Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques i Informàtica
Universitat de Barcelona

HAWKES PROCESSES IN FINANCE

Autor: Andrea Bosquet Rodríguez

Director: Dr. Jose Manuel Corcuera
Realitzat a: Departament de Matemàtiques i Informàtica

Barcelona, 19 de juny de 2019

Contents

Abstract

Finances are an important field where stochastic processes are applied. These processes allow to model different finance situations, such as price modeling or risk. The aim of this project is to study a type of stochastic processes, the Hawkes processes, which are an extension of Poisson processes that considers self-excitation, and see some of their application in the financial field.

2010 Mathematics Subject Classification. 60G10, 60G55, 91B26, 91B70

Acknowledgements

I would like to specially thank my supervisor Jose Manuel Corcuera for his guidance, help and advice during this research. Moreover, I would like to express my gratitude to the opportunity that gave me the Universitat de Barcelona to do an Exchange Program in Switzerland, which introduce me to the field of Stochastics. Finally, I would like to thank my friends and my family for their support and their patience.

Introduction

Hawkes processes are a generalization of homogeneous Poisson processes whose Poisson rate depends on time and the history, making that a new jump influences the probability of having a new jump.
These processes were first introduced by Alan Hawkes in 1971 [25] with the idea of modelling earthquakes: his idea was to model the replica effects after an earthquake. However, these processes have not come to standstill by this application, nowadays they are applied in many other fields, such as neurology, finance and social science. Particularly, the first one to introduce these processes in finance was C.G. Bowsher in 2003 [12]. Before Hawkes process was considered in finance, in most of its applications it was considered a Poisson process with more weight in the tail. After its apparition, the Poisson jump was replaced by a mutually-exciting Hawkes process which could propagate different effects between markets.
This project is divided in three different chapters: (i) Basic notions, where some general concepts that are needed later when talking about Hawkes processes, (ii) Hawkes processes, where an introduction to these processes is done together with different properties and characterizations and (iii) Finance applications, in which an example of application in this field is shown.
In chapter one, some general concepts of probability are defined. Then there are some general knowledge of martingales. Finally, an introduction to Point process is done.
In chapter two, first it is defined a Hawkes process and its extrapolation to m dimensions. Secondly, the concept of stationarity is given for these processes. Next, its first and second order are given. Following, different ways of representing a Hawkes processes are described. Then, different extensions considered in literature are introduced. Finally, a way of estimate the parameters given historical data is explained.
In chapter three, the market modelling application is described by using two different types of Hawkes processes.

Chapter 1

Basic notions

Prior to embarking upon Hawkes processes, it is precise to set some general concepts from the stochastic field, such as martingales and some results of point processes, which this last is the basis of Hawkes processes.

1.1 Introduction

Throughout this section, we shall provide some general definitions of the probability field which will let us to built the basis of this project. To this purpose, let us start off by giving some elemental definitions [19]:

Definition 1.1. Let Ω be a nonempty set, named the *sample space*, let $\mathcal{H}$ be a $\sigma-$algebra on Ω, also called the *event set*, and let $\mathbb{P} : \mathcal{H} \to [0,1]$ be a *probability measure*. The tuple $(\Omega, \mathcal{H})$ is called a *measurable space*, and the triple $(\Omega, \mathcal{H}, \mathbb{P})$ a *probability space*. Elements of the event set $\mathcal{H}$ are called *events*.

Definition 1.2. Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space, I an index set and let (χ, ξ) be a measurable space. For every $i \in I$, let $X_i : \Omega \to \chi$ be a random variable. The collection of random variables $(X_i)_{i\in I}$ (or $\{X_i : i \in I\}$) is called a *stochastic process*.

Definition 1.3. Let $(\Omega, \mathcal{H})$ be a measurable space and let $(\mathcal{H}_i)_{i\in I}$ be a family of σ-algebras on Ω such that $\mathcal{H}_i \subset \mathcal{H}_{i+1} \subset \mathcal{H}\ \forall i \in I$. Then, $\mathbb{H} = (\mathcal{H}_i)_{i\in I}$ is called a *filtration* of $(\Omega, \mathcal{H})$.

1.2 Martingales

As it will be seen in Section 1.3.2 and remarked in Section 2.4.3, a point process, and in particular a Hawkes process, can be decomposed in a martingale and

another addend. For this reason, we shall briefly introduce martingales. Following [24]:

Definition 1.4. Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a probability space and let $\mathbb{H}$ be a filtration of the measurable space $(\Omega, \mathcal{H})$. Let $I = \mathbb{N}$, then the stochastic process $(X_n)_{n\in\mathbb{N}}$ is a *martingale* with respect to $\mathbb{H}$ if it satisfies the following three properties:

- X_n is measurable with respect to $\mathcal{H}_n$, that is, $X_n^{-1}(B_n) = \{\omega \in \Omega : X(\omega) \in B_n\} \in \mathcal{H}_n \quad \forall B_n \subset \chi$.
- $E[|X_n|] < \infty$.
- $E(X_n|\mathcal{H}_m) = X_m$ a.s. $\forall m \leq n$.

It is worth highlighting that if $(X_n)_{n\geq 0}$ is a martingale, then $E[X_m] = E[X_n]$ $\forall m, n$.
Additionally, if the first two properties are satisfied and the third one is replaced by the inequality $E[X_n|\mathcal{H}_m] \geq X_m$ ($E[X_n|\mathcal{H}_m] \leq X_m$) a.s. $\forall n \geq m$, then $(X_n)_{n\in\mathbb{N}}$ is called a *submartingale* (*supermantigale*).

The reason behind introducing the stopping times is that they generalize processes which are not martingales, but locally they behave as martingales. It therefore follows the next definition:

Definition 1.5. Given a filtration $\mathbb{H} = (\mathcal{H}_t)_{t\in\mathbb{R}^+}$, then a random variable T taking values on $[0, \infty]$ is an $\mathbb{H}$-*stopping time* if and only if $\{T \leq t\} \in \mathcal{H}_t \ \forall t \geq 0$.

It is interesting to mention that a stopping time marks the time at which a determined process has a behaviour which is of interest. Depending on their predicting behaviour, there are two different types of stopping times [16]:

Definition 1.6. A stopping time T is *predictable* and predicted by a sequence of random times $\{T_n\}_{n\in\mathbb{N}}$ if $T_n \to T$ and $T_n < T \ \forall n \in \mathbb{N}$ on the event $\{T > 0\}$.
A stopping time T is *totally inaccessible* if $\mathbb{P}(T = S) = 0$ for every predictable stopping time S.

Observe that in the definition of predictable stopping time, it has not been mentioned that the sequence is made of stopping times, although it is implicit:

$$\{T \leq t\} = \bigcap_n \{T_n \leq t\} \in \mathcal{H}_t$$

We are now set to give the definition of local martingale:

Definition 1.7. Let $\{X(t) : t \geq 0\}$ be a stochastic process and let $\mathbb{H} = (\mathcal{H}_t)_{t\geq 0}$ be a filtration on the measurable space $(\overline{\mathbb{R}}^+, \mathcal{B}(\overline{\mathbb{R}}^+))$. Consider $\{T_n : n \geq 1\}$ an increasing family of $\mathbb{H}$-stopping times such that $\lim_{n\to\infty} T_n = +\infty$ and that for $n \geq 1$, $X(t \wedge T_n)$ is an $\mathcal{H}_t$-martingale. Then, $X(t)$ is an $\mathcal{H}_t$-*local martingale*.

Definition 1.8. A *semimartingale* is a process which can be decomposed as the sum of a local martingale and an adapted finite variation process, that is

$$M(t) = M(0) + X(t) + A(t)$$

where $M(0)$ is finite and $\mathcal{H}_0$-measurable, $X(t)$ is a local martingale and $A(t)$ is the process whose path have bounded variation on $[0, t]$ for each t.

Definition 1.9. Let $\mathbb{F}$ and $\mathbb{H}$ be two filtration such that $\mathbb{F} \subset \mathbb{H}$. Then, there is *immersion* between the filtration if any $\mathbb{F}$-local martingale is a $\mathbb{H}$-local martingale.

1.3 Point processes

Due to the fact that we will study a particular characterization of Point processes, it is time to put forward some fundamental definitions for the upcoming sections.

Definition 1.10. Let $\{N(t) \in \mathbb{N} : t \geq 0\}$ be a stochastic process such that $N(0) = 0$, it is almost surely finite and its trajectories are right-continuous step functions with increments of one unit. Then, $\{N(t) \in \mathbb{N} : t \geq 0\}$ is a *counting process*.

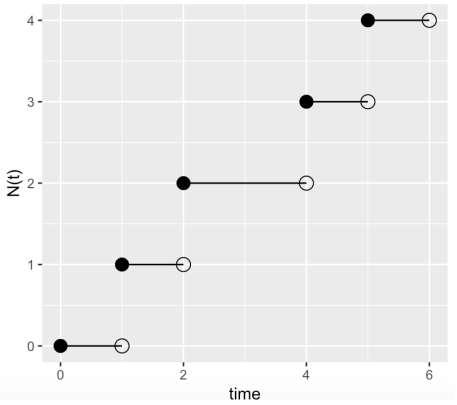

Figure 1.1: Example of counting process where the jumps have taken place in $T = \{1, 2, 4, 5\}$

It is interesting to highlight that a counting process just enumerate the times a determined event happens, such as when a person comes in a shop.
From this definition, the following properties are derived:

- $N(t) \geq 0 \; \forall t \geq 0$
- $N(t) \in \mathbb{N}$
- if $t_1 < t_2$ then $N(t_1) \leq N(t_2)$

So it turns out from the third property that the number of increments in the interval $(t_1, t_2]$ is obtained by considering $N(t_2) - N(t_1)$.

It is apparent that the information given by a counting process is completely equivalent to the information given by the times the jumps take place. These times can be defined as ordered random points in the time space whose weights are the same. This is quite remarkable since the study of the arrival times will allow to derive some properties of the counting process. Now, in order to bridge this relation, it therefore follows the next definition:

Definition 1.11. Let $T = \{T_1, T_2, ...\}$ be a sequence of random variables taking values in $[0, \infty)$ which has $\mathbb{P}(0 \leq T_1 \leq T_2 \leq ...) = 1$. If the subsets $T'_k = \{T_i : T_i \leq k, \; k < \infty\} \subset T$ have a finite cardinality, then T is an *ordered (simple) point process*.

It is remarkable that the condition $|T'_k| < \infty \; \forall k$ excludes processes that explode[1]; that is, having infinite jumps in a short period of time. Later, a criterion to ensure the non-explosion is given (see Lemma 2.14).
We will denote by $\mathcal{H}_t$, named the history, the list of time events $\{t_1, ..., t_n\}$ up to time t.
By defining T_i as the time at which the i-jump occurs and whose corresponding counting process is denoted as $\{N(t) : t \geq 0\}$, the following relation is obtained:

$$N(t) = \sum_{i \geq 1} \mathbb{1}_{\{T_i \leq t\}}$$

It is of interest to highlight that $N(t)$ is $\mathcal{H}$-adapted[2]. Moreover, since T_i is the first time for which $N(t) = i$, by the *debut theorem*[3], it leads that T_i is a stopping time.

Due to the fact that we will be mainly studying a generalization of an homogeneous Poisson process, the simplest example of point process, it is important to know their characterization and their properties.

[1] **Definition:** The *explosion* is defined as the minimum time interval $t - s < \infty$ that satisfies $N(t) - N(s) = \infty$ for $t - s < \infty$. Thus, if this time exists, we will say that the stochastic process $\{N(t) : t \geq 0\}$ explodes.

[2] **Definition:** A continuous process $\{N(t) : t \geq 0\}$ is *adapted* to a filtration $\mathbb{H}$ if $\forall t \geq 0$, $N(t)$ is measurable with respect to this filtration.

[3] **Theorem:** (*debut theorem*) Let X be an adapted right-continuous stochastic process such that is defined in the whole probability space. Then, if $K \in \mathbb{R}$, then the process $T : \Omega \to \overline{\mathbb{R}^+}$ defined as:

$$T(\omega) = \inf_{t \in \mathbb{R}^+} \{X(t + \omega) \geq K\}$$

is almost surely an stopping time.

