume that the remainder of the portfolio value, $X_t - \Delta_tS_t$, is invested in the money market for example in bonds or in a bank account, which gains an interest rate r per time unit. Then the differential dX_t consists of two components, the capital gain Δ_tdS_t on the stock position and the interest earnings $r(X_t - \Delta_tS_t)dt$ on the cash position. Hence,

$$\begin{aligned}
dX_t &= \Delta_tdS_t + r(X_t - \Delta_tS_t)dt \\
&= \Delta_t(\mu S_tdt + \sigma S_tdW_t) + r(X_t - \Delta_tS_t)dt \\
&= \Delta_t(\mu S_tdt - rS_tdt) + \sigma\Delta_tS_tdW_t + rX_tdt \\
&= \Delta_t(\mu - r)S_tdt + rX_tdt + \sigma\Delta_tS_tdW_t.
\end{aligned}$$

Lemma 6.0.3. The discounted portfolio value $e^{-rt}X_t$ satisfies the stochastic differential equation

$$d(e^{-rt}X_t) = \Delta_td\left(e^{-rt}S_t\right).$$

Proof. The SDE of the discounted portfolio value $e^{-rt}X_t$ can be computed by applying Itô-Doeblin formula to $f(t,x) = e^{-rt}x$. Note that $f_t(t,x) = -re^{-rt}x$, $f_x(t,x) = e^{-rt}$ and $f_{xx}(t,x) = 0$. Hence,

$$\begin{aligned}
d(e^{-rt}X_t) &= df(t,X_t) = f_t(t,X_t)dt + f_x(t,X_t)dX_t + \frac{1}{2}f_{xx}(t,X_t)d[X]_t \\
&= f_t(t,X_t)dt + f_x(t,X_t)dX_t \\
&= -re^{-rt}X_t dt + e^{-rt}dX_t \\
&= -re^{-rt}X_t dt + e^{-rt}\left(\Delta_t(\mu - r)S_t dt + rX_t dt + \sigma\Delta_t S_t dW_t\right) \\
&= -re^{-rt}X_t dt + \Delta_t(\mu - r)e^{-rt}S_t dt + re^{-rt}X_t dt + \sigma\Delta_t e^{-rt}S_t dW_t \\
&= \Delta_t(\mu - r)e^{-rt}S_t dt + \sigma\Delta_t e^{-rt}S_t dW_t \\
&= \Delta_t\left((\mu - r)e^{-rt}S_t dt + \sigma e^{-rt}S_t dW_t\right) \\
&= \Delta_t d\left(e^{-rt}S_t\right).
\end{aligned}$$

Therefore, the discounted portfolio value $e^{-rt}X_t$ satisfies the stochastic differential equation

$$d(e^{-rt}X_t) = \Delta_t d\left(e^{-rt}S_t\right).$$

■

Let us now consider the value of an option, more specifically an European call option. An European call option means that at maturity time T, the seller has the right, but not the obligation to sell the asset at the strike price. So, we will consider an European call option that pays $(S_T - K)^+$ at maturity time T with K the strike price, which is a constant and $K \geq 0$. We denote with $C(t,S_t)$ the value of the European call option at time t. Note that the value of the call option is a stochastic process. Our first step will be to determine the stochastic differential equation of $C(t,S_t)$.

Lemma 6.0.4. The stochastic differential equation of $C(t,S_t)$ is given by

$$dC(t,S_t) = \left(C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt + \sigma S_t C_x(t,S_t)dW_t.$$

Proof. We will apply the Itô-Doeblin formula to $C(t,x)$, so

$$\begin{aligned}
dC(t,S_t) &= C_t(t,S_t)dt + C_x(t,S_t)dS_t + \frac{1}{2}C_{xx}(t,S_t)d[S]_t \\
&= C_t(t,S_t)dt + C_x(t,S_t)\left(\mu S_t dt + \sigma S_t dW_t\right) + \frac{1}{2}C_{xx}(t,S_t)\sigma^2 S_t^2 dt \\
&= \left(C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt + \sigma S_t C_x(t,S_t)dW_t.
\end{aligned}$$

■

Now, we will derive the SDE for the discounted option price $e^{-rt}C(t,S_t)$.

Lemma 6.0.5. The stochastic differential equation of the discounted option price $e^{-rt}C(t,S_t)$ is given by

$$d(e^{-rt}C(t,S_t)) = e^{-rt}\left(-rC(t,S_t) + C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt + \sigma e^{-rt}S_t C_x(t,S_t)dW_t.$$

Proof. We will apply the Itô-Doeblin formula to $f(t,x) = e^{-rt}x$. Notice that the partial derivatives are $f_t(t,x) = -re^{-rt}x$, $f_x(t,x) = e^{-rt}$ and $f_{xx}(t,x) = 0$.

Hence,

$$
\begin{aligned}
d(e^{-rt}C(t,S_t)) &= df(t,C(t,S_t)) = f_t(t,C(t,S_t))dt + f_x(t,C(t,S_t))dC(t,S_t) + \frac{1}{2}f_{xx}(t,C(t,S_t))dC(t,S_t)dC(t,S_t) \\
&= f_t(t,C(t,S_t))dt + f_x(t,C(t,S_t))dC(t,S_t) \\
&= -re^{-rt}C(t,S_t)dt + e^{-rt}dC(t,S_t) \\
&= -re^{-rt}C(t,S_t)dt + e^{-rt}\left(\left(C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt + \sigma S_t C_x(t,S_t)dW_t\right) \\
&= e^{-rt}\left(-rC(t,S_t) + C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt + \sigma e^{-rt} S_t C_x(t,S_t)dW_t.
\end{aligned}
$$

■

To have a fair option value $C(0,S_0)$ at time 0, one should be able to start with an amount $X_0 = C(0,S_0)$ and build a portfolio process $(X_t)_{0\le t\le T}$ such that for all $t \in [0,T]$ it holds that $X_t = C(t,S_t)$. Hence, it holds that there is no-arbitrage in the financial market. In other words, we must have $e^{-rt}X_t = e^{-rt}C(t,S_t)$ for all $t \in [0,T]$. This holds if

$$d(e^{-rt}X_t) = d(e^{-rt}C(t,S_t)) \quad \text{for all} \quad t \in [0,T].$$

We know that

$$
\begin{aligned}
d(e^{-rt}X_t) &= \Delta_t\left((\mu - r)e^{-rt}S_t dt + \sigma e^{-rt}S_t dW_t\right) \quad \text{and} \\
d(e^{-rt}C(t,S_t)) &= e^{-rt}\left(-rC(t,S_t) + C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt + \sigma e^{-rt} S_t C_x(t,S_t)dW_t.
\end{aligned}
$$

Thus $d(e^{-rt}X_t) = d(e^{-rt}C(t,S_t))$ holds if and only if

(i)

$$\Delta_t \sigma e^{-rt} S_t dW_t = \sigma e^{-rt} S_t C_x(t,S_t) dW_t \tag{6.2}$$

So,

$$\Delta_t = C_x(t,S_t) \tag{6.3}$$

must hold for all $t \in [0,T]$. We call this equality the delta-hedging rule.

(ii)

$$\Delta_t(\mu - r)e^{-rt}S_t dt = e^{-rt}\left(-rC(t,S_t) + C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t)\right)dt \tag{6.4}$$

So, we must have

$$\Delta_t(\mu - r)S_t = -rC(t,S_t) + C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t). \tag{6.5}$$

We may now substitute $\Delta_t = C_x(t,S_t)$. Then the left-hand side of (6.5) becomes

$$\Delta_t(\mu - r)S_t = C_x(t,S_t)(\mu - r)S_t = \mu S_t C_x(t,S_t) - rS_t C_x(t,S_t). \tag{6.6}$$

Therefore,

$$\mu S_t C_x(t,S_t) - rS_t C_x(t,S_t) = -rC(t,S_t) + C_t(t,S_t) + \mu S_t C_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t). \tag{6.7}$$

So,

$$-rS_tC_x(t,S_t) = -rC(t,S_t) + C_t(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t). \tag{6.8}$$

Rearranging gives us the Black-Scholes SDE

$$rC(t,S_t) = C_t(t,S_t) + rS_tC_x(t,S_t) + \frac{1}{2}\sigma^2 S_t^2 C_{xx}(t,S_t) \quad \text{for all} \quad t \in [0,T]. \tag{6.9}$$

The solution to the Black-Scholes stochastic differential equation is a random process satisfying the terminal condition $c(T,S_T) = (S_T - K)^+$ is given by

$$C(t,x) = x\mathcal{N}(d_+(T-t,x)) - Ke^{-r(T-t)}\mathcal{N}(d_-(T-t,x)) \quad \text{for} \quad t \in [0,T) \quad \text{and} \quad x > 0, \tag{6.10}$$

where

$$d_\pm(\tau,x) := \frac{1}{\sigma\sqrt{\tau}}\left(\log\left(\frac{x}{K}\right) + \left(r \pm \frac{\sigma^2}{2}\right)\tau\right).$$

We call this equation the Black-Scholes option pricing formula some examples are given in Figures 6.1 and 6.2. The $\mathcal{N}$ denotes the cumulative standard normal distribution

$$\mathbb{P}(Z \leq y) = \mathcal{N}(y) := \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{y} e^{-\frac{z^2}{2}}\,dz = \frac{1}{\sqrt{2\pi}}\int_{-y}^{\infty} e^{-\frac{z^2}{2}}\,dz.$$

Written in a probabilistic way, we have that the value of the option is given by

$$C(t,x) = x\mathbb{P}(Z \leq d_+(T-t,x)) - Ke^{-r(T-t)}\mathbb{P}(d_-(T-t,x)) \tag{6.11}$$

in the case that $X_t = x$ is fixed.

The complete derivation of the Black-Scholes formula can be found in *Stochastic Calculus for Finance II - Continuous-Time Models* Section 5.2.5 [29].

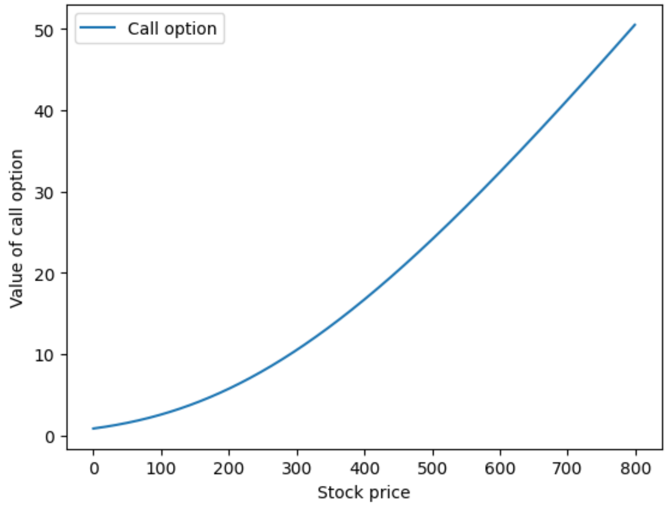

Figure 6.1: Black-Scholes option pricing formula with variable stock price S and constant strike price $K = 100$, interest rate $r = 0.1$, maturity time $T = 1$ and volatility $\sigma = 0.3$.