Example 1.12. An homogeneous Poisson process is defined as the process whose probability is given by:

$$\mathbb{P}[N(t+h)-N(t)=m|\mathcal{H}_t]=\begin{cases}\lambda h+o(h) & \text{if} \quad m=1\\ o(h) & \text{if} \quad m>1\\ 1-\lambda h+o(h) & \text{if} \quad m=0\end{cases}$$

where $\lambda>0$ is the *Poisson rate* which is constant and $o(h)$ is such that $\lim_{h\to 0}\frac{o(h)}{h}=0$. It is apparent that the probability of finding a point in the interval $(t,t+h]$ is independent of $\mathcal{H}_t$, thus, it is *memoryless*[4]. Moreover, by the second equation, simultaneous jumps are excluded.
It is time to put forward some fundamental results of homogeneous Poisson processes which will give a deeper knowledge of these type of processes.

Proposition 1.12.1: The time intervals between jumps in an homogeneous Poisson process, $T_{i+1}-T_i$, are independently exponential distributed.

Proof. Given the point process $\{T_k : k\in\mathbb{N}\}$ which describes the times at which jumps take place, if we denote by $\tau=T_{i+1}-T_i$ the time intervals between jumps, supposing that $t_2>t_1>0$, the survival function $S_\tau(\cdot)$ satisfies the independence property:

$$S_\tau(t_1+t_2)=\mathbb{P}(\tau>t_1)\mathbb{P}(\tau>t_1+t_2|\tau>t_1)=\mathbb{P}(\tau>t_1)\mathbb{P}(\tau>t_2)=S_\tau(t_1)S_\tau(t_2)$$

where in the second equality one has used that $t_2>t_1$ and the time has started over because of the memorylessness. Since the initial condition can only be $S_\tau(0)=1$, because of the independence:

$$S_\tau(0)=S_\tau(0+0)=(S_\tau(0))^2\to S_\tau(0)=0 \text{ or } S_\tau(0)=1$$

If $S_\tau(0)=0$, then $S_\tau(x)=0\ \forall x>0$ as $S_\tau(x)=S_\tau(x+0)=S_\tau(x)S_\tau(0)=0$. Thus $S_\tau(0)=1$. Furthermore:

$$S_\tau(1)=S_\tau(1/2+1/2)=(S_\tau(1/2))^2\geq 0\to S(1)=\alpha\geq 0$$

$$S_\tau(n)=S_\tau(1+...+1)=(S_\tau(1))^n=\alpha^n \quad \forall n\in\mathbb{N}$$

$$S_\tau(1)=S_\tau(1/n+...+1/n)=(S_\tau(1/n))^n\to S_\tau(1/n)=\alpha^{1/n}$$

[4]**Definition:** Let X be a random variable, we say that its probability distribution is *memoryless* if it satisfies:

$$\mathbb{P}[X>m+n|X>n]=\mathbb{P}[X>m]$$

Therefore $S_\tau(m/n) = \alpha^{m/n}$. By the continuity of random variables and because of the density of rationals in real numbers, $S_\tau(x) = \alpha^x \ \forall x \geq 0$. As $S_\tau(\cdot)$ is a probability, $S_\tau(x) = e^{-\lambda x}$ for $\lambda > 0$ (where the case $\alpha = 0$ has been excluded, otherwise, the density would be a Dirac point mass, which is not continuous). Hence, it is an exponential distribution, that is, $T_{i+1} - T_i \sim E_{i+1}$.
In particular, $T_n = E_1 + ... + E_n$ where E_i are independent exponential distributions. □

The reason behind introducing the above result is that it allows to completely characterize the distribution of jumps for these processes:

Proposition 1.12.2: For an homogeneous Poisson processes, the number of increments in the interval $(a, b]$, $N(b) - N(a)$, follows a Poisson distribution of parameter λ.

Proof. To this purpose, let us start off by seeing that the the n-jump follows a Gamma law with parameters n and λ, Gamma(n, λ), which corresponds to the following density function:

$$f_{T_n}(t) = \frac{\lambda^n}{(n-1)!} t^{n-1} e^{-\lambda t} \mathbb{1}_{\{t>0\}}$$

By using the induction method on n, we have that for the initial case, $n = 1$, T_1 follows an exponential law with parameter λ, as seen in the last proof, which is equivalent to a Gamma$(1, \lambda)$.
Suppose now that $T_n \sim$ Gamma(n, λ) (induction hypothesis). Since $E_{n+1} \sim$ Exp(λ) is independent of T_n, for $t > 0$, the distribution function is written as:

$$F_{T_{n+1}}(t) = \mathbb{P}(T_{n+1} \leq t) = \mathbb{P}(T_n + E_{n+1} \leq t) = \int_0^\infty \mathbb{P}(T_n + E_{n+1} \leq t | T_n = u) \cdot$$

$$\cdot f_{T_n}(u) du = \int_0^t \mathbb{P}(E_{n+1} \leq t - u) f_{T_n}(u) du = \int_0^t F_{E_{n+1}}(t-u) f_{T_n}(u) du$$

Thus:

$$f_{T_{n+1}}(t) = \frac{d}{dt} \int_0^t (1 - e^{-\lambda(t-u)}) f_{T_n}(u) du = \int_0^t \left((1 - e^{-\lambda(t-u)}) f_{T_n}(u) \right)' du + (1 - e^{-\lambda(t-t)}) \cdot$$

$$\cdot f_{T_n}(t) = \int_0^t \lambda e^{-\lambda(t-u)} \frac{\lambda^n}{(n-1)!} u^{n-1} e^{-\lambda u} du = \frac{\lambda^{n+1}}{(n-1)!} e^{-\lambda t} \int_0^t u^{n-1} du = \frac{\lambda^{n+1}}{n!} t^n e^{-\lambda t}$$

Now, we are ready to see that $N(t)$ follows a Poisson law. Denoting by c the interval length, $c = b - a$, and using the equivalence between the number of increments and the times in which these increments takes place, it turns out:

$$\mathbb{P}[N(b) - N(a) < k] = \mathbb{P}[T_1 + ... + T_k > c] = \int_c^\infty \frac{\lambda(\lambda u)^{k-1} e^{-\lambda u}}{(k-1)!} du = \sum_{s=0}^{k-1} \frac{e^{-\lambda c} (\lambda c)^s}{s!}$$

where in the last equality one has integrated by parts $(k-1)$ times. Then, it is now evident that $N(t) \sim Poisson(\lambda t)$, since:

$$\mathbb{P}(N(t)=n) = \mathbb{P}(N(t)-N(0)\geq n) - \mathbb{P}(N(t)-N(0)\geq n+1) =$$

$$\sum_{k=0}^{n} e^{-\lambda t}\frac{(\lambda t)^k}{k!} - \sum_{k=0}^{n-1} e^{-\lambda t}\frac{(\lambda t)^k}{k!} = e^{-\lambda t}\frac{(\lambda t)^n}{n!}$$

□

The natural question now is how to characterise a point process. Imagine, for example, that one wants to characterise the occurrence of earthquakes. One way to typify this situation is by knowing what is the probability of having an earthquake considering the past; in other words, the distribution function of the arrival time conditioning on the history $\mathcal{H}_u$. Being aware of this idea, the conditional cumulative distribution function of the next jump T_{k+1}, given the history up to last jump k taken at time u, $\mathcal{H}_u$, is:

$$F^*(t|\mathcal{H}_u) = \int_u^t \mathbb{P}[s \leq T_{k+1} \leq s+ds|\mathcal{H}_u]ds = \int_u^t f^*(s|\mathcal{H}_u)ds$$

Notation: The * as a superindex indicates conditioning on the history $\mathcal{H}_u$.
Supposing that the realisations are in $\{t_1, t_2, ..., t_k\}$ and using the law of total probabilities the arriving conditional distribution can be written as:

$$f(t_1, t_2, ..., t_k) = \prod_{i=1}^{k} f^*(t_i|\mathcal{H}_u)$$

1.3.1 Conditional intensity function

The reason behind introducing the conditional intensity function is that working with $f^*(\cdot|\mathcal{H}_i)$ can be very difficult in many situations; which creates the necessity of an alternative representation for a point process. It therefore follows the following definition:

Definition 1.13. The *conditional intensity function*, also known as the *complete intensity function*, is defined as the expected rate of jumps conditioned on $\mathcal{H}_t$:

$$\lambda^*(t|\mathcal{H}_t) = \frac{E[dN(t)|\mathcal{H}_t]}{dt} = \lim_{h\to 0^+} \frac{E[N(t+h)-N(t)|\mathcal{H}_t]}{h} = \lim_{h\to 0^+} \frac{\mathbb{P}(t\leq T < t+h|\mathcal{H}_t)}{h}$$

where $dN(t) = \lim_{h\to 0}[N(t+h)-N(t)]$.

It is worth highlighting that it only depends on information of the counting process in the past, which means it is $\mathcal{H}_t$-measurable. Moreover, notice that $\lambda^*(t)$ can only be non-negative.

An alternative way to define the conditional intensity function can be found in different papers, such as [28]; where the equivalence with the *hazard function* is given. The hazard function is the ratio of failure for small intervals of time, that is:

$$\lambda^*(t|\mathcal{H}_t) = \frac{f^*(t|\mathcal{H}_t)}{1 - F^*(t|\mathcal{H}_t)} \tag{1.1}$$

However, note that although they can have the same expression (as we will see next), they are not exactly the same: the hazard function is based on continuous values from a population, whereas the conditional intensity function considers the probability at which jumps occur.
Additionally, by considering Equation (1.1), it is possible to obtain the survival function in terms of the hazard function:

$$\lambda^*(t) = -\frac{\dot{S}^*(t|\mathcal{H}_t)}{S^*(t|\mathcal{H}_t)} = -\frac{d}{dt}\ln(S^*(t|\mathcal{H}_t))$$

Thus, it is now clear that the survival function follows an exponential distribution:

$$S^*(t|\mathcal{H}_t) = e^{-\int_0^t \lambda^*(u)du} \tag{1.2}$$

Note: In order to simplify notation in the following chapters, from now, the conditional intensity function is written as $\lambda^*(t)$ instead of $\lambda^*(t|\mathcal{H}_t)$.
In order to attain the equivalence between the conditional intensity function and the hazard function, a non-rigorous demonstration following [21] is shown next. However, a rigorous proof can be seen in Corollary 4.1.2 of [11] by taking the limit $s \to t$.

Proof. Consider an infinitesimal interval around t, say dt, then, taking into account Definition 1.11, where it is not possible to have simultaneous jumps and considering the probability of a first jump, one obtains:

$$\frac{f^*(t|\mathcal{H}_t)}{1 - F^*(t|\mathcal{H}_t)} = \frac{\mathbb{P}(\text{point in dt}|\mathcal{H}_t)}{\mathbb{P}(\text{point not before t}|\mathcal{H}_t)} = \frac{\mathbb{P}(\text{point in dt, point not before t }|\mathcal{H}_t)}{\mathbb{P}(\text{point not before t}|\mathcal{H}_t)} =$$

$$\mathbb{P}(\text{point in dt}|\text{point not before t}, \mathcal{H}_t) = \mathbb{P}(\text{point in dt}|\mathcal{H}_t) = E[N(dt)|\mathcal{H}_t] = \lambda^*(t)$$

where in the second equality we have considered that it is the first jump and in the forth equality, it is used that the information up to time t is already included in $\mathcal{H}_t$. □

Nevertheless, the conditional intensity function only characterises the finite dimensional point processes in case it exists [28].

The following result shows the complete characterization of the process given the jumps before an origin of time, result given by Alan Hawkes in [26]:

Lemma 1.14. Let t_0 be an origin of time and suppose that the jumps before this time follows a specific distribution. Then, there exists at most one orderly point process $\{N(t) : t \geq 0\}$ which satisfies Definition 1.11 for a given function $\lambda^*(\cdot)$ given the history.

Proof. See [26]. □

Going back to the information given by the conditional intensity function, depending on the future consequences of having a new jump, the point process can be classified in:

- *self-exciting* if a jump increases the conditional intensity function; that is, the fact that a new jump occurs increases the probability of having a new jump. For this reason, there will be a temporal clustering of the arrival time -a lot of increments in a short period of time-.

- *self-regulating* if a jump causes the conditional intensity function to decrease. That is, when a new increment occurs, the probability of having a new increment is decreased. In this case, the arrival times appear quite temporally regular.