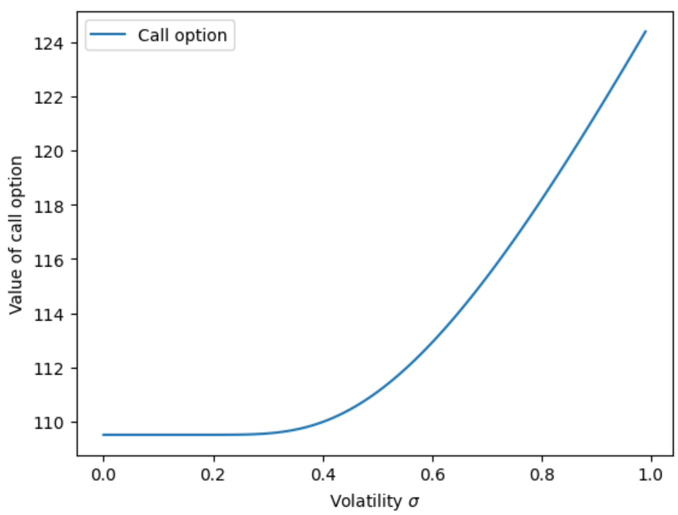

Figure 6.2: Black-Scholes option pricing formula with variable volatility σ and constant strike price $K = 100$, interest rate $r = 0.1$, maturity time $T = 1$ and stock price $S = 200$.

Lastly, we will note that the Black-Scholes model has its limitations. Namely, the Black-Scholes model is unable to adequately capture the crashes in the financial market. In the next chapter, we will see an extension of the Black-Scholes model proposed by Robert Merton. Merton extended the Black-Scholes model by incorporating a jump component in order to model those crashes. This jump component is a Poisson jump process.

Chapter 7

Merton Jump-Diffusion Model

In this chapter we will introduce the Merton jump-diffusion model, which is an extension of the Black-Scholes model in the sense that the model incorporates a jump component in order to capture the crashes in the financial market. The Merton jump-diffusion model was introduced in the nineteen-seventies by Robert Merton in his article *Option pricing when underlying stock returns are discontinuous* [17]. This chapter will rely on Mertons article [17], however changes are made in notation and calculations regarding the use of Itô-Doeblin formula are added.

The assumptions needed for the Black-Scholes model are still to be hold true throughout this chapter with the exception of the continuous stock price dynamics. It is now assumed that the stock price process follows a geometric Brownian motion with a jump component.

Assumptions:

(i) The short-term interest rate is known and constant through time.

(ii) The stock price process follows a geometric Brownian motion with constant drift and volatility.

(iii) The stock does not pay dividends.

(iv) The option is European. This means that the option can only be exercised on the maturity time T.

(v) There are no transaction costs in buying or selling the stock or the option.

(vi) One may borrow any fraction of the price of a security to buy it or to hold it at the short-term interest rate.

(vii) Short selling will not be penalized.

In the Merton jump-diffusion model it is assumed that the total change of the stock price S_t is a composition of two types of changes, namely the normal fluctuations in the price and the abnormal fluctuations. A normal fluctuation in the price may occur due to changes in the economic outlook or in a temporal imbalance of the supply and demand chain. These normal price fluctuations are modeled by a standard geometric Brownian motion with constant volatility. An abnormal price fluctuation is due to the availability of relevant and valuable information on a certain stock. This will be modeled according to a jump process, more specifically the jump component is modeled according to a Poisson distribution with parameter α. We may view the arrival of the valuable information about a certain stock as an event and we assume that these arrivals are independently and identically distributed. Then the stochastic differential equation of the stock price process

$(S_t)_{t\geq 0}$ is given by the Black-Scholes dynamics and the jump component J_t

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + J_t. \tag{7.1}$$

The jump component consist of arrivals and jump sizes. The arrivals are modeled according to a Poisson distribution with parameter α as stated before, hence $N_t \sim Pois(\alpha t)$. The jump size Y_t is log-normally distributed, so $Y_t > 0$. The J_t term can be specified as follows. Assume that the stock price equals S_t before the jump, then after the jump the stock price will be equal to $Y_t S_t$, were Y_t stands for the random variable percentage change. The change in price due to the jump may be calculated as

$$dS_t = Y_t S_t - S_t.$$

So

$$\frac{dS_t}{S_t} = Y_t - 1.$$

Then the jump component J_t may be written as $(Y_t - 1)dN_t$, hence

$$\frac{dS_t}{S_t} = \mu dt + \sigma dW_t + (Y_t - 1)dN_t. \tag{7.2}$$

Note that if a jump occurs, so $dN_t = 1$ this happens with probability αdt, then the jump component will be added to the stock price change. However, if no jump occurs, so $dN_t = 0$, which happens with probability $1 - \alpha dt$, then we will have precisely the SDE of the stock price process of the Black-Scholes model. Furthermore, the jump component introduces a drift term as well, since

$$\mathbb{E}[(Y_t - 1)dN_t] = \mathbb{E}[Y_t - 1]\mathbb{E}[dN_t] = \alpha k dt,$$

due to the independence of the jump size and arrivals and we defined $\mathbb{E}[Y_t - 1] = k$. We need to subtract this drift term from the jump component in order to obtain a pure jump process. Thus,

$$\begin{aligned} \frac{dS_t}{S_t} &= \mu dt + \sigma dW_t + (Y_t - 1)dN_t - \alpha k dt \\ &= (\mu - \alpha k)dt + \sigma dW_t + (Y_t - 1)dN_t. \end{aligned}$$

Hence, the stochastic differential equation of the stock price process $(S_t)_{t\geq 0}$ for the Merton jump-diffusion model is given by

$$dS_t = (\mu - \alpha k)S_t dt + \sigma S_t dW_t + S_t(Y_t - 1)dN_t, \tag{7.3}$$

with μ the drift term, σ the volatility, W_t a standard Brownian motion, k denotes the magnitude of the jump and N_t is a Poisson process with parameter $\alpha \geq 0$. We assume that the Brownian motion $(W_t)_{t\geq 0}$ and the jump component are independent and furthermore that the jump size and arrivals are independent. Notice that if $\alpha = 0$, then we obtain the following SDE

$$dS_t = \mu S_t dt + \sigma S_t dW_t,$$

which is precisely the SDE of the stock price process of the Black-Scholes model.

Lemma 7.0.1. If we assume that the parameters μ, α, k and σ are constant, then we may write

$$S_t = S_0 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2 - \alpha k)t + \sigma W_t\} \cdot \prod_{j=1}^{N_t} Y_j. \tag{7.4}$$

Then, (7.4) is a solution of the stochastic differential equation

$$dS_t = (\mu - \alpha k)S_t dt + \sigma S_t dW_t + S_t(Y_t - 1)dN_t.$$

Proof. In order to show that (7.4) is indeed a solution of the stochastic differential equation of $(S_t)_{t\geq 0}$, we will apply Itô-Doeblin formula. But, we will start by rewriting the SDE.

$$\begin{aligned} dS_t &= (\mu - \alpha k)S_t dt + \sigma S_t dW_t + S_t(Y_t - 1)dN_t \\ &= (\mu - \alpha k)S_t dt + \sigma S_t dW_t + \left(\prod_{j=1}^{dN_t} Y_j - 1\right) S_t \\ &= \left((\mu - \alpha k)dt + \sigma dW_t + \prod_{j=1}^{dN_t} Y_j - 1\right) S_t. \end{aligned}$$

Note that dN_t gives the number of jumps in a certain interval. If the interval is not too small multiple jumps can occur, for instance a jump with jump size Y_1 and a jump with jump size Y_2. In this case, $dN_t = 2$. Now, we divide both sides by S_t and note that the left-hand side looks like the derivative of the logarithm. Hence, we will apply Itô-Doeblin formula with jump component to $f(x) = \log(x)$. Note that $f_x(x) = \frac{1}{x}$ and $f_{xx}(x) = -\frac{1}{x^2}$. So,

$$\begin{aligned} d(\log(S_t)) = df(S_t) &= f_x(S_t)dS_t + \frac{1}{2}f_{xx}(S_t)d[S]_t + f\left(\left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - f(S_t) \\ &= \frac{1}{S_t}dS_t + \frac{1}{2} \cdot -\frac{1}{S_t^2}dS_t dS_t + \log\left(\left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - \log(S_t) \\ &= \frac{1}{S_t}dS_t + \frac{1}{2} \cdot -\frac{1}{S_t^2}dS_t dS_t + \log\left(\frac{\left(\prod_{j=1}^{dN_t} Y_j\right) S_t}{S_t}\right) \\ &= \frac{1}{S_t}dS_t + \frac{1}{2} \cdot -\frac{1}{S_t^2}dS_t dS_t + \log\left(\prod_{j=1}^{dN_t} Y_j\right) \\ &= \frac{1}{S_t}\left((\mu - \alpha k)S_t dt + \sigma S_t dW_t\right) - \frac{1}{2} \cdot \frac{1}{S_t^2}\left(\sigma^2 S_t^2 dt\right) + \log\left(\prod_{j=1}^{dN_t} Y_j\right) \\ &= (\mu - \alpha k)dt + \sigma dW_t - \frac{1}{2}\sigma^2 dt + \log\left(\prod_{j=1}^{dN_t} Y_j\right) \\ &= (\mu - \frac{1}{2}\sigma^2 - \alpha k)dt + \sigma dW_t + dN_t + \log\left(\prod_{j=1}^{dN_t} Y_j\right) \end{aligned}$$

Integration on both sides yields

$$\log(S_t) - \log(S_0) = (\mu - \frac{1}{2}\sigma^2 - \alpha k)t + \sigma W_t + \log\left(\prod_{j=1}^{N_t} Y_j\right).$$

$$\log\left(\frac{S_t}{S_0}\right) = (\mu - \frac{1}{2}\sigma^2 - \alpha k)t + \sigma W_t + \log\left(\prod_{j=1}^{N_t} Y_j\right).$$

Taking the exponent on both sides and then multiply with S_0 yields the desired statement

$$\begin{aligned}
\frac{S_t}{S_0} &= \exp\left\{(\mu - \frac{1}{2}\sigma^2 - \alpha k)t + \sigma W_t + \log\left(\prod_{j=1}^{N_t} Y_j\right)\right\} \\
S_t &= S_0 \cdot \exp\left\{(\mu - \frac{1}{2}\sigma^2 - \alpha k)t + \sigma W_t + \log\left(\prod_{j=1}^{N_t} Y_j\right)\right\} \\
&= S_0 \cdot \exp\{(\mu - \frac{1}{2}\sigma^2 - \alpha k)t + \sigma W_t\} \cdot \prod_{j=1}^{N_t} Y_j.
\end{aligned}$$

■

We will now derive the stochastic differential equation of the option price. Let the value of the option be depending on the stock price S_t and the time t, so the value of the option is denoted as $C(t,S_t)$ and is a stochastic process. Then the stochastic differential equation of the option price can be derived by applying the Itô-Doeblin formula with jump component to $C(t,x)$.