In order to be more familiar with these two types of process, an example of each one is shown next:

Example 1.15. The behavior of an earthquake can be modelled with a self-exciting process, whereas for modelling an important loss of money, a self-regulating process would be an ideal choice.

Particularly, in Chapter 2 some explicit expressions of these two types of conditional intensity function are given.

1.3.2 Compensator

Throughout this subsection, more properties of counting processes are given by considering their compensator, which is defined as:

Definition 1.16. The *compensator* of a counting process is defined as the integral of its conditional intensity function:

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

Notice that it can be interpreted as the conditional mean of $N(t)$ given the past. Calling to mind Equation (1.2), the survival function can be given in terms of the compensator:

$$S(t) = e^{-\int_0^t \lambda^*(u)du} = e^{-\Lambda(t)}$$

So as to see the importance of the compensators for a counting process, firstly, it is necessary to see that the counting process is a submartingale:

Proposition 1.17. The counting process defined in terms of the arrival times, $N(t) = \sum_{i\geq 1} \mathbb{1}_{\{T_i \leq t\}}$, is a submartingale.

Proof. Following the conditions on Definition 1.4, it is implicit to see the $N(t)$ is $\mathcal{H}_t$-measurable (as $\{T_i \leq t\} \in \mathcal{H}_t$), and consequently, $N(t)$ is $\mathbb{H}$-adapted.
Trivially, $E[|N(t)|] < \infty$ since the increments can only be of one unit, so:

$$E[|N(t)|] = E[N(t)] = \mathbb{P}[N(t) = 1] < \infty$$

And finally, $E[N(t)|\mathcal{H}_s] \geq N(s)$ because:

$$E[N(t)|\mathcal{H}_s] = E\left[\sum_{i=1}^{\infty} \mathbb{1}_{\{T_i \leq t\}}|\mathcal{H}_s\right] = \sum_{i=1}^{\infty} E\left[\mathbb{1}_{\{T_i \leq t\}}|\mathcal{H}_s\right] =$$

$$\sum_{i=1}^{\infty} \mathbb{1}_{\{T_i \leq s\}} + \sum_{i=1}^{\infty} E\left[\mathbb{1}_{\{s < T_i \leq t\}}|\mathcal{H}_s\right] \geq \sum_{i=1}^{\infty} \mathbb{1}_{\{T_i \leq s\}} = N(s)$$

where in the inequality it is used that $E\left[\mathbb{1}_{\{s<T_i\leq t\}}|\mathcal{H}_s\right] \geq 0$. □

Then, $N(t) = \sum_{i\geq 1} \mathbb{1}_{\{T_i \leq t\}}$ can be decomposed into a zero mean martingale $M(t)$ and a unique $(\mathcal{H}_t)$-predictable increasing process, $\Lambda(t)$. This decomposition is known as *Doob-Meyer decomposition*:

$$N(t) = M(t) + \Lambda(t) \tag{1.3}$$

where $E[M(t)] = 0$.

Proof. See [35]. □

More details of this representation are shown in upcoming sections.
It is time to put forward some important results for the characterization of point processes, and in particular, of counting processes [29]:

Lemma 1.18. Let $\Lambda(t)$ be the compensator of a given process X. Then for any stopping time T:

- $\Delta\Lambda(T) = 0$ if T is a totally inaccessible stopping time.
- $\Delta\Lambda(T) = E[\Delta X(T)|\mathcal{H}_{T^-}]$ if T is a predictable stopping time.

where Δ denotes increments of the variable that comes next.

Proof. See Lemma 2 of [29]. □

This result is quite remarkable since the study of the variations of the compensator are described in terms of the expected value of the point process.

Lemma 1.19. Let $N(t)$ be a counting process whose compensator is given by $\Lambda(t)$. Then, $N(\infty) < \infty$ if and only if $\Lambda(\infty) < \infty$ with probability 1.

Proof. Let τ be the first time at which $\Lambda(\tau) \geq n$. As the compensator is an increasing function, it is clear that $\forall t < \tau \ \Lambda(t) < n$. Moreover, as the counting process can only increase by increments of one unit, the same is for the compensator because of Lemma 1.18. Thus, $\Lambda(\tau) < n + 1$ and the following inequality is obtained:

$$E[N(\tau)] = E[\Lambda(\tau) + M(\tau)] = E[\Lambda(\tau)] + E[M(\tau)] = E[\Lambda(\tau)] < n + 1$$

which means that $N(\tau) < \infty$[5]. Then, if $\Lambda(\infty) < \infty$ one obtains that $N(\infty) < \infty$ letting $n \to \infty$.
Conversely, define τ as the first time at which $N(\tau) = n$, hence $N(\tau) \leq n$. Then, the Doob-Meyer decomposition yields to:

$$E[\Lambda(\tau)] = E[N(\tau)] \leq n$$

Therefore, supposing that $N(\infty) < n$, that is for $\tau = \infty$, and letting $n \to \infty$ one obtains that $\Lambda(\infty) < \infty$ almost surely. □

Apart from a relation to determine when the process is finite, it also characterizes when the counting process is constant:

[5] **Property:** If $E(X) < \infty$ then $X < \infty$ almost surely.

Lemma 1.20. Let $N(t)$ be a counting process with compensator $\Lambda(t)$. Then, $N(t)$ is almost-surely constant on all intervals on which the compensator is constant.

Proof. Let $s, \epsilon > 0$ and let $T_{s,\epsilon}$ denote the stopping time defined as:

$$T_{s,\epsilon} = \inf\{t \geq s : \Lambda(t) \geq \Lambda(s) + \epsilon\}$$

As $T_{s,\epsilon}$ is predictable, there exists a sequence of stopping times $T_n < T_{s,\epsilon}$ increasing to $T_{s,\epsilon}$. Letting n increase to infinity, then:

$$E[N(T_{s,\epsilon}) - N(s)] = \lim_{n\to\infty} E[N(T_n) - N(s)] = \lim_{n\to\infty} E[\Lambda(T_n) - \Lambda(s)] \leq \epsilon$$

Suppose now that for any $t > s$, $\Lambda(t) = \Lambda(s)$. For the right-continuity of the compensator, $T_{s,\epsilon}$ is strictly greater than t:

$$E\left[\sup_{t>s} \mathbb{1}_{\{\Lambda(t)=\Lambda(s)\}}(N(t) - N(s))\right] \leq E[N(T_{s,\epsilon}) - N(s)] \leq \epsilon$$

Nevertheless, this inequality holds for any $\epsilon > 0$, thus, $E\left[\sup_{t>s} \mathbb{1}_{\{\Lambda(t)=\Lambda(s)\}}(N(t) - N(s))\right] = 0$ almost surely. Hence, $N(t) = N(s)\ \forall t > s$ for which $\Lambda(t) = \Lambda(s)$. □

Chapter 2

Hawkes processes

In the last chapter, we shed some light on point processes with the characterization of the conditional intensity function. Now, it is time to familiarize with Hawkes processes for the upcoming sections.

2.1 Introduction

In order to bridge last section with Hawkes processes, its definition is given. This shows that the univariate Hawkes process is a generalization of an homogeneous Poisson process (see Example 1.12), where the Poisson rate depends on time and its history:

Definition 2.1. Let $\{N(t) : t \geq 0\}$ be a counting process with associated filtration $\mathbb{H} = \{\mathcal{H}_t : t \geq 0\}$ which satisfies:

$$\mathbb{P}[N(t+h) - N(t) = m | \mathcal{H}_t] = \begin{cases} \lambda^*(t)h + o(h) & \text{if} \quad m = 1 \\ o(h) & \text{if} \quad m > 1 \\ 1 - \lambda^*(t)h + o(h) & \text{if} \quad m = 0 \end{cases}$$

where the conditional intensity function of the process, also known as *jump intensity*, can be written as:

$$\lambda^*(t) = \lambda + \int_0^t \mu(t-u) dN(u)$$

$\lambda > 0$ is known as *background intensity* and $\mu : \mathbb{R} \to [0, \infty)$ is the *excitation function* with $\mu(t) = 0$ if $t < 0$ -that is, it is causal- and $\mu \neq 0$ to avoid the trivial case -which corresponds to an homogeneous Poisson process-. Such a process $N(\cdot)$ is a *linear Hawkes process*.

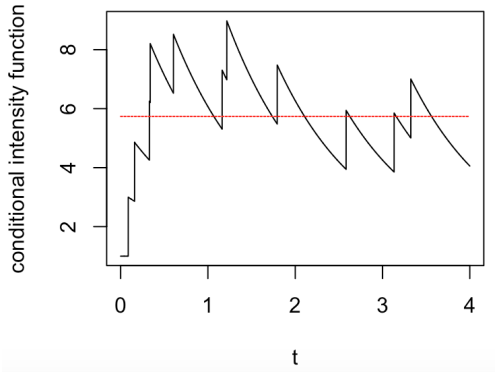

Figure 2.1: Representation of the conditional intensity function of a Hawkes process. In red, the expected value of the conditional intensity function.

It is interesting to mention that in the particular situation that the conditional intensity does not depend on the history, a non homogeneous Poisson process would be obtained, which its probability has the following form:

$$\mathbb{P}[N(b) - N(a) = k] = \frac{\left(\int_a^b \lambda(t)dt\right)^k}{k!} e^{-\int_a^b \lambda(t)dt}$$

To the purpose of being more familiar with the relation between the conditional intensity function and the time the jumps take place, we start off by giving the probability of the first jump. By using Equation (1.2):

$$\mathbb{P}(T_0 \leq s|\mathcal{H}_0) = 1 - \exp\left(-\int_0^s \lambda^*(u)du\right) =$$

$$= 1 - \exp\left(-\lambda s - \int_{-\infty}^0 [\mu(s-u) - \mu(-u)]dN(u)\right)$$

Being aware of the two types of point processes when considering the conditional intensity function, processes which are considered in this project are the self-exciting ones. However, an example of conditional intensity function for a self-regulated process is presented (examples of conditional intensity functions for self-exciting processes can be seen in Example 2.4 and 2.5):

Example 2.2. Let $\{T_1, T_2, ...\}$ be the times at which the jumps take place. Suppose the following expression for the conditional intensity function:

$$\lambda^*(t) = \exp\left(\mu t - \sum_{T_i<t} \alpha\right)$$

where $\mu, \alpha > 0$. Observe that the intensity increases as time draws on, but each time a new jump appears, it is multiplied by a constant $e^{-\alpha} < 1$ and thus, the probability of a new jump decreases immediately after a jump has appeared; that is, a self-regular point process.

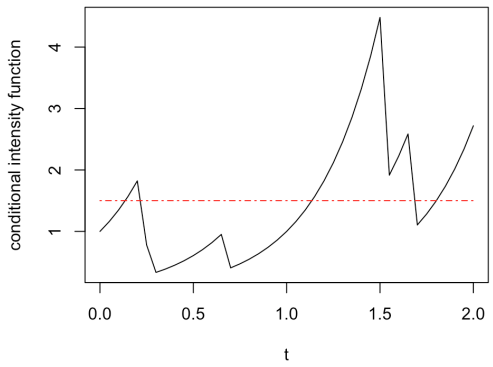

Figure 2.2: Representation of the conditional intensity function of a self-regulating process. Jumps have taken place at $T = \{0.2, 0.3, 0.7, 1.5, 1.7\}$ In red, the expected value of the conditional intensity function.

Rigorously, a self-exciting process is defined as, which is to be expected that the definition matches well with the one given in Chapter 1:

Definition 2.3. The process is said to be *self-exciting* if $Cov[N(b) - N(a), N(c) - N(b)] > 0$ where $0 < a \leq b < c$.

It is worth highlighting that, for self-exciting processes, the jump intensity is a stochastic process where each previous jump increases the jump intensity. This shows the dependency on the past, which is different from homogeneous and non homogeneous Poisson process as their distribution is memoryless. Moreover, by definition, instantaneous jumps are very improbable as seen in homogeneous Poisson process. Since a new jump depends on the last jump and the distribution of jumps is completely determined by the conditional intensity function, the pair $(N(t), \lambda^*(t))$ forms a Markov process[1] [20].