Lemma 7.0.2. The stochastic differential equation of the option price $C(t,S_t)$ is given by

$$\begin{aligned}
dC(t,S_t) = &\left(C_t(t,S_t) + \frac{1}{2}C_{xx}(t,S_t)\sigma^2 S_t^2 + C_x(t,S_t)(\mu - \alpha k)S_t\right)dt + C_x(t,S_t)\sigma S_t dW_t \\
&+ \alpha\mathbb{E}[C\left(t, \left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - C(t,S_t)] \qquad (7.5)
\end{aligned}$$

Proof. We will apply the Itô-Doeblin formula with jump component to $C(t,x)$. So,

$$\begin{aligned}
dC(t,S_t) &= C_t(t,S_t)dt + C_x(t,S_t)dS_t + \frac{1}{2}C_{xx}(t,S_t)d[S]_t + \alpha\mathbb{E}[C\left(t, \left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - C(t,S_t)] \\
&= C_t(t,S_t)dt + C_x(t,S_t)dS_t + \frac{1}{2}C_{xx}(t,S_t)dS_t dS_t + \alpha\mathbb{E}[C\left(t, \left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - C(t,S_t)] \\
&= C_t(t,S_t)dt + C_x(t,S_t)\left((\mu - \alpha k)S_t dt + \sigma S_t dW_t\right) + \frac{1}{2}C_{xx}(t,S_t)\sigma^2 S_t^2 dt \\
&\quad + \alpha\mathbb{E}[C\left(t, \left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - C(t,S_t)]
\end{aligned}$$

$$= \left(C_t(t,S_t) + C_x(t,S_t)(\mu - \alpha k)S_t + \frac{1}{2}C_{xx}(t,S_t)\sigma^2 S_t^2 \right) dt + C_x(t,S_t)\sigma S_t dW_t$$
$$+ \alpha \mathbb{E}[C\left(t, \left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - C(t,S_t)]$$

Hence, the SDE of the option price is given by

$$dC(t,S_t) = \left(C_t(t,S_t) + \frac{1}{2}C_{xx}(t,S_t)\sigma^2 S_t^2 + C_x(t,S_t)(\mu - \alpha k)S_t \right) dt + C_x(t,S_t)\sigma S_t dW_t$$
$$+ \alpha \mathbb{E}[C\left(t, \left(\prod_{j=1}^{dN_t} Y_j\right) S_t\right) - C(t,S_t)]$$

■

In Figure 7.1, we have modeled the Merton jump-diffusion model. We see that there are lots of fluctuations/jumps over the time period of one year. For instance, roughly around day 310 a large jump occurs. In Figure 7.2, the points are not connected with straight lines as in Figure 7.1 to make the jumps more visible.

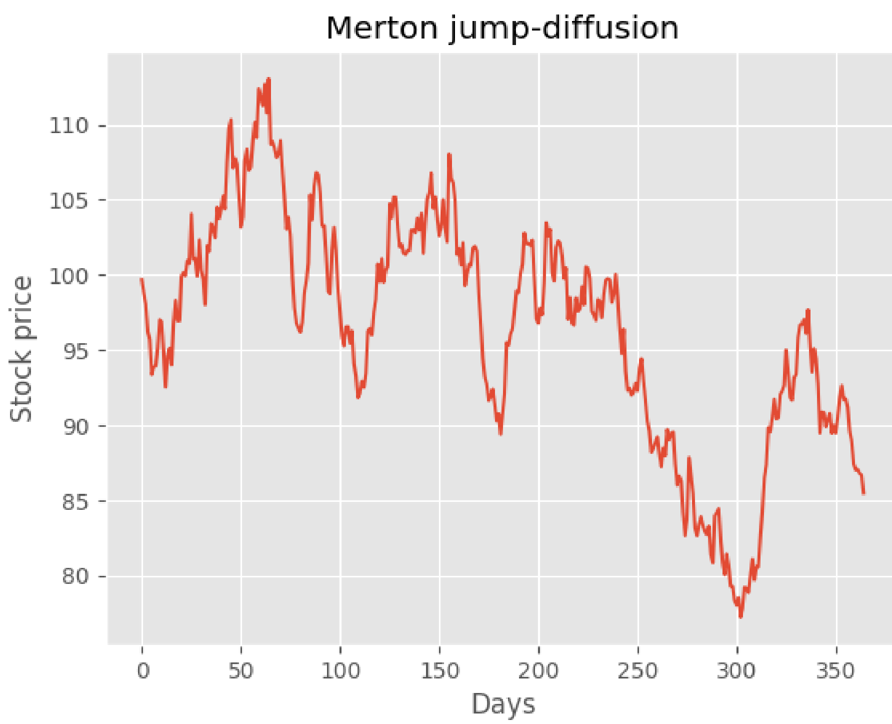

Figure 7.1: A realization of the Merton jump-diffusion model over one year based on the code given in Appendix 11.4.

In spite of all that is stated before, the Merton jump-diffusion model is incapable to account for jump propagation. Jump propagation means that if an event happens in one region of the world, then the likelihood of an event happening in other parts of the world increases, this is also known as financial contagion. In the next chapter, we will define the Hawkes jump-diffusion model, this model will account for jump propagation also known as financial contagion. The Hawkes jump-diffusion model extends the Merton jump-diffusion model by replacing the Poisson jump process by a mutually exciting Hawkes process.

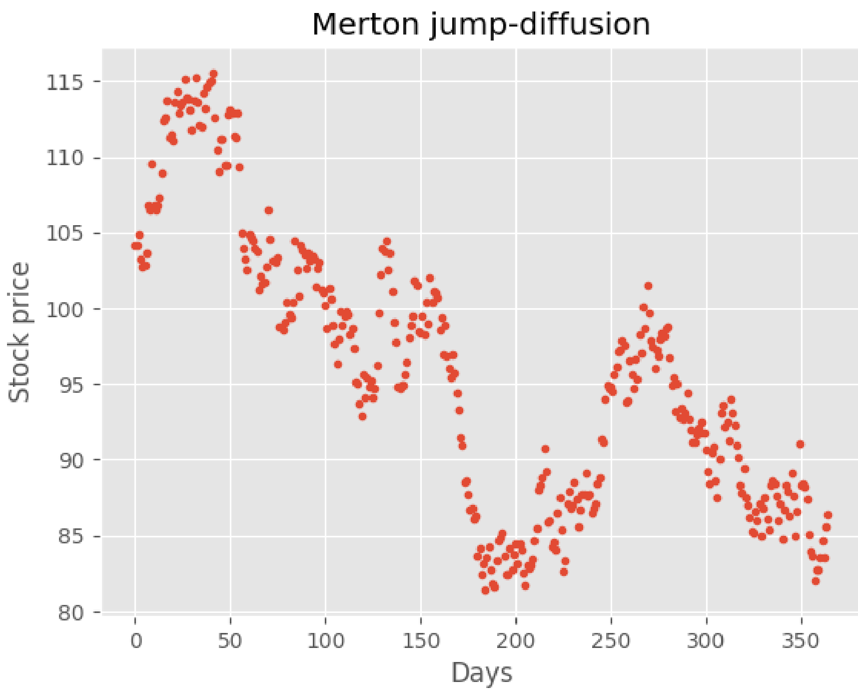

Figure 7.2: A realization of the Merton jump-diffusion model over one year.

Chapter 8

Hawkes Jump-Diffusion Model

8.1 Hawkes Jump-Diffusion Model

This section is largely based on the article of Y. Ait-Sahalia [1]. We will introduce the Hawkes jump-diffusion model. The Hawkes jump-diffusion model extends the Merton jump-diffusion model by replacing the Poisson jump process by a mutually exciting Hawkes process, which allows for financial contagion. As we have seen, the Merton jump-diffusion model extended, in his turn, the Black-Scholes model by incorporating a jump component in order to model the crashes in the market. This was the case because the Black-Scholes model is unable to adequately capture those crashes.

We will start with stating the definition of the Hawkes jump-diffusion model without any explanation. The rest of this section will be devoted to deriving the dynamics stated in the definition of the Hawkes jump-diffusion model.

Definition 8.1.1 (Hawkes jump-diffusion model)**.** The Hawkes jump-diffusion model is defined as

$$\begin{cases} dX_{i,t} = \mu_i dt + \sqrt{V_{i,t}} dW^X_{i,t} + Z_{i,t} dN_{i,t}, & i = 1,...,n \\ dV_{i,t} = \kappa_i(\theta_i - V_{i,t})dt + \eta_i \sqrt{V_{i,t}} dW^V_{i,t}, & i = 1,...,n \\ d\lambda_i(t) = -\beta_i (\nu_i - \lambda_i(t))\, dt + \sum_{j=1}^n \alpha_{i,j} dN_{j,t}, & i = 1,...,n. \end{cases} \tag{8.1}$$

with $X_{i,t}$ the asset log-returns, $V_{i,t}$ the instantaneous variance and $\lambda_i(\cdot)$ the exponentially decaying intensities.

Let us consider the asset log-returns $X_{i,t}$ defined as

$$dX_{i,t} = \mu_i dt + \sigma_i dW_{i,t} + Z_{i,t} dN_{i,t}, \quad i = 1,...,n, \tag{8.2}$$

with μ_i the drift term, σ_i the volatility, $W_t := (W_{1,t},...,W_{n,t})^T$ the n-dimensional vector of standard Brownian motions, $Z_t := (Z_{1,t},...,Z_{n,t})^T$ the n-dimensional vector of jump sizes that are independently distributed with distributions F_{Z_i} and $N_t := (N_{1,t},...,N_{n,t})^T$ the n-dimensional vector of Hawkes processes, which is the jump process. The n-dimensional vector of Hawkes processes is the mutually exciting Hawkes process defined in Definition 4.1.1 with conditional intensity functions

$$\lambda_i(t) = \nu_i + \sum_{j=1}^n \int_0^t h_{i,j}(t-s) dN_{j,s}, \quad i = 1,...,n.$$

In the described model (8.2), we assume that the drift term μ_i and the volatility σ_i are constant parameters. Furthermore, we assume that the vectors W_t, Z_t and N_t are mutually independent. This model can be extended by allowing for stochastic volatility, which is known as the Heston model:

$$dX_{i,t} = \mu_i dt + \sqrt{V_{i,t}} dW^X_{i,t} + Z_{i,t} dN_{i,t}, \quad i = 1, ..., n, \tag{8.3}$$

with $V_{i,t}$ the instantaneous variance that follows the following dynamics

$$dV_{i,t} = \kappa_i(\theta_i - V_{i,t})dt + \eta_i \sqrt{V_{i,t}} dW^V_{i,t},$$

where κ_i, θ_i and η_i are assumed to be constant parameters.