Remembering that the counting process and the arrival times are equivalent, one can express the conditional intensity function in terms of the past arrival times

[1]**Definition:** A *Markov process* is a collection $\{X(t)\}_{t \in \mathbb{R}^+}$ of random variables with values on a countable set E. Two conditions are imposed:

- Markov property: $\mathcal{P}[X(t+s) = j | X(u), 0 \leq u \leq t] = \mathcal{P}[X(t+s) = j | X(t)] \ \forall s, t \geq 0 \ i, j \in E$
- Homogeneity: $\mathbb{P}[X(t+s) = j | X(t) = i] = \mathbb{P}[X(s) = j | X(0) = i] \ s, t \geq 0, \ i, j \in E$.

Please see [19] for more details.

$\{T_1, T_2, ...\}$:

$$\lambda^*(t) = \lambda + \sum_{T_i<t} \mu(t - T_i)$$

Thus, the particular structure of the conditional intensity function is found when specifying the background intensity and the excitation function. Particularly, A.Hawkes in [26] considered two different self-exciting excitation functions: (i) an exponential decay and (ii) a potential decay. These examples are shown next:

Example 2.4. In the case of an exponential decay with parameters α and β, the excitation function has the particular form of: $\mu(t) = \alpha e^{-\beta t}$. Hence, this brings to the conditional intensity function:

$$\lambda^*(t) = \lambda + \int_{-\infty}^{t} \alpha e^{-\beta(t-s)} dN(s) = \lambda + \sum_{T_i<t} \alpha e^{-\beta(t-T_i)}$$

Notice that α gives information of the quantity that increases the conditional intensity when there is a new jump; whereas β gives information of the jump's influence decay from past to future events.

Example 2.5. For the potential decay, the excitation function is particularly expressed as $\mu(t) = \frac{k}{(c+(t-s))^p}$, which depends on three positive parameters: s, k and p. This leads to:

$$\lambda^*(t) = \lambda + \int_{-\infty}^{t} \frac{k}{(c + (t-s))^p} dN(s) = \lambda + \sum_{T_i<t} \frac{k}{(c + (t - T_i))^p}$$

When one wants to use Hawkes processes to model, it is necessary to fix an origin of time so as to know the impact of an event. For this reason, it is useful to consider an initial condition, $\lambda^*(t_0) = \lambda_0^*$, and by this, the conditional intensity function is the solution of a stochastic differential equation. Specifically, for Example 2.4, the stochastic differential equation is the following:

$$d\lambda^*(t) = \beta(\lambda - \lambda^*(t))dt + \alpha dN(t) \quad t \geq 0$$

whose solution is of the form:

$$\lambda^*(t) = e^{-\beta t}(\lambda_0 - \lambda) + \lambda + \int_0^t \alpha e^{\beta(t-s)} dN(s) \quad t \geq 0$$

2.1.1 m-variate Hawkes processes

Notation: The operation $*$ corresponds to the *convolution product*, which is defined as:

$$F * g(t) = \int_{-\infty}^{\infty} F(u) g(t - u) du$$

When refering to finance, sometimes there are some indicators that are preferible to be considered all together at the same time. For example, when one wants to model the mid-price changes by taking into account the best bid and the best ask (see Section 3.1 for more details).
This can be completely defined because Definition 2.1 is completely extendable to m dimensions by taking an m-variate counting process $N(t) = \{N_i(t) : t \geq 0\}_{i=1,...,m}$, where each $N_i(\cdot)$ has a different conditional intensity function $\lambda_i^*(\cdot)$ and each one represents a univariate Hawkes process. Thus, this is possible by considering collections of one-dimensional Hawkes processes in which a jump of one of these Hawkes process can increase the probability of a jump for another one, that is, they can excite one to another. Therefore, it is to be expected that the following definition matches well with the one given for the one dimensional process:

Definition 2.6. Let $\mathcal{N} = \{N_1(t), ..., N_m(t)\}$ be a collection of counting processes and let $T = \{T_{ij} : i \in \{1, ..., m\}, j \in \mathbb{N}\}$ be the random arrival times for each counting process. Supposing that the conditional intensity function of $N_i(\cdot)$ is:

$$\lambda_i^*(t) = \lambda_i + \sum_{j=1}^{m} \int_{-\infty}^{t} \mu_{ij}(t-u) dN_j(u) = \lambda + dN * \mu(t)$$

where $\lambda = \{\lambda_i\}_{i=1,...,m}$ is a vector of exogenous intensities with $\lambda_i > 0 \ \forall i$; and $\mu = \{\mu_{ij}\}_{i=1,...,m}$ is a matrix-valued kernel where $\mu_{ij} : \mathbb{R} \to [0, \infty)$ are L^1-integrable where $\mu(t) = 0 \ \forall t < 0$. Then, $\mathcal{N}$ is a *mutually exciting Hawkes process*.

Definition 2.7. The *conditional intensity matrix* of a m-variate Hawkes process $\{N_i(t) : t \geq 0\}$ is defined as:

$$\kappa_{ij}^* dt = E[dN_i(t) | dN_j(0) = 1] - \epsilon_{ij}\delta(t) - \Lambda_i dt$$

where Λ_i is the compensator of the Hawkes process $N_i(t)$ and ϵ_{ij} is the Kronoecker delta[2].

It is of interest to mention that μ has positive components -that is $\mu_i \geq 0 \ \forall i$- and causal components -that is, for $t < 0$, $\mu_i = 0 \ \forall i$-.
Additionally, by considering the arrival times of each counting process $t = \{T_{ij} : i \in \{1, ..., m\}, j \in \mathbb{N}\}$, the conditional intensity function is equivalently written as:

$$\lambda_i^*(t) = \lambda_i + \sum_{j=1}^{m} \sum_{T_{ij} < t} \mu_{ij}(t - T_{ij})$$

[2]**Definition:** The *Kronoecker delta* is defined as:

$$\epsilon_{ij} = \begin{cases} 1 & \text{if} \quad i = j \\ 0 & \text{if} \quad i \neq j \end{cases}$$

As for the univariate Hawkes process, an example follows:

Example 2.8. In the particular case where the excitation functions are exponential decays (expression equivalent to Example 2.4), the conditional intensity function is given by:

$$\lambda_i^*(t) = \lambda_i + \sum_{j=1}^{m} \int_{-\infty}^{t} \alpha_{ij} e^{-\beta_i(t-u)} dN_j(u) = \lambda_i + \sum_{j=1}^{m} \sum_{T_{jk}<t} \alpha_{ij} e^{-\beta_i(t-T_{jk})}$$

where α_{ij}, β_i are non-negative constants. This yields to the differential equation:

$$d\lambda_i^*(t) = \beta_i(\lambda_i - \lambda_i^*(t))dt + \sum_{j=1}^{m} \alpha_{ij} dN_j(t)$$

It must be emphasized that this equation takes into account the activity dependency on time, where many jumps will be clustered in a short period of time; and the dependency on the space of the counting processes, where an adverse shock in a market propagates to the others. So in the multivariate case, as in the univariate one, one can have a set of similar differential stochastic equations whose solutions are the components of conditional intensity function with exponential kernel of each process. Thus, equivalently, for the multivariate Hawkes process, the $(2m)$-uple $\{N(t), \lambda_1^*(t), ..., \lambda_m^*(t)\}$ is a Markov chain in the situation of exponential kernels, but not in the case of potential decays, for example.

It must be pointed out that the interest of introducing the multivariate Hawkes processes relies on the possibility of self-excitation and cross-section which gives feedback, that is, new jumps vary the probability of new jumps. Hence, when a first jump is occurred, and in the situation of having a mutually exciting process, the process is self-feeding and contagious. Furthermore, they introduce asymmetry: not all jumps have the same consequences.

2.2 Stationarity

In this section we will focus on stationary Hawkes processes which are of interest, since they are easy to study and because of their invariance along time. For example, in finance, processes that do not change in time have been studied because there exists an interest in different modelings of stable markets.
To this purpose, let us start off by giving the following definitions, which introduce this property are considered in the general case of a m-variate Hawkes process, and they can be particularized to one dimension by setting $m = 1$.

Definition 2.9. Let $\mathcal{N} = \{N_i(t) : t \geq 0\}_{i=1,...,m}$ be a m-variate Hawkes process, $\mathcal{N}$ is *stationary* if a translation in time does not change its distribution. That is, for $h > 0$:

$$\{N_1(t+h), N_2(t+h), ..., N_m(t+h)\} \sim \{N_1(t), N_2(t), ..., N_m(t)\}$$

Nevertheless, from Theorem 1 (unbounded Lipschitz dynamics) of [13], an equivalent definition of stationarity for a Hawkes process is considered in [22].

Definition 2.10. Let $F(t) = \{f^{ij}(t)\}_{i,j=1}^{m}$ be a square matrix of scalar functions. Its *spectral radius* is given by:

$$||F|| = max\left\{||k_i(t)|| : k_i \text{ eigenvalue of F(t) } i = 1, ..., m\right\} =$$

$$= max\left\{\int dt|k_i(t)| : k_i \text{ eigenvalue of F(t) } i = 1, ..., m\right\}$$

where the integral is taken over all possible values of t.

Specifically, for a univariate Hawkes process, the spectral ratio is known as *branching ratio*, which is given by:

$$n := \int_0^\infty |\mu(s)|ds$$

Definition 2.11. Let $\{N(t) : t \geq 0\}$ be a Hawkes process. We say it has asymptotically stationary increments and $\lambda^*(t)$ is *asymptotically stationary* if the kernel μ satisfies the *stability condition*, that is, $||\mu|| = \left\{||\mu_{ij}||\right\}_{i,j=1}^{m} < 1$.

Let's illustrate with examples the stability condition referring to the two possible excitation functions shown in Example 2.4 and 2.5.

Example 2.12. Imagine a bivariate simetric Hawkes process whose kernel is represented by an exponential decay. In this situation there are two type of excitations: (i) the self-excitation which corresponds to the influence on their probability when a jump occurs and (ii) the cross-excitation which takes into account the influence of one to the other. In this case, the kernel is of the form:

$$\mu = \begin{pmatrix} \mu_s(t) & \mu_c(t) \\ \mu_c(t) & \mu_s(t) \end{pmatrix}$$

where each component has the following form:

$$\mu_i(t) = \alpha_i e^{-\beta_i t} \mathbb{1}_{\{t>0\}}$$

with $\alpha_i, \beta_i > 0$. Note that, as in the one dimensional case, α_i gives information of the interaction strength and β_i controls the relaxation time of perturbations from

past to future jumps. In order to be the process stable, the stability condition must be satisfied:

$$||\mu|| = ||\mu_s + \mu_c|| < 1 \Leftrightarrow \frac{\alpha_s}{\beta_s} + \frac{\alpha_c}{\beta_c} < 1$$

as the eigenvalues of μ are $k_1 = \mu_s + \mu_c$ and $k_2 = \mu_s - \mu_c$.

Example 2.13. Suppose now the case of a univariate Hawkes process with a power law as excitation function:

$$\mu(t) = \frac{k}{(c + (t-s))^p} \mathbb{1}_{\{t>s\}} \equiv \mu(u) = \frac{k}{c^p (1 + u/c)^p} \mathbb{1}_{\{u>0\}}$$

with $k, c > 0$ and the last expression is the one that will be used from now. So as to satisfy the stability condition, the following criteria must be satisfied:

$$\int_0^\infty \frac{k/c}{(1 + t/c)^p} dt = \frac{k}{p-1} < 1 \Leftrightarrow k < (p-1)$$

This indicates that $p > 1$ as k is positive by definition; otherwise, the process is not stationary.

It is of interest to remark that for a self-exciting process, the value of the branching ratio, apart from giving a stability criteria, also gives a criteria of whether the process explodes [28]:

Proposition 2.14. Let n be the branching ratio of a Hawkes process. We say that the Hawkes process explodes if and only if $n \geq 1$.

Proof. Firstly, we start by defining the expected value of the conditional intensity function as:

$$g(t) = E[\lambda^*(t)] = E\left[\lambda + \int_0^t \mu(t-s) dN(s)\right] = \lambda + \int_0^t \mu(t-s) E[dN(s)] \quad (2.1)$$

where the last equality is because of Fubini's theorem[3]. On the other hand, by Definition 1.13:

$$\lambda^*(s) = \frac{E[dN(s)|\mathcal{H}(s)]}{ds}$$

[3]**Theorem:** *(Fubini's theorem)* Let $(\Omega, \mathcal{H}, \mathbb{P})$ be a σ-finite measurable space and $\{X(t)\}_{t \in T}$ a set of measurable real-valued random variables. Then:

$$E\left[\int_T X(t) d\mu(t)\right] = \int_T E[X(t)] d\mu(t)$$

if $X(t)$ is non-negative $\forall t \in T$ or if $\int_T E(|X(t)|) d\mu(t) < \infty$.

Then, if we calculate the expected value of this expression and use the tower property[4], it yields to:

$$g(s) = E[\lambda^*(s)] = \frac{E[E[dN(s)|\mathcal{H}(s)]]}{ds} = \frac{E[dN(s)]}{ds}$$

Therefore, by replacing this result in Equation (2.1), we have:

$$g(t) = E[\lambda^*(s)] = \lambda + \int_0^t \mu(t-s)g(s)ds = \lambda + \int_0^t g(t-s)\mu(s)ds = \lambda + \mu * g$$

which corresponds to a renewal equation[5]. Reasoning over the range of cases:

1. $n < 1$ (deflective case): By Asmussen's Proposition[6]:

$$g(t) = E[\mu^*(t)] \to \frac{\lambda}{1-n} \quad t \to \infty$$

2. $n > 1$ (excessive case) and $n = 1$: As the number of jumps each time can only be 0 or 1 because instantaneous events have been excluded, the following result holds:

$$\lambda^*(t) = \lambda + \sum_{t_i \le t;\, dN(t_i)=1} \int_0^t \mu(s)ds = \lambda + \sum_{t_i \le t;\, dN(t_i)=1} n \to \infty \quad \text{as } t \to \infty$$

   Thus, the process explodes.

□

2.3 First and second order

The possibility to know the first and second order allows to characterize the process. Particularly, thanks to the structure of the conditional intensity function, it is possible to characterize its properties in an analytical way which is what we are going to do all along this section.

[4] $E(X) = E(E(X|Y))$.

[5] **Definition:** A *renewal equation* is an equation of the form $u = a + (F * u)$ where $(F * u)(t) = \int_0^t F(t-s)du(s)$ and whose solution is given by $u(t) = (R * a)(t)$. $R(t)$ is the *renewal function*, which is defined as $R(t) = \sum_{n \ge 0} F^{*n}(t)$.

[6] **Proposition**: Suppose the renewal equation $Z = z + Z * F$. If in the deflective case z is bounded and $z(\infty) = \lim_{t\to\infty} z(t)$ exists, then

$$Z(t) \to \frac{z(\infty)}{1 - ||F||} = Z(\infty)$$

Proof: See proposition 7.4 in [2].

2.3.1 Covariance and Power spectral densities

Bearing in mind that when the branching ratio satisfies $n \in (0,1)$, the expected value $E[dN(t)]$ and the covariance $Cov[dN(t), dN(t+s)]$ do not depend on time since the process is stationary, it is interesting to see how the process behaves in the first and second order. Because of the result of Proposition 2.14 for the deflective case (see the proof), the expected value of the conditional intensity matrix for the univariate Hawkes process is constant:

$$g = g(t) = E[\lambda^*(t)] = \frac{E[dN(t)]}{dt} = \frac{\lambda}{1-n} \tag{2.2}$$

where $dN(t) = \lim_{h \to 0}[N(t+h) - N(t)]$.
In the case of the second order, it follows the next definition.

Definition 2.15. Let $N(t) = \{N_i(t) : t > 0, i \in \{1, ..., m\}\}$ be a m-variate Hawkes process and let $s > 0$, the *autocovariance density* is defined as:

$$R(s) = Cov\left(\frac{dN(t)}{dt}, \frac{dN(t+s)}{ds}\right) = \frac{E[dN(t+s)dN^T(t)]}{dtds} - g \cdot g^T$$

Although it is symmetric with respect to 0, $R(s) = R(-s)$, it cannot be extended to $\mathbb{R}$ as there is an atom[7] in 0 ($s > 0$). In the case of simple point processes, that is there are no multiple events (which means that the increments $dN(t)$ can be either 0 or 1), it results that:

$$E[(dN(t))^2] = E[dN(t)] = g\delta(s)$$

which would correspond to the case of $s = 0$. Hence, in order to define the covariance in all $\mathbb{R}$, the complete covariance density is proposed:

Definition 2.16. The *complete covariance density* can be defined as:

$$R^{(c)}(s) = g\delta(s) + R(s)$$

Note it is defined $R(0)$ so that the complete covariance density is continuous.

For continuous signals over time, such as an asymptotic stationary process, one important characteristic is the description of how the variations of these signals are distributed over frequency and how they change. This description allows to see how weak or strong are these changes. It therefore follows the next definition:

[7] **Definition:** An *atom* is a measurable set which has a positive measure and it does not contain a subset with a positive measure.

Definition 2.17. The *power spectral density function* is defined as:

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-is\omega} R^{(c)}(s)ds = \frac{1}{2\pi}\left[g + \int_{-\infty}^{\infty} e^{-is\omega} R(s)ds\right]$$

It is interesting to see the particular expression that takes the covariance density and the power spectral density for Hawkes processes. This informations is given in the following proposition:

Proposition 2.18. Particularly, for a stationary mutually exciting process, the covariance density matrix is given by:

$$R(s) = E\left[\left(\lambda + \int_{-\infty}^{t+s} \mu(t+s-u)dN(u)\right)\frac{dN^T(t)}{dt}\right] - g\cdot g^T$$

which gives a power spectral density matrix:

$$S(\omega) = \frac{1}{2\pi}\frac{1}{\mathbb{I} - \mathcal{L}\{\mu(t)\}(\omega)}\frac{g}{\mathbb{I} - \mathcal{L}\{\mu(t)\}(-\omega)}$$

Proof. See Section 2 in [25]. □

Note that in the particular case of a self-exciting process, this result is written as:

$$S(\omega) = \frac{g}{2\pi(1-\mathcal{L}\{\mu(t)\}(\omega))(1-\mathcal{L}\{\mu(t)\}(\omega))} = \frac{g}{2\pi|1-\mathcal{L}\{\mu(t)\}(\omega)|^2}$$

Let us give now the specification for an exponential decay:

Theorem 2.19. *(Hawkes process power spectral density)* Let $\{N(t) : t > 0\}$ be a univariate Hawkes process whose excitation function has the form of an exponential decay with $\alpha < \beta$ (see Example 2.4). Then, the covariance density for $s > 0$ is:

$$R(s) = \frac{\alpha\beta\lambda(2\beta-\alpha)}{2(\beta-\alpha)^2}e^{-(\beta-\alpha)s}$$

And its power spectra density function has the form of:

$$S(\omega) = \frac{\lambda\beta}{2\pi(\beta-\alpha)}\left(1 + \frac{\alpha(2\beta-\alpha)}{(\beta-\alpha)^2+\omega^2}\right)$$

Proof. By the definition of the covariance density for $s \in \mathbb{R}\setminus\{0\}$ we have that:

$$R(s) = Cov\left(\frac{dN(t)}{dt}, \frac{dN(t+s)}{ds}\right) = E\left[\frac{dN(t)}{dt}\frac{dN(t+s)}{ds}\right] - g^2$$

By using the tower property and Definition 1.13, it is obtained:

$$E\left[\frac{dN(t)}{dt}\frac{dN(t+s)}{ds}\right] = E\left[E\left[\frac{dN(t)}{dt}\frac{dN(t+s)}{ds}|\mathcal{H}(t+s)\right]\right] =$$

$$= E\left[\frac{dN(t)}{dt}E\left[\frac{dN(t+s)}{ds}|\mathcal{H}(t+s)\right]\right] = E\left[\frac{dN(t)}{dt}\lambda^*(t+s)\right] =$$

$$= E\left[\frac{dN(t)}{dt}\left(\lambda + \int_{-\infty}^{t+s}\mu(t+s-u)dN(u)\right)\right]$$

Therefore:

$$R(s) = E\left[\frac{dN(t)}{dt}\left(\lambda + \int_{-\infty}^{t+s}\mu(t+s-u)dN(u)\right)\right] - g^2 = \lambda g +$$

$$+E\left[\frac{dN(t)}{dt}\left(\int_{-\infty}^{t+s}\mu(t+s-u)dN(u)\right)\right] - g^2 = g\lambda +$$

$$+E\left(\int_{-\infty}^{s}\mu(s-v)\frac{dN(t)}{dt}\frac{dN(t+v)}{dv}dv\right) - g^2 = \lambda g +$$

$$+\int_{-\infty}^{s}\mu(s-v)E\left[\frac{dN(t)}{dt}\frac{dN(t+v)}{dv}\right]dv - g^2 = \lambda g + \int_{-\infty}^{s}\mu(s-v)(R^{(c)}(v)+g^2)dv -$$

$$-g^2 = \lambda g + g\mu(s) + \int_{-\infty}^{s}\mu(s-v)R(v)dv + ng^2 - g^2 =$$

$$= g\mu(s) + \int_{0}^{\infty}\mu(s+v)R(v)dv + \int_{0}^{s}\mu(s-v)R(v)dv$$

where n is the branching ratio. Now, taking the Laplace transformation[8] (for more details see appendix A.2 of [28]), it is possible to solve this equation and by taking into account that the excitation function is an exponential decay, we have that:

$$\mathcal{L}\{R(s)\}(r) = \frac{\alpha g(2\beta-\alpha)}{2(\beta-\alpha)(r+\beta-\alpha)} = \frac{\alpha\beta\lambda(2\beta-\alpha)}{2(\beta-\alpha)^2(r+\beta-\alpha)}$$

By inverting this expression, the result that we were looking for is obtained:

$$R(s) = \frac{\alpha\beta\lambda(2\beta-\alpha)}{2(\beta-\alpha)^2}e^{-(\beta-\alpha)s}$$

Moreover, by taking the definition of the power spectra density function, one obtains:

[8] **Definition:** Let $f:\mathbb{R}\to\mathbb{R}$ be a L^1-scalar function, then its *Laplace transformation* is defined as:

$$\mathcal{L}\{f(t)\} = \int_{-\infty}^{\infty}dtf(t)e^{zt}$$

All Laplace transformations in this section and the next one have been calculated using [34].

$$S(\omega) = \frac{1}{2\pi}\left[g + \int_0^\infty e^{-is\omega}R(s)ds + \int_0^\infty e^{i\omega s}R(s)ds\right] = \frac{1}{2\pi}[g + \mathcal{L}\{R(s)\}(i\omega) +$$

$$+\mathcal{L}\{R(s)\}(-i\omega)] = \frac{\lambda\beta}{2\pi(\beta-\alpha)}\left(1 + \frac{\alpha(2\beta-\alpha)}{(\beta-\alpha)^2+\omega^2}\right)$$

□

2.3.2 Kernel inversion

Throughout this section, we are going to find some general results for a mutually exciting process by giving an alternative characterization of the first and second order. This characterization is based on the Laplace transformation which will make some problems easier and the kernel inversion. Therefore, we have the following definition:

Definition 2.20. Let $\{N(t) : t \geq 0\}$ be an asymptotically stationary Hawkes process, then the *Kernel inversion* $\Phi(t)$ is denoted as the solution of the following renewal equation:

$$\mu(t) + \Phi(t) * \mu(t) = \Phi(t)$$

As a consequence of the asymptotically stationarity, $\Phi(t)$ exists and it can be expressed as an infinite sum of convolutions:

$$\Phi(t) = \mu(t) + \mu(t) * \mu(t) + \mu(t) * \mu(t) * \mu(t) + ... = \sum_{i\geq 1} \mu^{*i}(t)$$

It is worth highlighting that this solution could have also be obtained by seeing that the equation is a renewal equation. Then, its Laplace transformation is given by:

$$\mathcal{L}\{\Phi(t)\}(z) = \frac{1}{\mathbb{I} - \mathcal{L}\{\mu(t)\}(z)} - \mathbb{I}$$

where $\mathbb{I}$ is the identity matrix.
It is also interesting to mention that by the definition of the Laplace transformation, one has an alternative expression for the spectral radius:

$$\mathcal{L}\{\mu\}(0) = \int_{-\infty}^{\infty} dt\mu(t) = ||\mu||$$

Moreover, with the kernel inversion, it is possible to acquire an alternative representation for the first and second order, which is shown in the following result [4]:

Proposition 2.21. *(First and Second order statistics)* Let $\{N(t) : t \geq 0\}$ be an asymptotically stationary Hawkes process. Then:

- The *average intensity* g is given by:

$$g = \frac{E[dN(t)]}{dt} = (\mathbb{I} + \mathcal{L}\{\Phi(0)\})\lambda = \frac{\lambda}{\mathbb{I} - ||\mu||}$$

- The autocovariance density matrix given in Definition 2.15, has as Laplace transformation:

$$\mathcal{L}\{R\}(z) = (\mathbb{I} + \mathcal{L}\{\Phi\}(-z))\Sigma(\mathbb{I} + \mathcal{L}\{\Phi\}^T(z))$$

  where Σ is a diagonal matrix whose non-zero elements are given by $\Sigma_{ii} = g_i$, that is the average intensity of the stationary Hawkes process $\{N_i(t) : t \geq 0\}$.