The manageability of the jump part of the extended model and thus the possibility of estimating the jump component depends on the parametrization of the conditional intensities $\lambda_i(\cdot)$. Therefore, we will choose the excitation function $h_{i,j}(\cdot)$ to be exponential, so

$$h_{i,j}(t) = \alpha_{i,j} e^{-\beta_i t},$$

with $\alpha_{i,j} \geq 0$ and $\beta_i > 0$ for all $i, j = 1, ..., n$. This excitation function can be interpret as follows. Whenever a jump occurs in the asset prices, the conditional intensities $\lambda_i(\cdot)$ will increase by $\alpha_{i,j}$ and then decays exponentially at rate β_i back towards ν_i.

We may write the exponentially decaying intensities $\lambda_i(\cdot)$ as follows

$$\begin{aligned}\lambda_i(t) &= \nu_i + \sum_{j=1}^n \int_0^t \alpha_{i,j} e^{-\beta_i(t-s)} dN_{j,s} \\ &= \nu_i + \sum_{j=1}^n \alpha_{i,j} e^{-\beta_i t} \int_0^t e^{\beta_i s} dN_{j,s}, \quad i = 1, ..., n.\end{aligned}$$

Lemma 8.1.1. The exponentially decaying intensity function $\lambda_i(\cdot)$ satisfies the following stochastic differential equation

$$d\lambda_i(t) = -\beta_i (\nu_i - \lambda_i(t)) dt + \sum_{j=1}^n \alpha_{i,j} dN_{j,t}, \quad i = 1, ..., n. \tag{8.4}$$

Proof. The exponentially decaying intensity function $\lambda_i(\cdot)$ can be written as

$$\lambda_i(t) = \nu_i + \sum_{j=1}^n \alpha_{i,j} e^{-\beta_i t} \int_0^t e^{\beta_i s} dN_{j,s}, \quad i = 1, ..., n.$$

Taking the derivatives on both sides yields

$$\begin{aligned}d\lambda_i(t) &= \left(\sum_{j=1}^n -\alpha_{i,j} \beta_i e^{-\beta_i t} \int_0^t e^{\beta_i s} dN_{j,s} \right) dt + \sum_{j=1}^n \alpha_{i,j} e^{-\beta_i t} \cdot e^{\beta_i t} dN_{j,t} \\ &= \left(-\beta_i \sum_{j=1}^n \int_0^t \alpha_{i,j} e^{-\beta_i(t-s)} dN_{j,s} \right) dt + \sum_{j=1}^n \alpha_{i,j} dN_{j,t} \\ &= \left(-\beta_i \left(\sum_{j=1}^n \int_0^t \alpha_{i,j} e^{-\beta_i(t-s)} dN_{j,s} + \nu_i - \nu_i \right) \right) dt + \sum_{j=1}^n \alpha_{i,j} dN_{j,t}\end{aligned}$$

$$= \left(-\beta_i \left(\lambda_i(t) - \nu_i\right)\right) dt + \sum_{j=1}^{n} \alpha_{i,j} dN_{j,t}$$
$$= -\beta_i \left(\nu_i - \lambda_i(t)\right) dt + \sum_{j=1}^{n} \alpha_{i,j} dN_{j,t}, \quad i = 1, ..., n.$$

Hence, $d\lambda_i(t) = -\beta_i \left(\nu_i - \lambda_i(t)\right) dt + \sum_{j=1}^{n} \alpha_{i,j} dN_{j,t}, \quad i = 1, ..., n.$

■

We note that the jump component of the model is able to provide for jump clustering and jump propagation. Jump clustering means that the jumps are more or less concentrated in short periods of time. When an event happens in one region of the world and that results in an increased likelihood of an event happening in other regions of the world, then we call that jump propagation. This is also known as financial contagion, where an event might stand for the booms and crashes in the stock market.

Putting the three stochastic differential equations together gives us the Hawkes jump-diffusion model. Hence, the Hawkes jump-diffusion model is defined as

$$\begin{cases} dX_{i,t} = \mu_i dt + \sqrt{V_{i,t}} dW_{i,t}^X + Z_{i,t} dN_{i,t}, & i = 1, ..., n \\ dV_{i,t} = \kappa_i (\theta_i - V_{i,t}) dt + \eta_i \sqrt{V_{i,t}} dW_{i,t}^V, & i = 1, ..., n \\ d\lambda_i(t) = -\beta_i \left(\nu_i - \lambda_i(t)\right) dt + \sum_{j=1}^{n} \alpha_{i,j} dN_{j,t}, & i = 1, ..., n. \end{cases} \tag{8.5}$$

The univariate model is given by

$$\begin{cases} dX_t = \mu dt + \sqrt{V_t} dW_t^X + Z_t dN_t \\ dV_t = \kappa (\theta - V_t) dt + \eta \sqrt{V_t} dW_t^V \\ d\lambda(t) = -\beta \left(\nu - \lambda(t)\right) dt + \alpha dN_t. \end{cases} \tag{8.6}$$

Figure 8.1 gives a realization of a self-exciting/linear Hawkes process with corresponding conditional intensity and stock price. The time will be on the horizontal axis and on the vertical axis we have starting from the top, the cumulative number of arrivals, the conditional intensity function and lastly the stock price.

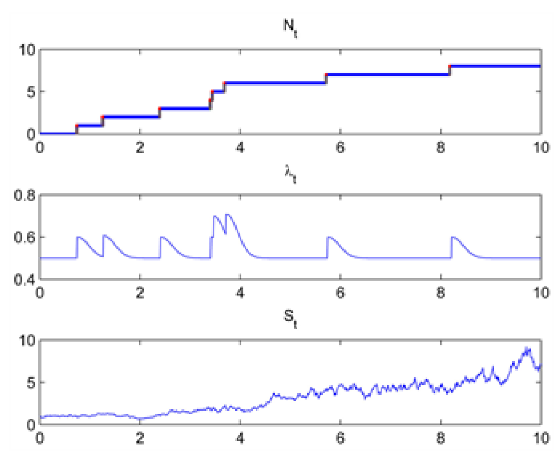

Figure 8.1: A realization of the Hawkes process N_t with corresponding conditional intensity function $\lambda(t)$ given on the top and middle. On the bottom, a graph of the corresponding stock price S_t is shown. This figure is taken from [4].

In the next section, we will prove the Law of Large Numbers and the Central Limit Theorem for the univariate case, where we do not allow for stochastic volatility.

8.2 Limit Theorems

This section is largely based on the article of Y. Seol [26] with the exception of the proof of Theorem 8.2.2 and examples 8.2.3 and 8.2.4. Furthermore, there are details added in the proof of Theorem 8.2.1 and mistakes corrected in both Theorems, 8.2.1 and 8.2.2.

Let us consider the log-stock price process X_t defined as

$$X_t = \alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i,$$

with α the drift term, β the volatility, W_t the standard Brownian motion as defined in Definition 2.0.1, Y_i are independent and identically distributed (i.i.d.) $\mathbb{R}$-valued random variables and $\sum_{i=1}^{N_t} Y_i$ is a compound Hawkes process, where N_t is defined as a linear Hawkes process with conditional intensity function given by

$$\lambda(t) := \nu + \int_0^t h(t-s)dN_s.$$

Note that N_t and Y_i are independent. We will prove that under certain assumptions the Law of Large Numbers (LLN) and the Central Limit Theorem (CLT) holds for this specific log-stock price process X_t. Before we prove the LLN and the CLT, we will define the assumptions and recall the LLN and the CLT in the case of a linear Hawkes process and a marked Hawkes process.

Assumptions:

(i) The random variables Y_i have finite expectation and variance, hence $Var(Y_1) < \infty$.

(ii) The conditional intensity function is linear and increasing, so $\lambda(x) = \nu + x$ for some strictly positive ν.

(iii) It holds that the L^1-norm of the excitation function h is less than one, hence $||h||_{L^1} < 1$ with $||h||_{L^1} = \int_0^\infty h(t)dt < \infty$.

The second assumption assures that the Hawkes process has a nice immigration-birth representation, whereas the first and third assumption makes sure that the limits are well-defined and makes sure that the Hawkes process does not explode.

Recall that in the linear Hawkes process, the Law of Large Numbers and the Central Limit Theorem were given by

$$\frac{N_t}{t} \xrightarrow{\mathbb{P}} \frac{\nu}{1-||h||_{L^1}} \quad \text{as} \quad t \to \infty$$

and

$$\sqrt{t}\left(\frac{N_t}{t} - \frac{\nu}{1-||h||_{L^1}}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu||h||_{L^1}}{(1-||h||_{L^1})^3}\right) \quad \text{as} \quad t \to \infty,$$

respectively. And in the marked Hawkes process the LLN and CLT were defined as

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{\nu}{1-\mathbb{E}[H(\xi)]} \quad \text{as} \quad t \to \infty$$

and

$$\sqrt{t}\left(\frac{N_t}{t} - \frac{\nu}{1-\mathbb{E}[H(\xi)]}\right) \rightsquigarrow \mathcal{N}\left(0, \frac{\nu(1+Var(H(\xi)))}{(1-\mathbb{E}[H(\xi)])^3}\right) \quad \text{as} \quad t \to \infty.$$

We will now state the Law of Large Numbers and the Central Limit Theorem for the Hawkes jump-diffusion model.

Theorem 8.2.1 (Law of Large Numbers for Hawkes jump-diffusion model)**.** Assume that the provided assumptions are satisfied and let X_t be the log-stock price process defined as

$$X_t = \alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i,$$

then we have that

$$\frac{X_t}{t} \xrightarrow{\mathbb{P}} \alpha + \mu \mathbb{E}[Y_1] \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mu := \frac{\nu}{1 - ||h||_{L^1}}.$$

Proof. Let us consider

$$\frac{X_t}{t} = \alpha + \beta \frac{W_t}{t} + \frac{1}{t} \sum_{i=1}^{N_t} Y_i.$$

We will look at the terms one by one. Starting with α. We have that $\alpha \xrightarrow{\mathbb{P}} \alpha$, since α is a constant. By Lemma 2.0.2, we have that $\lim_{t\to\infty} \frac{W_t}{t} = 0$ almost surely. Now, we will look at $\sum_{i=1}^{N_t} Y_i$. By Wald's equation, we have that

$$\begin{aligned}
\mathbb{E}[\sum_{i=1}^{N_t} Y_i] &= \sum_{k=0}^{\infty} \mathbb{E}[\sum_{i=1}^{k} Y_i | N_t = k] \mathbb{P}(N_t = k) = \sum_{k=0}^{\infty} \mathbb{E}[\sum_{i=1}^{k} Y_i] \mathbb{P}(N_t = k) \\
&= \sum_{k=0}^{\infty} \sum_{i=1}^{k} \mathbb{E}[Y_i] \mathbb{P}(N_t = k) \quad \text{due to finite expectation and linearity} \\
&= \sum_{k=0}^{\infty} \sum_{i=1}^{k} \mathbb{E}[Y_1] \mathbb{P}(N_t = k) \quad \text{since } Y_i \text{ are i.i.d.} \\
&= \sum_{k=0}^{\infty} k \cdot \mathbb{E}[Y_1] \mathbb{P}(N_t = k) = \mathbb{E}[Y_1] \sum_{k=0}^{\infty} k \cdot \mathbb{P}(N_t = k) \\
&= \mathbb{E}[Y_1] \mathbb{E}[N_t]
\end{aligned}$$