Notice that the expression of the average intensity is completely in accordance with Equation (2.2). With this result, a linear prediction is obtained providing that one has the conditional intensity function.
Now we are ready to see how this result can help on finding the first and second order with two examples:

Example 2.22. Suppose an asymptotically stationary bivariate Hawkes process with exponential decay like the one of Example 2.12. Then, the Laplace transformation of the excitation function can be written equivalently as:

$$\mathcal{L}\{\mu\}(z) = \frac{1}{2}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} \mathcal{L}\{\mu_s\}(z) + \mathcal{L}\{\mu_c\}(z) & 0 \\ 0 & \mathcal{L}\{\mu_s\}(z) - \mathcal{L}\{\mu_c\}(z) \end{pmatrix}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$$

where

$$\mathcal{L}\{\mu_s\}(z) = \frac{\beta_s\alpha_s}{\beta_s - z} \qquad \mathcal{L}\{\mu_c\}(z) = \frac{\beta_c\alpha_c}{\beta_c - z}$$

Observe that from the first representation the kernel matrix is diagonal in the basis of the symmetric and antisymmetric combinations $N^{\pm}(t) = \frac{1}{\sqrt{2}}(N_1(t) \pm N_2(t))$. Therefore, in this basis, by assuming that $\lambda = (\lambda_0, \lambda_0)$, the average intensity has both components equal to g_0:

$$g_0 = \frac{\lambda_0}{1 - \alpha_s - \alpha_c}$$

And the correlation matrix in this basis is of the form:

$$R^{\pm}(s) = \frac{E[(dN_1(t) \pm dN_2(t))(dN_1(t+s) \pm dN_2(t+s))]}{2dtds} - g_0^2$$

whose Laplacian transformation is:

$$\mathcal{L}\{R\}(z) = [\mathbb{I} + \mathcal{L}\{\mu\}(-z)]^{-1}\Sigma[\mathbb{1} - \mathcal{L}\{\mu\}(z)]^{-1}$$

After replacing by the Laplacian of the excitation function, one obtains:

$$\mathcal{L}\{R^{\pm}\}(z) = \frac{g_0}{(1 - \mathcal{L}\{\mu_s\}(-z) \mp \mathcal{L}\{\mu_c\}(-z))(1 - \mathcal{L}\{\mu_s\}(z) \mp \mathcal{L}\{\mu_c\}(z))}$$

Example 2.23. Suppose now a univariate Hawkes process with a power law excitation function as the one Example 2.13. In this situation, the Laplace transformation of the excitation function is given by:

$$\mathcal{L}\{\mu\}(z) = kc^{-p+1}e^{-cz}(-cz)^{p-1}\Gamma(-p+1, -cz)$$

where $\Gamma(n, m) = \int_m^\infty t^{n-1}e^{-t}dt$ is the incomplete Gamma function. The Laplace transformation of the Kernel inversion is:

$$\mathcal{L}\{\Phi\}(z) = \frac{kc^{-p+1}e^{-cz}(-cz)^{p-1}\Gamma(-p+1, -cz)}{1 - kc^{-p+1}e^{-cz}(-cz)^{p-1}\Gamma(-p+1, -cz)}$$

As seen before, the Hawkes process is asymptotically stationary if and only if $k < p - 1$, which leads to an average intensity of:

$$g = \frac{\lambda}{1 - ||\mu||} = \lambda\left(\frac{p-1}{p-1-k}\right)$$

Note that the average intensity is between a total intensity equals to the exogeneous one ($k = 0$); that is the non-interacting case, and an increasingly larger number of events ($k \sim p - 1$); which corresponds to a process near the instability point.
Then, in this example the Laplace transformation of the covariance matrix is:

$$\mathcal{L}\{R\}(z) = \frac{g}{(1 - \mathcal{L}\{\mu\}(z))(1 - \mathcal{L}\{\mu\}(z))}$$

This expression cannot be inverted analytically[9].

Additionally, there exists an alternative way to find the covariance matrix, which is by taking into account the conditional intensity matrix (Definition 2.7), which is shown in the next proposition:

[9]However, for $z << 1$, this equation can be inverted by using Tauberian theorems (theorems that gives conditions in order to sum divergent series) [5].

Proposition 2.24. There exists a relation between the conditional average and the correlation matrix for an asymptotically stationary m-variate Hawkes process:

$$R(t) = \Sigma\kappa^T(t) \ \forall t > 0$$

where Σ is a diagonal matrix with non-zero elements equal to $\Sigma_{ii} = g_i$, being g_i the average intensity of the process N_i and $\kappa^*(t)$ is the conditional intensity matrix.

Proof. See Section II.B of [4]. □

The reason behind introducing the first and second order with its different expressions is that they completely characterize a Hawkes process by using a Wiener-Hopf equation[10]. Accordingly, the following result is shown [8]:

Theorem 2.25. *(Wiener-Hopf equation)* Let $\{N(t) : t > 0\}$ be an asymptotically stationary Hawkes process. Then, the matrix function $\chi(t) = \mu(t)$ is the unique solution of the Wiener-Hopf system:

$$\kappa(t) = \chi(t) + \chi(t) * \kappa(t) \quad \forall t > 0$$

Proof. The uniqueness can be seen in Appendix A of [8]. □

This result has a lot of importance since, when the average intensity $g(t)$ and the conditional intensity matrix $\kappa(t)$ are fixed, it shows the uniqueness of a Hawkes process for these given observables. However, this process does not always exist as it cannot reproduce the linear properties for systems in which inhibition is important, that is, self-regulating processes [5].
Thus, whereas this result states that correlations and average intensities uniquely fix the interactions; Proposition 2.21 states that by fixing the excitation function and the exogeneous intensity, the correlation is uniquely determined.

It is also remarkable to see what happens when the background intensity function depends on time, $\lambda(t)$. In this situation [4]:

[10] **Definition:** A *Wiener-Hopf equation* corresponds to an integral equation on $[0, \infty)$ that depends on the difference of two arguments:

$$u(x) - \int_0^\infty k(x-s)u(s)ds = f(x) \quad \forall x \in [0, \infty)$$

.

Theorem 2.26. Let $\{N_1, ..., N_m\}$ be an asymptotically stationary mutually exciting process. Assuming that $h(t) = \int_0^t \lambda(s)ds < \infty \quad \forall t > 0$, one has:

$$E[N(t)] = h(t) + \int_0^t \sum_{n\geq 1} \mu^{*n} h(s)ds$$

where $\mu^{*n} = \mu * ... * \mu$ (n times).

Proof. Note that $\sum_{n\geq 1} \mu^{*n}$ is L^1-integrable. By induction, it is obtained that:

$$\int_0^\infty \mu^{*n}(t)dt = ||\mu||^n$$

and as the Hawkes process is asymptotically stationary, $\sum_{n\geq 1} K^n$ is finite when $0 < K < 1$, it yields:

$$\int_0^\infty \sum_{n\geq 1} \mu^{*n}(t)dt = \frac{||\mu||}{\mathbb{1} - ||\mu||}$$

Additionally, using the Doob-Meyer decomposition given in Equation (1.3), it is obtained:

$$E[N(t)] = E\left[\int_0^t \lambda^*(s)ds\right] = E\left[\int_0^t \lambda(s)ds\right] + E\left[\int_0^t ds \int_0^s \mu(s-u)dN(u)\right]$$

Considering Fubini's theorem:

$$\int_0^t ds \int_0^s \mu(s-u)dN(u) = \int_0^t \left(\int_u^t \mu(s-u)ds\right) dN(u) = \int_0^t \left(\int_0^{t-u} \mu(x)dx\right) dN(u)$$

After integrating by parts, the final result is obtained:

$$E[N(t)] = E\left[\int_0^t \lambda(s)ds\right] + E\left[\int_0^t \mu(t-s)N(s)ds\right] = \int_0^t \lambda(s)ds + \int_0^t \mu(t-s)d(E[N(s)])$$

where in the last equality the Fubini's theorem has been used. Observe that this corresponds to a renewal equation, hence the solution is:

$$E[N(t)] = \left(\sum_{n\geq 0} \mu^{*n}(t) * h(t)\right) = h(t) + \int_0^t \sum_{n\geq 1} \mu^{*n} h(s)ds$$

For more details, see Theorem 1 in [4]. □

It is to be expected that this definition matches well with Equation (2.2) when $\lambda(t) = \lambda$.

2.4 Representation of a Hawkes process

In this section different ways of representing a Hawkes process are presented, which will lead to different views of seeing them.

2.4.1 Immigrant-Birth representation

The population of a country can vary because of births, deaths or immigration. Suppose that the arrival of immigrants can be modelled by an homogeneous Poisson process of rate λ. Additionally, an individual can produce $m \in \mathbb{N}$ descendants independently from other individuals. Hence, we can mind that the number of births is modelled by an inhomogeneous Poisson process: imagine an individual that enters in $t_i \in \mathbb{R}^+$, the rate at which they produce offspring is $\mu(t - t_i)$ for $t > t_i$.
Denote by Z_i the number of offspring in the i^{th} generation and suppose the case where initially there is only one individual, $Z_0 = 1$. Then, the first-generation of offspring corresponds to a Poisson process of parameter n, that is, $Z_1 \sim Pois(n)$. In this situation:

$$n = \int_0^\infty dt\mu(t) = \int_0^\infty \alpha e^{-\beta s} ds = \frac{\alpha}{\beta}$$

Conditioning on the knowledge of Z_1, the ratio of offspring of an immigrant is i.i.d with density $\frac{\mu(t-t_i)}{n}$. This corresponds to the rate of offspring per person over the frequency of occurrence that there is offspring.
It is now apparent the equivalence between this branching representation[11] and a Hawkes process, since each new individual regardless of its origin (birth or immigration) can be comparable to a new jump in a Hawkes process.

With this representation in mind, we can derive an asymptotic characteristic [28]:

Theorem 2.27. *(Hawkes process with asymptotic normality)* Let $\{N(t) : t \geq 0\}$ be an asymptotically stationary univariate Hawkes process whose conditional intensity function is given by Definition 2.1. If

$$\int_0^\infty s\mu(s)ds < \infty$$

[11] **Definition:** A *branching process* is a Markov chain $\{X(n) : n \in \mathbb{N}\}$ on $\mathbb{N}$. Its transition function is defined by a given probability function $\{p_k : k \in \mathbb{N}\}$, with $p_k \geq 0$ and $\sum p_k = 1$ by

$$\mathbb{P}[X(n+1) = j | X(n) = i] = \begin{cases} p_j^{*i} & \text{if} \quad i \geq 1 \quad j \geq 0 \\ \delta_{0j} & \text{if} \quad i = 0 \quad j \geq 0 \end{cases}$$

where δ_{ij} is the Kronecker delta and $\{p_k^{*i} : k \in \mathbb{N}\}$ is the i-fold convolution of $\{p_k : k \in \mathbb{N}\}$ [3].

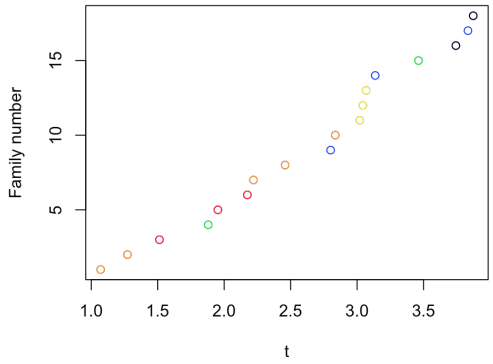

Figure 2.3: Example of a representation of the arrival of new population. Different colours represent offspring from different immigrants, which in this case there are six different immigrants.