So, we have that $\mathbb{E}[\sum_{i=1}^{N_t} Y_i] = \mathbb{E}[Y_1]\mathbb{E}[N_t]$. Furthermore, due to assumption part (ii) we have that N_t is a linear Hawkes process. Therefore, by the LLN for linear Hawkes processes 3.3.1, we obtain that

$$\frac{N_t}{t} \xrightarrow{a.s.} \frac{\nu}{1 - ||h||_{L^1}} \quad \text{as} \quad t \to \infty.$$

Hence, we have that

$$\frac{1}{t} \sum_{i=1}^{N_t} Y_i \xrightarrow{\mathbb{P}} \frac{\nu}{1 - ||h||_{L^1}} \mathbb{E}[Y_1] = \mu \mathbb{E}[Y_1] \quad \text{as} \quad t \to \infty.$$

Using Lemma 1.1.4, we obtain that

$$\frac{X_t}{t} \xrightarrow{\mathbb{P}} \alpha + \mu \mathbb{E}[Y_1] \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mu := \frac{\nu}{1 - ||h||_{L^1}}.$$

■

Theorem 8.2.2 (Central Limit Theorem for Hawkes jump-diffusion model)**.** Assume that the provided assumptions are satisfied and let X_t be the log-stock price process defined as

$$X_t = \alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i,$$

then we have that

$$\frac{X_t - (\alpha + \mu \mathbb{E}[Y_1])t}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0, \beta^2 + \mu Var(Y_1) + (\mathbb{E}[Y_1])^2 \sigma^2) \quad \text{as} \quad t \to \infty$$

with $\mu := \frac{\nu}{1 - ||h||_{L^1}}$ and $\sigma^2 := \frac{\nu}{(1 - ||h||_{L^1})^3}$.

Proof. Let us consider

$$\frac{X_t}{t} = \alpha + \beta \frac{W_t}{t} + \frac{1}{t} \sum_{i=1}^{N_t} Y_i.$$

By Lemma 8.2.1, we obtained that

$$\frac{X_t}{t} \xrightarrow{\mathbb{P}} \alpha + \mu \mathbb{E}[Y_1] \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mu := \frac{\nu}{1 - ||h||_{L^1}}.$$

So:

$$\begin{aligned}
\sqrt{t}\left(\frac{X_t}{t} - (\alpha + \mu \mathbb{E}[Y_1])\right) &= \sqrt{t}\left(\frac{X_t - (\alpha + \mu \mathbb{E}[Y_1])t}{t}\right) \\
&= \frac{X_t - (\alpha + \mu \mathbb{E}[Y_1])t}{\sqrt{t}}
\end{aligned}$$

Plugging in $X_t = \alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i$ gives

$$\begin{aligned}
\frac{\alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i - (\alpha + \mu \mathbb{E}[Y_1])t}{\sqrt{t}} &= \frac{\alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i - \alpha t - \mu \mathbb{E}[Y_1]t}{\sqrt{t}} \\
&= \frac{\beta W_t + \sum_{i=1}^{N_t} Y_i - \mu \mathbb{E}[Y_1]t}{\sqrt{t}} \\
&= \frac{\beta W_t}{\sqrt{t}} + \frac{\sum_{i=1}^{N_t} Y_i - \mu \mathbb{E}[Y_1]t}{\sqrt{t}}
\end{aligned}$$

First, we look at $\frac{\beta W_t}{\sqrt{t}}$. We will show, using characteristic functions, that $\frac{\beta W_t}{\sqrt{t}}$ is normally distributed with mean 0 and variance β^2. Remember that $W_t \sim \mathcal{N}(0,t)$ and thus $\frac{W_t}{\sqrt{t}} \sim \mathcal{N}(0,1)$. Define $Z := \frac{W_t}{\sqrt{t}}$ with Z the standard normal. The corresponding characteristic function is

$$\varphi_Z(\theta) = e^{-\frac{1}{2}\theta^2}.$$

So, we obtain:

$$\varphi(\theta) = \mathbb{E}[e^{i\theta \frac{\beta W_t}{\sqrt{t}}}] = \mathbb{E}[e^{i\theta \beta Z}] = e^{-\frac{1}{2}\theta^2 \beta^2}.$$

Therefore, $\frac{\beta W_t}{\sqrt{t}} \sim \mathcal{N}(0, \beta^2)$. Now, we will look at the second term, $\frac{\sum_{i=1}^{N_t} Y_i - \mu \mathbb{E}[Y_1]t}{\sqrt{t}}$. First, we will rewrite the numerator.

$$\begin{aligned}\sum_{i=1}^{N_t} Y_i - \mu\mathbb{E}[Y_1]t &= \sum_{i=1}^{N_t} Y_i + \mathbb{E}[Y_1](\mathbb{E}[N_t] - \mu t - \mathbb{E}[N_t]) \\ &= \sum_{i=1}^{N_t} Y_i + \mathbb{E}[Y_1](\mathbb{E}[N_t] - \mu t) - \mathbb{E}[N_t]\mathbb{E}[Y_1]\end{aligned}$$

So:

$$\frac{\sum_{i=1}^{N_t} Y_i + \mathbb{E}[Y_1](\mathbb{E}[N_t] - \mu t) - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} = \frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} + \frac{\mathbb{E}[Y_1](\mathbb{E}[N_t] - \mu t)}{\sqrt{t}}$$

We have that

$$\frac{\mathbb{E}[Y_1](\mathbb{E}[N_t] - \mu t)}{\sqrt{t}} \to 0 \quad \text{as} \quad t \to \infty,$$

since $\frac{N_t}{t} \xrightarrow{\mathbb{P}} \mu$ as $t \to \infty$ by the LLN for linear Hawkes processes 3.3.1 and therefore the difference $\mathbb{E}[N_t] - \mu t$ will go to zero as $t \to \infty$. Furthermore, note that $\mathbb{E}[Y_1]$ is also deterministic. So, the entire expression will go to zero in the limit.

Now, we look at $\frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}}$. We will show that

$$\frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0, \mu Var(Y_1) + (\mathbb{E}[Y_1])^2\sigma^2) \quad \text{as} \quad t \to \infty.$$

Let $\tilde{X}_n := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - \mathbb{E}[Y_1])$ and let $\tilde{Y}_t := \mathbb{E}[Y_1]\left(\frac{N_t - \mathbb{E}[N_t]}{\sqrt{t}}\right)$.

By assumption, $(\tilde{X}_n)_{n\in\mathbb{N}/\{0\}}$ and $(\tilde{Y}_t)_{t>0}$ are independent.

We may write $\frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} = \sqrt{\frac{N_t}{t}}\tilde{X}_{N_t} + \tilde{Y}_t$, since

$$\begin{aligned}\sqrt{\frac{N_t}{t}}\tilde{X}_{N_t} + \tilde{Y}_t &= \sqrt{\frac{N_t}{t}} \cdot \frac{1}{\sqrt{N_t}} \sum_{i=1}^{N_t} (Y_i - \mathbb{E}[Y_1]) + \mathbb{E}[Y_1]\left(\frac{N_t - \mathbb{E}[N_t]}{\sqrt{t}}\right) \\ &= \frac{1}{\sqrt{t}} \sum_{i=1}^{N_t} (Y_i - \mathbb{E}[Y_1]) + \frac{\mathbb{E}[Y_1](N_t - \mathbb{E}[N_t])}{\sqrt{t}} \\ &= \frac{\sum_{i=1}^{N_t} Y_i - N_t\mathbb{E}[Y_1]}{\sqrt{t}} + \frac{N_t\mathbb{E}[Y_1] - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} \\ &= \frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}}.\end{aligned}$$

By the classical CLT 1.1.8, we have that

$$\tilde{X}_n \rightsquigarrow \tilde{X} \quad \text{as} \quad n \to \infty \quad \text{with} \quad \tilde{X} \sim \mathcal{N}(0, Var(Y_1))$$

Under certain conditions, we have that

$$\tilde{Y}^* := \frac{N_t - \mathbb{E}[N_t]}{\sqrt{t}} \rightsquigarrow \tilde{Y}^* \quad \text{as} \quad t \to \infty \quad \text{with} \quad \tilde{Y}^* \sim \mathcal{N}(0, \sigma^2).$$

Those specific conditions can be read in ([10], Theorem 2.2, p. 40). So by Slutsky 1.1.5, we obtain

$$\tilde{Y}_t \rightsquigarrow \tilde{Y} \quad \text{as} \quad t \to \infty \quad \text{with} \quad \tilde{Y} \sim \mathcal{N}(0, (\mathbb{E}[Y_1])^2\sigma^2),$$

since $\mathbb{E}[Y_1]$ is just a constant.
By the LLN for linear Hawkes processes, we obtain that $\frac{N_t}{t} \xrightarrow{\mathbb{P}} \mu$ as $t \to \infty$. So by the Continuous Mapping Theorem 1.1.3, we have that

$$\sqrt{\frac{N_t}{t}} \xrightarrow{\mathbb{P}} \sqrt{\mu} \quad \text{as} \quad t \to \infty.$$

Then by Slutsky 1.1.5 and the version of Continuous Mapping Theorem for compositions, we have that

$$\sqrt{\frac{N_t}{t}} \tilde{X}_{N_t} \rightsquigarrow \mathcal{N}(0, \mu Var(Y_1)) \quad \text{as} \quad t \to \infty.$$

Due to independence, we can apply Lemma 1.1.2. Hence, we obtain the joint convergence

$$\left(\sqrt{\frac{N_t}{t}} \tilde{X}_{N_t}, \tilde{Y}_t\right) \rightsquigarrow \left(\tilde{X}^*, \tilde{Y}\right) \quad \text{as} \quad t \to \infty \quad \text{with} \quad \tilde{X}^* \sim \mathcal{N}(0, \mu Var(Y_1)) \quad \text{and} \quad \tilde{Y} \sim \mathcal{N}(0, (\mathbb{E}[Y_1])^2 \sigma^2).$$

Then, again by Slutsky, we have

$$\sqrt{\frac{N_t}{t}} \tilde{X}_{N_t} + \tilde{Y}_t \rightsquigarrow \mathcal{N}(0, \mu Var(Y_1) + (\mathbb{E}[Y_1])^2 \sigma^2) \quad \text{as} \quad t \to \infty.$$

So

$$\frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0, \mu Var(Y_1) + (\mathbb{E}[Y_1])^2 \sigma^2) \quad \text{as} \quad t \to \infty,$$

with $\mu := \frac{\nu}{1 - ||h||_{L^1}}$ and $\sigma^2 := \frac{\nu}{(1 - ||h||_{L^1})^3}$. We apply Slutksy once more to obtain the desired result. Hence,

$$\begin{aligned} \frac{X_t - (\alpha + \mu \mathbb{E}[Y_1])t}{\sqrt{t}} &= \frac{\beta W_t}{\sqrt{t}} + \frac{\sum_{i=1}^{N_t} Y_i - \mathbb{E}[N_t]\mathbb{E}[Y_1]}{\sqrt{t}} + \frac{\mathbb{E}[Y_1](\mathbb{E}[N_t] - \mu t)}{\sqrt{t}} \\ &\rightsquigarrow \mathcal{N}(0, \beta^2 + \mu Var(Y_1) + (\mathbb{E}[Y_1])^2 \sigma^2) \quad \text{as} \quad t \to \infty \end{aligned}$$

with $\mu := \frac{\nu}{1 - ||h||_{L^1}}$ and $\sigma^2 := \frac{\nu}{(1 - ||h||_{L^1})^3}$.

■

Example 8.2.3 (Normal distribution)**.** Let us recall that the log-stock price process X_t is defined as

$$X_t = \alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i.$$

Suppose now that $Y_1 \sim \mathcal{N}(0,1)$. Hence, we have $\mathbb{E}[Y_1] = 0$ and $Var(Y_1) = 1$. By the Law of Large Numbers (8.2.1), we obtained

$$\frac{X_t}{t} \xrightarrow{\mathbb{P}} \alpha + \mu \mathbb{E}[Y_1] \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mu := \frac{\nu}{1 - ||h||_{L^1}}.$$

In this case,

$$\frac{X_t}{t} \xrightarrow{\mathbb{P}} \alpha \quad \text{as} \quad t \to \infty.$$

The Central Limit Theorem (8.2.2) is in general given by

$$\frac{X_t - (\alpha + \mu \mathbb{E}[Y_1])t}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0, \beta^2 + \mu Var(Y_1) + (\mathbb{E}[Y_1])^2 \sigma^2) \quad \text{as} \quad t \to \infty$$

with $\mu := \frac{\nu}{1-||h||_{L^1}}$ and $\sigma^2 := \frac{\nu}{(1-||h||_{L^1})^3}$.
In this case, it will be

$$\frac{X_t - \alpha t}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0, \beta^2 + \mu) \quad \text{as} \quad t \to \infty.$$

Now, we will take the excitation function $h(\cdot)$ to be exponential, so $h(t) = \alpha e^{-\beta t}$ with $\alpha, \beta > 0$ some constants. Then:

$$\begin{aligned} ||h||_{L^1} &= \int_0^\infty h(t)dt = \int_0^\infty \alpha e^{-\beta t} dt = \alpha \int_0^\infty e^{-\beta t} dt \\ &= \alpha \left[-\frac{1}{\beta} e^{-\beta t} \right]_{t=0}^{t=\infty} = \alpha(0 + \frac{1}{\beta}) = \frac{\alpha}{\beta}. \end{aligned}$$

So

$$\begin{aligned} \mu &:= \frac{\nu}{1-||h||_{L^1}} = \frac{\nu}{1-\frac{\alpha}{\beta}} = \frac{\nu\beta}{\beta - \alpha} \quad \text{if} \quad \alpha \neq \beta \quad \text{and} \\ \sigma^2 &:= \frac{\nu}{(1-||h||_{L^1})^3} = \frac{\nu}{(1-\frac{\alpha}{\beta})^3} = \frac{\nu\beta^3}{(\beta-\alpha)^3} \quad \text{if} \quad \alpha \neq \beta. \end{aligned}$$

♦

Example 8.2.4 (Exponential distribution)**.** Let us recall that the log-stock price process X_t is defined as

$$X_t = \alpha t + \beta W_t + \sum_{i=1}^{N_t} Y_i.$$

Suppose $Y_1 \sim Exp(\lambda)$. The probability density function is given by

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if} \quad x \geq 0 \\ 0 & \text{if} \quad x < 0 \end{cases}.$$

Then the mean equals

$$\begin{aligned} \mathbb{E}[Y_1] &= \int_0^\infty x f(x) dx = \int_0^\infty x \lambda e^{-\lambda x} dx = \lambda \int_0^\infty x e^{-\lambda x} dx \\ &= \lambda \left(\left[-\frac{x}{\lambda} e^{-\lambda x} \right]_{x=0}^{x=\infty} - \int_0^\infty -\frac{1}{\lambda} e^{-\lambda x} dx \right) \quad \text{partial integration} \\ &= \lambda \left(0 + \frac{1}{\lambda} \int_0^\infty e^{-\lambda x} dx \right) = \lambda \left(0 + \frac{1}{\lambda} \left[-\frac{1}{\lambda} e^{-\lambda x} \right]_{x=0}^{x=\infty} \right) \\ &= \lambda \left(\frac{1}{\lambda}(0 + \frac{1}{\lambda}) \right) = \frac{1}{\lambda}. \end{aligned}$$

The second moment equals

$$\begin{aligned} \mathbb{E}[Y_1^2] &= \int_0^\infty x^2 f(x) dx = \int_0^\infty x \lambda e^{-\lambda x} dx = \lambda \int_0^\infty x^2 e^{-\lambda x} dx \\ &= \frac{2}{\lambda^2} \quad \text{by applying two times partial integration.} \end{aligned}$$

Then the variance is given by

$$Var(Y_1) = \mathbb{E}[Y_1^2] - (\mathbb{E}[Y_1])^2 = \frac{2}{\lambda^2} - \left(\frac{1}{\lambda} \right)^2 = \frac{1}{\lambda^2}.$$

Now, we can apply the Law of Large Numbers (8.2.1) and the Central Limit Theorem (8.2.2) to obtain

$$\frac{X_t}{t} \xrightarrow{\mathbb{P}} \alpha + \frac{\mu}{\lambda} \quad \text{as} \quad t \to \infty \quad \text{with} \quad \mu := \frac{\nu}{1 - ||h||_{L^1}}$$

and

$$\frac{X_t - (\alpha + \frac{\mu}{\lambda})t}{\sqrt{t}} \rightsquigarrow \mathcal{N}(0, \beta^2 + \frac{\mu}{\lambda^2} + \frac{\sigma^2}{\lambda^2}) \quad \text{as} \quad t \to \infty$$

with $\mu := \frac{\nu}{1-||h||_{L^1}}$ and $\sigma^2 := \frac{\nu}{(1-||h||_{L^1})^3}$.

Notice that if λ is a large number, the variance is mostly dependent on β, since $\frac{\mu}{\lambda^2} + \frac{\sigma^2}{\lambda^2}$ will be very small and might tend to zero. On the other hand, if λ is very small, then the variance will be quite large. This implies that the spread will be very wide.

To be more concrete, we will take the excitation function $h(\cdot)$ to be the power law function, so $h(t) = \frac{k}{(c+t)^p}$ with $k, c, p > 0$ some constants. Let $p = 2$, $k = 1$ and $c = 2$, then $h(t) = \frac{1}{(2+t)^2}$. So:

$$||h||_{L^1} = \int_0^\infty h(t)dt = \int_0^\infty \frac{1}{(2+t)^2} dt = \left[-\frac{1}{2+t} \right]_{t=0}^{t=\infty} = (0 + \frac{1}{2}) = \frac{1}{2}.$$

Hence,

$$\mu := \frac{\nu}{1 - ||h||_{L^1}} = \frac{\nu}{1 - \frac{1}{2}} = 2\nu \quad \text{and}$$
$$\sigma^2 := \frac{\nu}{(1 - ||h||_{L^1})^3} = \frac{\nu}{(1 - \frac{1}{2})^3} = 8\nu.$$

♦

Chapter 9

Numerical Study

In order to simulate a Hawkes process we need to use a Python package called "hawkes". This package provides us with the tools to simulate the conditional intensity functions. The Python code relies on the tutorial of Takahiro Omi [20] although changes are made in the code especially in the code of the power law examples. Below one finds an example were we simulate the exponentially decaying intensity with parameters $\alpha = 1.0$, $\beta = 1.1$ and the background intensity equal to 0.1. Please note that we use λ instead of μ as the background intensity.

```
!pip install hawkes
import Hawkes as hk
import numpy as np
from matplotlib import pyplot as plt

#Model set up for conditional intensity function
model = hk.simulator()
model.set_kernel('exp') #Exponentially decaying intensity
model.set_baseline('const') #Background intensity (constant)
parameters = {'mu':..., 'alpha':..., 'beta':...} #Given set of parameter values
model.set_parameter(parameters)

#Simulation
parameters = {'mu':0.1, 'alpha':1.0, 'beta':1.1}
interval = [0,100]
model = hk.simulator().set_kernel('exp').set_baseline('const').set_parameter(
                                          parameters)
T = model.simulate(interval)
model.plot_N() #Plot of time and the number of arrivals
model.plot_l() #Plot of time and conditional intensity function
```

A realization of a Hawkes process with exponentially decaying intensity, meaning the excitation function is exponential, is shown in figure 9.1. In this example, we have taken the excitation function $h(\cdot)$ to be

$$h(t) = 1.0e^{-1.1t}$$

and the exponentially decaying intensity $\lambda(\cdot)$ is then

$$\lambda(t) = 0.1 + \int_0^t 1.0e^{-1.1(t-s)} dN_s = 0.1 + \sum_{t_i < t} 1.0e^{-1.1(t-t_i)}.$$

An arrival during the time interval $[0, 100]$ will increases the intensity by $\alpha = 1.0$ and then decays exponentially at rate $\beta = 1.1$ towards the background intensity $\nu = 0.1$. Note that the lines on top are the realizations of the arrival times, so at those times an event/jump occurs.

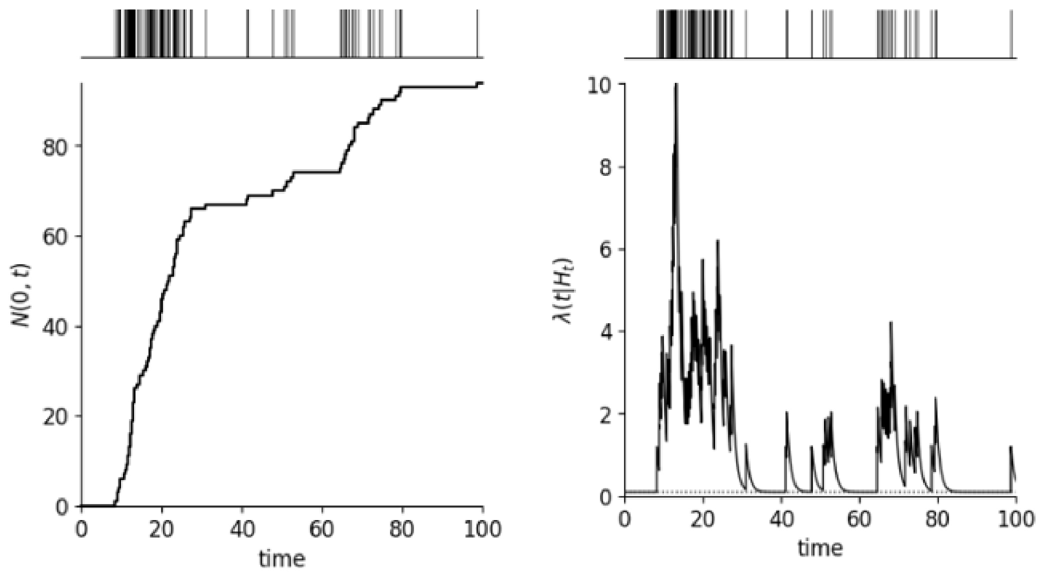

Figure 9.1: Realization Hawkes process (left) with corresponding exponentially decaying intensity (right).