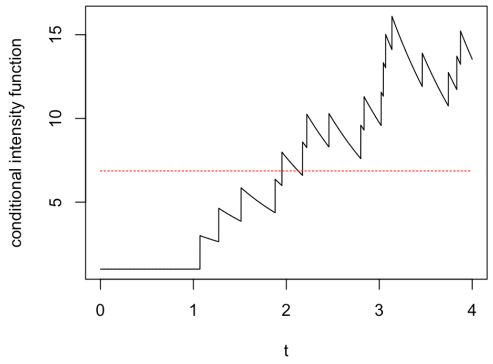

Figure 2.4: Corresponding conditional intensity function of the immigrant representation shown in Figure 2.3. In red the expected value of the conditional intensity function.

then the number of jumps in t is asymptotically normally distributed, that is:

$$\mathbb{P}\left(\frac{N(t)-\lambda t/(1-n)}{\sqrt{\lambda t/(1-n)^3}} \leq x\right) \to \Phi(x) \quad \text{when } t \to \infty$$

where in this case $\Phi(x)$ corresponds to the cumulative distribution function of the normal distribution with mean 0 and standard deviation 1.

Proof. See Theorem 2 in [26]. □

Going back to the immigrant representation, we have the following property:

Proposition 2.28. Under the immigrant representation and supposing that $Z_0 = 1$, the i^{th}-generation satisfies:

$$E[Z_i] = n^i$$

Proof. Induction will be done in order to proof this result:
For $i = 0$, as there is one immigrant, $E[Z_0] = E[1] = 1 = n^0$.
Suppose that for i the proposition is satisfied, $E[Z_i] = n^i$, then:

$$E[Z_{i+1}] = E[Z_1 \cdot Z_i] = E[Z_1] \cdot E[Z_i] = n \cdot n^i = n^{i+1}$$

where in the first equality, it has been used that $Z_i = Z_{i-1} \cdot Z_i$, since the number of individuals in the i^{th}-generation is equal to the number of individuals in the $(i-1)^{th}$-generation multiplied by the 1^{st}-generation that each of these individual can have, as offspring of different individuals is independent. The second equality has been obtained by the independence of each individual. Finally, in the third equality, the induction hypothesis has been used. □

Hence in the case of one immigrant ($Z_0 = 1$), the expected number of descendants is:

$$E\left[\sum_{i=1}^{\infty} Z_i\right] = \sum_{i=1}^{\infty} E[Z_i] = \sum_{i=1}^{\infty} n^i = \begin{cases} \frac{n}{1-n} & \text{if} \quad n < 1 \\ \infty & \text{if} \quad n \geq 1 \end{cases}$$

It is worth highlighting that in the situation of $n \in (0, 1)$, which corresponds to the asymptotically stationary situation, this result can be seen as the ratio between the number of descendants for one immigrant and the number of members in their family.

2.4.2 Cluster representation

In the general case of a m-variate spatial Hawkes process, the cluster interpretation is similar to the immigrant representation but jumps represent a spatial location in space.
Cluster processes are point processes which are built conditioning on the realization of a center process, generally a Poisson point process. It turns out that the process can be seen as a cluster process in which the process of cluster centers $N_c(t)$ is the Poisson process of rate λ formed by the arrival of immigrants. Thus, for each point of $N_c(t)$, one has a cluster corresponding to the descendants of all generations of the immigrant.
First of all, we will see the existence of a *Poisson cluster process*[12] given a self-exciting process:

[12]**Definition:** A *Poisson cluster process* corresponds to a compound Poisson process where each point is replaced by a cluster of points and the original point is seen as the center of the cluster. The cluster sizes are independent and identically distributed and the cluster center is not observed.

Lemma 2.29. Given a univariate Hawkes process, if the background intensity $\lambda > 0$ and the branching ratio satisfies $n \in (0,1)$, then there exists a Poisson cluster process with rate $\lambda' = \frac{\lambda}{1-n}$.

Proof. See Lemma 1 of [26]. □

It is worth mentioning that the Poisson rate corresponds to the average intensity of the process. Notice that this lemma shows the existence of a self-exciting process given two parameters: the background intensity and the branching ratio. Additionally, this representation allows to see the uniqueness of the process: if there exists two process with the same conditional intensity, they must come from the same distribution. This fact is shown by the following lemma, which is a particularization of Lemma 1.14 for a Hawkes process:

Lemma 2.30. There exists at most one asymptotically stationary orderly point process with finite rate whose intensity function is given by Definition 2.1 and the branching ratio satisfies $n \in (0,1)$.

Proof. See Lemma 2 of [26]. □

The reason behind introducing these two results is that they allow to build self-exciting processes using Poisson cluster processes following the next idea.
Let N_c be a Poisson point process with Poisson rate λ_c, which is the starting point (in the immigrant representation, it would represent an immigrant that appears spontaneously). At each center point $t_c \in \mathbb{R}$ of N_c, a point process is generated independently $N(\cdot|t_c)$. So, the cluster process N is defined as the set of all immigrants and offspring generated independently at each center point t_c of N_c: $N(A) = N_c(N(A|\cdot)) \quad \forall A \in \mathcal{B}(\mathbb{R})$.
The following proposition gives the proceed of how a Hawkes process can be built following this representation [5]:

Proposition 2.31. *(Clustering representation)* Let $n \in \mathbb{N}$ and let $[0,\tau]$ be a time interval, not necessarily finite, in which we define a sequence of events according to the following steps:

1. For each $1 \leq i \leq n$, consider $\{t_m^{(0)}, i\}_{m=1}^{M_i}$ which is a set of events built from an homogeneous Poisson process of rate λ_i in the interval $[0,\tau]$.

2. For each event of type j, $(t_m^{(0)}, j)$ and for each $1 \leq i \leq n$, a new sequence of first-generation events $\{t_m^{(1)}, i\}_{m=1}^{M_i^{(1)}}$ is generated with Poisson rate $\mu_{ij}(t - t_m)$ in the interval $[t_m^{(0)}, \tau]$.

3. Repeat (2) from $n-1$ to generation n to obtain the sequence $\{t_m^{(n)}, k_m^{(n)}\}_{m=1}^{M^{(n)}}$ until more events in the time interval $[0,\tau]$ cannot be generated.

Then, the events

$$\{t_m, k_m\}_{m=1}^{M} = \bigcup_{n=0}^{\infty} \{t_m^{(n)}, k_m^{(n)}\}_{m=1}^{M^{(n)}}$$

are the ones that generate the Hawkes process in $[0,\tau]$.

2.4.3 Martingale representation

Hawkes processes can also be represented in terms of martingales by considering the Doob-Meyer decomposition (Equation (1.3)); which is the background of the following theorem [4]:

Theorem 2.32. *(Martingale representation)* Let $\{N_i(t) : t \geq 0, 1 \leq i \leq m\}$ be a m-variate Hawkes process, the stochastic process defined as:

$$M(t) = N(t) - \Lambda(t)$$

where $\Lambda(t)$ is the compensator, is a martingale with respect to the canonical filtration[13] of the Hawkes process. Moreover, if the Hawkes process is asymptotically stationary, then the conditional intensity function can be denoted by:

$$\lambda^*(t) = \lambda + \int_0^t \Phi(t-s)\lambda ds + \int_0^t \Phi(t-s) dM(s)$$

by using the summation representation of the kernel inversion.

Particularly, in the asymptotic regime ($t >> 1$), the last result is simplified by:

$$\lambda^*(t) \to g + \int_0^t \Phi(t-s) dM(s)$$

because of the average intensity's definition.
Thanks to this result, the average intensity can be calculated following two different ways:

[13] **Definition:** The *canonical filtration* for a random variable $\{X(t) : t \geq 0\}$ is the filtration

$$F_t^X = \sigma(X(u) : 0 \leq u \leq t)$$

This is the smallest filtration for which $\{X(t) : t \geq 0\}$ is adapted.

Example 2.33. Consider an asymptotically stationary Hawkes process. Then:

$$E[\lambda^*(t)|\mathcal{H}_s] = \lambda + \int_0^s \mu(t-u)dN(u) + \int_s^t du\mu(t-u)E[\lambda^*(u)|\mathcal{H}_s]$$

Using the martingale representation, one obtains:

$$E[\lambda^*(t)|\mathcal{H}_s] = \lambda + \int_0^t \Phi(t-s)\lambda ds + \int_0^s \Phi(t-s)dM(s)$$

where it has been used that the expected value of the martingale is zero.

This representation is suitable for high-frequency data modeling because of its structure.

2.5 Extensions

During this section we will present extensions of linear Hawkes processes, which provide different characteristics.

2.5.1 Marked Hawkes process

Marked Hawkes process are quite used in finance modelling such as in modeling commerce with different volumes [15]. These processes are defined by using a mark variable:

Definition 2.34. Let $\{N(t) : t \geq 0\}$ be a Hawkes process where for each arrival time t_{ij}, it is considered a mark variable ξ_{ij}. Then, the conditional intensity function of a *marked Hawkes process* is written as:

$$\lambda_i^*(t) = \lambda_i + \sum_{j=1}^{d} \mu_{ij}(t - t_{ij}, \xi_{ij})$$

It is worth mentioning that the fact of labeling with different marks gives a certain dynamism in the conditional intensity function, as different marks have different effects in future jumps.
In general, it is assumed that marks of different events are i.i.d. random variables and they come from the same distribution $f(\xi)$. Typically, the choice of the excitation function is such that the two variables are independent, i.e, the excitation function can be factorized as $\mu_{ij}(t, \xi) = \mu_{ij}(t)\chi_{ij}(\xi)$.

Example 2.35. Let $\xi_i \geq 0$ be a variable that denotes the magnitude of an earthquake which occurs at time t_i. In order to model this situation, a possible conditional intensity function that can be taken is of the form [23]:

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

where the parameters $\alpha, \beta, \gamma > 0$ and the mark density follows an exponential distribution, that is:

$$f^*(\xi|\mathcal{H}_t) = \delta e^{-\delta\xi}$$

As these marks do not depend on history, they are unpredictable.
An alternative representation of the conditional intensity function is considering the time and the marks at the same time:

$$\lambda^*(t,\xi) = \left(\lambda + \alpha \sum_{t_i<t} e^{\beta\xi_i} e^{-\gamma(t-t_i)}\right) \delta e^{-\delta\xi}$$

The idea behind this model is that earthquakes cause aftershocks, this is why there is the factor $\alpha e^{\beta\xi_i}$, which increases intensity. Moreover, this model makes a relation between the intensity of the earthquake and the probability of having another one.

2.5.2 Quasi-stationary Hawkes process

Another possible extension is the dependence on time of the background intensity $\lambda_i(t)$ as in Theorem 2.26. In particular, it allows to model a non-stationary system in which the excitation matrix is independent of time. This type of extension is useful when modeling the intra-day seasonalities or the side effects in consecutive days [5], for example.

Suppose the non asymptotically stationary regime, that is, a process whose $||\mu|| \geq 1$. While in the case of $||\mu|| > 1$, the average intensity grows exponentially in time, in $||\mu|| = 1$ the process can have a finite average event rate. This situation corresponds to *quasi-stationarity*. From this type of extension, two important results are obtained [14]:

Proposition 2.36. Consider a univariate Hawkes process $\{N(t) : t \geq 0\}$, such that $||\mu|| = 1$ and $\lambda = 0$. If

$$\int_0^\infty dt t\mu(t) < \infty$$

then the average conditional intensity is either 0 or $+\infty$. This property is known as the *degeneracy of critical, short-range Hawkes.*

Thus, in the univariate case, two trivial cases are obtained: (i) for $g = 0$ there are no events and (ii) for $g = \infty$ there is an explosion.

Theorem 2.37. *(Existence of critical stationarity)* Suppose a quasi-stationary Hawkes process such that $\lambda = 0$ and

$$\sup_{t\geq 0} t^{1+\gamma}\mu(t) \leq R \qquad \lim_{t\to\infty} t^{1+\gamma}\mu(t) = r$$

where $r, R > 0$ and $\gamma \in (0, 1/2)$. Then, the average conditional intensity is finite.