As an other example, we can take the excitation function to be the power law function with parameters $k = 0.8$, $p = 1.5$ and $c = 5.0$ and constant background intensity equal to 0.1. The realization of a Hawkes process with this particular excitation function is given in figure 9.2. So, in this case the excitation function $h(\cdot)$ is chosen to be

$$h(t) = \frac{0.8}{(5.0+t)^{1.5}}$$

and then the corresponding conditional intensity is

$$\lambda(t) = 0.1 + \int_0^t \frac{0.8}{(5.0+(t-s))^{1.5}} dN_s = 0.1 + \sum_{t_i<t} \frac{0.8}{(5.0+(t-t_i))^{1.5}}.$$

Each event during the time-period $[0,100]$ will excite the process in such a way that the probability of the next event happening increases and then decreases according to the excitation function, $h(t) = \frac{0.8}{(5.0+t)^{1.5}}$, towards the given background intensity $\nu = 0.1$.

```
#Simulation power law
parameters = {'mu':0.1, 'k':0.8, 'p':1.5, 'c':5.0}
interval = [0,100]
model = hk.simulator().set_kernel('pow').set_baseline('const').set_parameter(
                                              parameters)
T = model.simulate(interval)
model.plot_N() #Plot of time and the number of arrivals
model.plot_l() #Plot of time and conditional intensity function
```

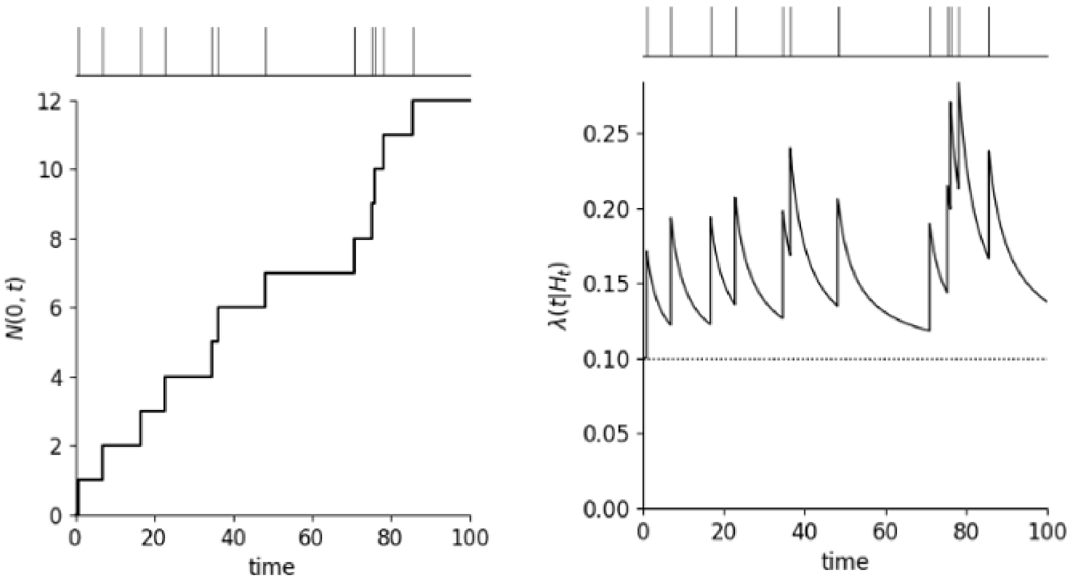

Figure 9.2: Realization Hawkes process (left) with corresponding conditional intensity function; power law function (right).

We also may estimate the parameters and the log-likelihood in both cases. We start with the case of exponentially decaying intensity. We started with parameters $\alpha = 0.25$, $\beta = 1.0$ and the background intensity equal to 0.1. So, the excitation function is given by $h(\cdot)$ to be

$$h(t) = 0.25e^{-1.0t}$$

and the exponentially decaying intensity $\lambda(\cdot)$ is then

$$\lambda(t) = 0.1 + \int_0^t 0.25e^{-1.0(t-s)}dN_s = 0.1 + \sum_{t_i<t} 0.25e^{-1.0(t-t_i)}.$$

The estimated parameters, which we will denote with a hat notation, are approximately $\hat{\alpha} = 0.234$ and $\hat{\beta} = 0.997$. The estimated background intensity is equal to 0.092. The log-likelihood was given to be approximately equal to -362.1.

```
#Parameter estimation
parameters = {"mu":0.1, "alpha":0.25, "beta":1.0}
interval = [0,1000]
model_simulation = hk.simulator().set_kernel('exp').set_baseline('const').
                                   set_parameter(parameters)
T_simulation = model_simulation.simulate(interval)

model = hk.estimator().set_kernel('exp').set_baseline('const')
model.fit(T_simulation,interval)
print("parameters:",model.parameter) #Estimated parameter values
print("log-likelihood:",model.L) #Log-Likelihood
```

For the conditional intensity given as the power law, we chose the excitation function to be given by

$$h(t) = \frac{0.8}{(5.0+t)^{1.5}}$$

and then the corresponding conditional intensity is

$$\lambda(t) = 0.1 + \int_0^t \frac{0.8}{(5.0+(t-s))^{1.5}} dN_s = 0.1 + \sum_{t_i<t} \frac{0.8}{(5.0+(t-t_i))^{1.5}}.$$

We found that the estimated parameters were approximately $\hat{k} = 3.295$, $\hat{p} = 1.779$ and $\hat{c} = 11.098$. The estimated background intensity was approximately 0.101 and the log-likelihood was -587.0.

```
#Parameter estimation power law
parameters = {"mu":0.1, "k":0.8, "p":1.5, "c":5.0}
interval = [0,1000]
model_simulation = hk.simulator().set_kernel('pow').set_baseline('const').
                                        set_parameter(parameters)
T_simulation = model_simulation.simulate(interval)

model = hk.estimator().set_kernel('pow').set_baseline('const')
model.fit(T_simulation,interval)
print("parameters:",model.parameter) #Estimated parameter values
print("log-likelihood:",model.L) #Log-Likelihood
```

Chapter 10

Conclusion

One of the goals of the thesis is that it is a self-contained literature overview of the Hawkes process and the properties of the Hawkes process. This goal is achieved by studying various articles and books and making sure that every notion is included in this thesis. However, the main goal was defining, explaining and proving certain properties of the Hawkes process and the generalizations of the (linear) Hawkes process, such as the mutually exciting Hawkes process and the marked Hawkes process as well as introducing financial applications of the Hawkes process. We achieved this goal as well.

My own contribution involves finding the appropriate literature, studying this literature in detail and extended the proofs given in this literature. Furthermore, I constructed examples and numerical examples that satisfy the conditions of the Law of Large Numbers and the Central Limit Theorem. Moreover, this thesis gives an complete overview of the Hawkes process and its generalizations as well as the properties of the Hawkes process.

Below, one can find a summary of what we have studied during this master thesis.

As we saw the Hawkes process has applications in financial mathematics. The Hawkes jump-diffusion model is able to account for financial contagion were the more common financial models as the Black-Scholes model and the Merton jump-diffusion model failed to do so. We saw that the Hawkes jump-diffusion model is in fact an extension of the Merton jump-diffusion model were the Poisson jump process is replaced by a mutually exciting Hawkes process. Also, the Merton jump-diffusion model is in fact an extension of the Black-Scholes model that incorporated a jump component. We were able to provide a proof for the Law of Large Numbers and the Central Limit Theorem in case of the Hawkes jump-diffusion model. Furthermore, we have seen that the Black-Scholes model and the Merton jump-diffusion model needed certain assumptions on the financial market to derive the formula for the value of an option in terms of the price of a stock. The above was explained in Part three of this thesis, more specific in Chapters 6, 7 and 8.

Chapter 4 dealt with the mutually exciting Hawkes process, the same process that was incorporated in the Hawkes jump-diffusion model. For the mutually exciting Hawkes process we proved the likelihood function and the log-likelihood function. Besides the mutually exciting Hawkes processes we encountered the self-exciting Hawkes process in Chapter 3. For this process we proved the Law of Large Numbers and the Central Limit Theorem as well as the likelihood function. Furthermore, we saw that there were different choices for the excitation function, for instance the exponentially decaying intensity, so an exponential excitation function, and the power law function. For the exponentially decaying intensity function, we showed that it satisfied a certain stochastic differential equation that later showed up in the Hawkes jump-diffusion model. There was one other generalization of the Hawkes process discussed in this thesis, namely the marked Hawkes process. For the marked Hawkes process the proof of the Central Limit Theorem was provided. Also, we gave an example of marked Hawkes processes in geological sciences. In this example, we discussed how the aftershocks of an earthquake could be modeled using the ETAS model. All these chapters were part of the second part of the thesis, the Hawkes processes.

The last chapter in this thesis involved some numerical examples. We provided the Python code to simulate a realization of a Hawkes process when the excitation function is exponential or satisfies the power law function. We also showed how to obtain the estimated parameter values and the log-likelihood.

Chapter 11

Appendix

11.1 Code Chapter 1.3

The Python code below is based on `https://fromosia.wordpress.com/2017/03/19/stochastic-poisson-process/` accessed on 5-5-2023.

```
#Simulation of Homogeneous Poisson Process

import numpy as np
import matplotlib.pyplot as plt

# Prepare data
N = 50 # step
mu = [1, 5, 10]
X_T = [np.random.poisson(lam, size=N) for lam in mu]
S = [[np.sum(X[0:i]) for i in range(N)] for X in X_T]
X = np.linspace(0, N, N)

# Plot the graph
graphs = [plt.step(X, S[i], label="Mu = %d"%mu[i])[0] for i in range(len(mu))]
plt.legend(handles=graphs, loc=2)
plt.title("Homogeneous Poisson Process", fontdict={'fontname': 'Times New Roman', '
                                          fontsize': 21}, y=1.03)
plt.ylim(0)
plt.xlim(0)
plt.xlabel('Time (t)')
plt.ylabel('Cumulative number of events')
plt.show()
```

11.2 Code Chapter 4.1

The Python code below can also be accessed at Github, more specifically `https://github.com/X-DataInitiative/tick/blob/TICK-367-oscar-penalization/doc/modules/code_samples/simulation/plot_hawkes_multidim_simu.py`. For further explanation, see the following article by E. Bacry et al. [3].

```
#Simulation of 3-dimensional Hawkes process

!pip install tick

import numpy as np
import matplotlib.pyplot as plt

from tick.hawkes import SimuHawkesExpKernels
from tick.plot import plot_point_process

n_nodes = 3  #Dimension of the Hawkes process
adjacency = 0.2 * np.ones((n_nodes, n_nodes)) #Intensities of exponential kernels;
                                          \alpha_{ij}
adjacency[0, 1] = 0
decays = 3 * np.ones((n_nodes, n_nodes)) #Decays of exponential kernels; \beta_{ij}
baseline = 0.5 * np.ones(n_nodes) #Background intensities; \nu_i
hawkes = SimuHawkesExpKernels(adjacency=adjacency, decays=decays,
                              baseline=baseline, verbose=False, seed=2398)

run_time = 100
hawkes.end_time = run_time
dt = 0.01
hawkes.track_intensity(dt)
hawkes.simulate()

fig, ax = plt.subplots(n_nodes, 1, figsize=(16, 8), sharex=True, sharey=True)
plot_point_process(hawkes, n_points=50000, t_min=10, max_jumps=30, ax=ax)
fig.tight_layout()
```

11.3 Code Chapter 6

The figures shown in Chapter 6 are created using the following code, which is largely based on the Python code on https://www.codearmo.com/python-tutorial/options-trading-black-scholes-model (Accessed on 20-5-2023)

```
#Option pricing formula of Black-Scholes model
import numpy as np
from scipy.stats import norm
import matplotlib.pyplot as plt

N = norm.cdf

#Value of European call option
def BS_CALL(S, K, T, r, sigma):
    d1 = (np.log(S/K) + (r + sigma**2/2)*T) / (sigma*np.sqrt(T))
    d2 = d1 - sigma * np.sqrt(T)
    return S * N(d1) - K * np.exp(-r*T)* N(d2)

#Effect on option value whenever stock price ($S$) varies
K = 100                  #Strike price
r = 0.1                  #Interest rate
T = 1                    #Maturity time
sigma = 0.3              #Volatility
S = np.arange(60,140,0.1) #Stock price

calls = [BS_CALL(s, K, T, r, sigma) for s in S]
plt.plot(calls, label='Call option')
plt.xlabel('Stock price')
plt.ylabel('Value of call option')
plt.legend()

#Effect on option value whenever the volatility ($\sigma$) varies
K = 100
r = 0.1
T = 1
Sigmas = np.arange(0.0, 1.0, 0.01)
S = 200

calls = [BS_CALL(S, K, T, r, sig) for sig in Sigmas]
plt.plot(Sigmas, calls, label='Call option')
plt.xlabel('Volatility $\sigma$')
plt.ylabel('Value of call option')
plt.legend()
```

11.4 Code Chapter 7

The figure shown in Chapter 7 are created using the following code, which is largely based on the Python code on https://www.codearmo.com/python-tutorial/merton-jump-diffusion-model-python (Accessed on 25-5-2023)

```
#Merton jump-diffusion model
import matplotlib.pyplot as plt
plt.style.use('ggplot')
import numpy as np

def merton_jump_paths(S, T, r, sigma, alpha, m, d, steps, Npaths):
    size=(steps,Npaths)
    dt = T/steps
    pois_rv = np.multiply(np.random.poisson(alpha*dt, size=size),
                          np.random.normal(m, d, size=size)).cumsum(axis=0)
    geo = np.cumsum(((r -  sigma**2/2 -alpha*(m + d**2*0.5))*dt +\
                               sigma*np.sqrt(dt) * \
                               np.random.normal(size=size)), axis=0)

    return np.exp(geo+pois_rv)*S


S = 100          #Current stock price
T = 1            #Maturity time
r = 0.1          #Interest rate
sigma = 0.3      #Volatility
m = 0            #Mean of jump size Y_t
d = 0.5          #Standard deviation of jump size Y_t
alpha = 1        #Rate of Poisson process
steps = 365      #Number of time steps
Npaths = 1       #Number of paths to simulate


MJD = merton_jump_paths(S, T, r, sigma, alpha, m, d, steps, Npaths)

plt.plot(MJD)
plt.xlabel('Days')
plt.ylabel('Stock price')
plt.title('Merton jump-diffusion')
```

Bibliography

[1] Y. Aït-Sahalia, J. Cacho-Diaz and R.J. Laeven, (2015), *Modeling financial contagion using mutually exciting jump processes*, Journal of Financial Economics, 117:3, pp. 585-606

[2] E. Bacry, I. Mastromatteo and J.-F. Muzy, (2015), *Hawkes processes in Finance*, Market Microstructure and Liquidity, DOI: 10.1142/S2382626615500057

[3] E. Bacry, M. Bompaire, P. Deegan, S. Gaiffas, S. Poulsen, (2018), *tick: a Python Library for Statistical Learning, with an emphasis on Hawkes Processes and Time-Dependent Models*, Journal of Machine Learning Research, 18, pp. 1-5

[4] B. Bian, X. Chen, X. Zeng, (2019), *Optimal Portfolio Choice in a Jump-Diffusion Model with Self-Exciting*, Journal of Mathematical Finance, 9, pp. 345-367. doi: 10.4236/jmf.2019.93020.

[5] F. Black and M. Scholes, (1973), *The Pricing of Options and Corporate Liabilities*, The Journal of Political Economy, 81:3, https://doi-org.proxy.library.uu.nl/10.1086/260062

[6] F. Boshuizen, A. van der Vaart, H. van Zanten, K. Banachewicz, P. Zareba, and E. Belitser, (2021), *Lecture notes for the course Stochastic Processes for Finance*, Vrije Universiteit Amsterdam

[7] R. Cont and P. Tankov, (2004), *Financial Modelling with Jump Processes*, Financial Mathematics Series, Chapman and Hall/CRC

[8] D. Daley and D. Vere-Jones, (2003), *An Introduction to the Theory of Point Processes*, 2nd edition, Vol. I, Springer: New York

[9] D. Daley and D. Vere-Jones, (2007), *An Introduction to the Theory of Point Processes*, 2nd edition, Vol. II, Springer: New York

[10] R. Fierro, V. Leiva and J. Moller, (2015), *The Hawkes process with different exciting functions and its asymptotic behavior*, Journal of Applied Probability, 52(1), pp. 37-54, DOI: 10.1239/jap/1429282605

[11] A.G. Hawkes, (1971), *Spectra of some self-exciting and mutually-exciting point processes*, Biometrika, 58, pp. 83–90

[12] A.G. Hawkes, (1974), *A cluster process representation of a self-exciting process*, Journal of Applied Probability, Volume 11, No., pp. 493–503.

[13] A.G. Hawkes, (2018), *Hawkes processes and their applications to finance: a review*, Quantitative Finance, 18:2, pp. 193-198, DOI: 10.1080/14697688.2017.1403131

[14] Y. Jiao, (2020), *From Probability to Finance: Lecture Notes of BICMR Summer School on Financial Mathematics*, Springer Singapore, DOI: 10.1007/978-981-15-1576-7

[15] P.J. Laub, T. Taimre and P.K. Pollett, (2015), *Hawkes Processes*, arXiv:1507.02822v1

[16] P.J. Laub, Y. Lee and T. Taimre, (2021), *The Elements of Hawkes Processes*, Springer Cham, DOI: 10.1007/978-3-030-84639-8

[17] R.C. Merton, (1976), *Option pricing when underlying stock returns are discontinuous*, Journal of Financial Economics, 3:1, pp. 125-144

[18] P. Morters and Y. Peres, (2010), *Brownian Motion*, Cambridge University Press, DOI: 10.1017/CBO9780511750489

[19] Y. Ogata, (1999), *Seismicity Analysis through Point-process Modeling: A Review*, Pure and applied geophysics, 155, pp. 471–507

[20] T. Omi, https://github.com/omitakahiro/Hawkes/, https://colab.research.google.com/github/omitakahiro/Hawkes/blob/master/notebooks/tutorial.ipynb, (Both accessible on 1-3-2023)

[21] J.G. Rasmussen, (2018), *Temporal point processes: the conditional intensity function*, https://doi.org/10.48550/arXiv.1806.00221

[22] M.-A. Rizoiu, S. Mishra, Q. Kong, M. Carman, L. Xie, (2018), *SIR-Hawkes: Linking Epidemic Models and Hawkes Processes to Model Diffusions in Finite Populations*, arXiv:1711.01679v3

[23] J.S. Rosenthal, (2006), *A first look at rigorous probability theory*, World Scientific Publishing Co. Pte. Ltd., Hackensack, New Jersey, Second edition

[24] S.M. Ross, (2010), *Introduction to Probability Models*, Academic Press, Tenth Edition, ISBN: 978-0-12-375686-2

[25] W.M. Ruszel, (2021), *Lecture notes on Mathematical Statistics WISB263*, Utrecht University

[26] Y. Seol, (2020), *Limit theorems for a jump-diffusion model with Hawkes jumps*, Mathematical communications, Volume 26, pp. 53–67

[27] Y. Seol, (2019), *Limit theorems for Hawkes processes with uniform immigrants*, J. Korean Math. Soc. 56, No. 4, pp. 935–946, https://doi.org/10.4134/JKMS.j180483

[28] R. Shcherbakov, (2021), *Statistics and forecasting of aftershocks during the 2019 Ridgecrest, California, earthquake sequence*, Journal of Geophysical Research: Solid Earth, 126, e2020JB020887, https://doi.org/10.1029/2020JB020887

[29] S.E. Shreve, (2004), *Stochastic Calculus for Finance II - Continuous Time Models*, Springer Finance, ISBN: 9781441923110

[30] P.J.C. Spreij, (2021), *Measure Theoretic Probability*, Lecture notes of the course Measure Theoretic Probability, Chapter 1

[31] H.J.T. Unwin, I. Routledge, S. Flaxman, M.A. Rizoiu, S. Lai, J. Cohen, et al., (2021), *Using Hawkes Processes to model imported and local malaria cases in near-elimination settings.* PLoS Computational Biology, 17(4), https://doi.org/ 10.1371/journal.pcbi.1008830

[32] A.W. van der Vaart, (1995), *Mathematische Statisiek*, Lecture notes of the course Asymptotic Statistics, Chapter 1

[33] H. Wang, (2021), *Large and moderate deviations for a discrete-time marked Hawkes process*, arXiv:2003.05772v2

[34] H. Wang, (2022), *Limit theorems for a discrete-time marked Hawkes process*, Statistics and Probability Letters, Volume 184, https://doi.org/10.1016/j.spl.2022.109368

[35] L. Wasserman, (2006), *All of Nonparametric Statistics*, Springer New York, p. 8, https://doi-org.proxy.library.uu.nl/10.1007/0-387-30623-4

[36] S. Wheatley, V. Filimonov, and D. Sornette, (2016), *The Hawkes process with renewal immigration and its estimation with an EM algorithm*, Computational Statistics and Data Analysis, 94, pp. 120-135

[37] L. Zhu and D. Karabash, (2012), *Limit Theorems for Marked Hawkes Processes with Application to a Risk Model*, Stochastic Models, Volume 31, Issue 3,
DOI: 10.1080/15326349.2015.1024868.