2.5.3 Non-linear Hawkes process

This extension is presented so as to notice that not all Hawkes processes present the same structure. For this reason, we shall briefly introduce them:

Definition 2.38. Let $\{N(t) : t \geq 0\}$ be a counting process whose conditional intensity function is given by:

$$\lambda_i^*(t) = \Psi\left(\sum_{j=1}^{m} \int_{-\infty}^{t} \mu_{ij}(t-u)dN(u)\right)$$

where $\Psi : \mathbb{R}^m \to [0,\infty)$, $\mu_{ij} : \mathbb{R}^m \to [0,\infty)$. Then, $N(t)$ is a *non-linear Hawkes process*.

Note that if $\Psi(x) = \lambda_i + x$, $N(t)$ is the linear Hawkes process described in Definition 2.6.
However, this extension is not used a lot because of its complexity and for being quite recent. Nevertheless, it is worth showing an example on a possible application:

Example 2.39. Considering the one dimensional situation with $\Psi(x) = k\mathbb{1}_{[0,K-1/2]}(x)$ and $\mu(x) = \mathbb{1}_{[0,b]}(x)$, it is obtained the $M/M/K/0$ queue input. This corresponds to a queue with Poisson jumps of intensity $k > 0$, service time $b > 0$, no waiting room and K servers.

Additionally, in [13] some applications to the neuron modelling are presented. The interest of applying these processes rely on the existence of asymptotically stationary non-linear Hawkes processes which is given by the following theorem:

Theorem 2.40. *(Increasing kernel)* Suppose a nonnegative, nondecreasing and left-continuous function $\Psi(x)$ such that

$$\Psi(x) \leq \lambda + \alpha x \quad x \in \mathbb{R}$$

where $\alpha \geq 0$. Let $\mu : \mathbb{R} \to \mathbb{R}^+$ be a casual function that satisfies

$$\alpha \int_{\mathbb{R}} \mu(t)dt < 1$$

Then, there exists an asymptotically stationary non linear Hawkes process defined by $\Psi(x)$.

2.6 Parameter estimation

In this section, the estimation of the parameters that Hawkes processes depend on is done by knowing a set of arrival times $\{t_1, t_2, ..., t_n\}$. Specially, the method that is explained corresponds to the *maximum likelihood estimation* which consists of finding the likelihood function[14] and then, estimating the model parameters by maximizing this function [28].

Theorem 2.41. *(Hawkes process likelihood)* Consider a univariate Hawkes process in a finite time interval $\{N(t) : t \in [0, \tau]\}$ and let $\{t_1, t_2, ..., t_k\}$ be the arrival times in this temporal interval. Then, the likelihood is written as:

$$L(k|\mathcal{H}_\tau) = \left[\prod_{i=1}^{k} \lambda^*(t_i)\right] \exp\left(-\int_0^\tau \lambda^*(u)du\right)$$

Proof. Assume that the process is observed up to the k^{th} arrival, then the joint density function is defined as:

$$L(k|\mathcal{H}_\tau) = f(t_1, t_2, ..., t_k) = \prod_{i=1}^{k} f^*(t_i|\mathcal{H}_{t_{i-1}})$$

because the jumps are i.i.d.. Using Equation 1.2 one has:

$$F^*(t|\mathcal{H}_s) = 1 - \exp\left(-\int_{t_k}^{t} \lambda^*(u)du\right) \qquad f^*(t|\mathcal{H}_s) = \lambda^*(t) \exp\left(-\int_{t_k}^{t} \lambda^*(u)du\right)$$

By replacing this in the likelihood expression, it yields:

$$L(k|\mathcal{H}_\tau) = \prod_{i=1}^{k} f^*(t_i|\mathcal{H}_s) = \left[\prod_{i=1}^{k} \lambda^*(t_i)\right] \exp\left(-\int_0^{t_k} \lambda^*(u)du\right)$$

Suppose now that the observation is done in a subinterval $[0, t_k] \subset [0, \tau]$. In this situation, the likelihood includes the probability of not seeing any jump in the interval $(t_k, \tau]$, which means:

$$L(k|\mathcal{H}_\tau) = \left[\prod_{i=1}^{k} f^*(t_i)\right] (1 - F^*(\tau|\mathcal{H}_s)) = \left[\prod_{i=1}^{k} \lambda^*(t_i)\right] \exp\left(-\int_0^\tau \lambda^*(u)du\right)$$

□

[14] **Definition:** The *likelihood* of a finite realization of a Poisson process is the probability of having a given number of observations in a time period conditioning on the parameters that depends the model, that is:

$$L(N(t)|\mathcal{H}_t) = \mathbb{P}[\mathcal{H}_t|N(t)]$$

Once the likelihood is obtained, we have to maximize it in order to estimate the parameters. This maximization would be done depending on the type of likelihood, for example by calculating the derivatives and using numerical methods (for example, Newton-Raphson method).

Example 2.42. In the particular case of having a univariate Hawkes process with exponential decay, it is easier to work with the log-likelihood which, considering the interval $[0, t_k]$, is written as:

$$l = \sum_{i=1}^{k} \log(\lambda^*(t_i)) - \int_0^{t_k} \lambda^*(u)du = \sum_{i=1}^{k} log(\lambda^*(t_i)) - \Lambda(t_k)$$

Using Definition 1.16 and dividing the integral in different subsets, it yields to:

$$\Lambda(t_k) = \int_0^{t_k} \lambda^*(u)du = \int_0^{t_1} \lambda^*(u)du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda^*(u)du =$$

$$= \int_0^{t_1} \lambda du + \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \lambda + \sum_{t_j<u} \alpha e^{-\beta(u-t_j)}du = \lambda t_k + \alpha \sum_{i=1}^{k-1} \int_{t_i}^{t_{i+1}} \sum_{j=1}^{i} e^{-\beta(u-t_j)} =$$

$$= \lambda t_k + \alpha \sum_{i=1}^{k-1} \sum_{j=1}^{i} \int_{t_i}^{t_{i+1}} e^{-\beta(u-t_j)}du = \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k-1} \sum_{j=1}^{i} [e^{-\beta(t_{i+1}-t_j)} - e^{-\beta(t_i-t_j)}] =$$

$$= \lambda t_k - \frac{\alpha}{\beta} \sum_{i=1}^{k} [e^{-\beta(t_k-t_i)} - 1]$$

Replacing in the log-likelihood expression, it is obtained:

$$l = \sum_{i=1}^{k} \log \left[\lambda + \alpha \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)} \right] - \lambda t_k + \frac{\alpha}{\beta} \sum_{i=1}^{k} \left[e^{-\beta(t_k-t_i)} - 1 \right]$$

However, as there are two summations, computationally it is difficult. But, for $i \in \{2, ..., k\}$ the function

$$A(i) = \sum_{j=1}^{i-1} e^{-\beta(t_i-t_j)} = e^{-\beta t_i + \beta t_{i-1}} \left(1 + \sum_{j=1}^{i-1} e^{-\beta t_{i-1} + \beta t_j} \right) = e^{-\beta(t_i-t_{i-1})}(1 + A(i-1))$$

simplifies the expression by assuming the particular condition $A(1) = 0$:

$$l = \sum_{i=1}^{k} log(\lambda + \alpha A(i)) - \lambda t_k + \frac{\alpha}{\beta} \sum_{i=1}^{k} \left[e^{-\beta(t_k-t_i)} - 1 \right]$$

Its derivatives and the Hessian can be seen in [30], and then, depending on the data it will be used different methods for the approximation of the parameters.

Chapter 3

Finance applications

Although initially Hawkes processes were not created for financial modeling, an important application in different financial situations such as a microstructure dynamics, order arrival rate modeling and high-frequency data has been observed. In this section, we will use the knowledge shown in last chapter to see an example of application in the finance field. Particularly, two different ways of modeling the market impact at a microstrutural level are presented by using Hawkes processes.

3.1 Market impact

The market impact, that is the market reaction, describes the changes in prices immediately after an order takes place in a short period of time. In general, price changes because of incoming and cancelled orders. Hence, the behaviour of these changes are of interest since they induced extra costs, specially for large volumes of trades; for example, in the case of a metaorder[1]: when the trading is being executed, the order is becoming less attractive (price raises) and when the execution is over, the market impact is relaxed. This is based on the fact that when an asset is bought, the price will increase; and when it is sold, the price will decrease. This variation depends on the liquidity of the asset: the more liquidity of an asset, the less the price is affected by an order. Furthermore, it is possible that other trades of other assets affect the price of the asset considered; that is known as cross-impact. Hence, the aim is to optimise the rate of trading so as to minimise the market impact and the volatility risk.

[1] **Definition:** A large trading order that is divided in small pieces and executed incrementally is called a *metaorder*.

3.1.1 Bivariate Hawkes process for price modeling

In the first representation, the upward price changes, which is the price change due to the purchase, and the downward price changes, which is the price change due to the sales, are considered [4]. For this, it is taken into account the discrete nature of price variations and that these variations do not depend on the size of the jump. In this situation, the price variations are divided in two different processes, the first one is affected by the purchases, this corresponds to the upward price changes; and the second one, by the sales, that is, the downward price changes. Hence, two counting processes forming a 2-dimensional Hawkes process are taken into consideration, $(N_1(t), N_2(t))$, where $N_1(t)$ is related to the arrival times of upward price changes and the second one, to the arrival times of downward price changes. Thus, the price is given by:

$$P(t) = P(0) + N_1(t) + N_2(t) \qquad t > 0$$

Considering the conditional intensity functions $\lambda_1^*(t)$ and $\lambda_2^*(t)$ and knowing that, in the microstructure level, the price tends to its mean value, that is, it is highly mean reverting, the 2-dimensional Hawkes process has only the crossing terms in the kernel [9]:

$$\lambda_1^*(t) = \lambda + dN_2 * \mu(t) \qquad \lambda_2^*(t) = \lambda + dN_1 * \mu(t)$$

Observe that the more the price increases $P(t)$ (respectively decreases), the greater the conditional intensity function $\lambda_2^*(t)$ (resp. $\lambda_1^*(t)$) is. This corresponds to the mean reversion property. Note that these two counting processes are assymptotically stable if the condition $||\mu|| < 1$ holds.
Suppose that we want to model the impact of a metaorder corresponding to a continuous flow of buying orders with a trading rate per unit of time $r(t)$ which starts at time t_0 and ends at time $t_0 + \tau$ (hence $r(t) = 0 \ \forall t \notin [t_0, t_0 + \tau]$). In order to explain the main dynamics of the market impact, a linear model is assumed, that is, the impact of a metaorder is the sum of the impact of its child orders.
With these assumption, an extra term in the conditional intensity function must be added, which accounts for the impact of the buying order, that is

$$\lambda_1^*(t) = \lambda + dN_2 * \mu(t) + \int_{t_0}^{t} f(r(s)) g_1(s - t_0) ds$$

$$\lambda_2^*(t) = \lambda + dN_1 * \mu(t) + \int_{t_0}^{t} f(r(s)) g_2(s - t_0) ds$$

where f accounts for the instantaneous impact and it satisfies that $f(0) = 0$. Additionally, $g_1(t)$ is the impact kernel and $g_2(t)$ is the cross-impact kernel. Note that

the second term of each expression takes into account the time dependency of the order, while the third one considers the volume.
In order to describe the market impact, a particularization of the kernels is done. As the impact is really focalised on time, the upward impact of a single buying order can be considered as instantaneous, which means that there is an instantaneous jump in the price when there is an order. Hence, as in [4], the impact kernel can be assumed as a Dirac function:

$$g_1(t) = g_{1i}(t) = \delta(t)$$

On the other hand, taking into account that the market reacts to a new trade as if it provoked an upward jump. Then, it is reasonable that the cross-impact kernel is given by:

$$g_2(t) = g_{2i}(t) = C\frac{\mu(t)}{||\mu||}$$

where C is the ratio between the impact decay and the impact amplification. Depending on its value, three different situations are possible:

1. C=0: no opposing reaction, thus, there is a permanent effect on price from metaorders.
2. C=1: the opposing reaction has the same effect as the impact amplification. Thus, one expects that they are compensated asymptotically, that means that the total effect of the metaorder on prices is 0.
3. $C \in (0,1)$, there is an opposing reaction but it is smaller than the impact amplification.

Using the Hawkes impact model, the market impact of a metaorder is given by:

$$\eta(t) = E[P(t)] \quad \forall t \geq t_0$$

Thanks to Proposition 1 of [4], a general result for the impact of a metaorder can be obtained:

Proposition 3.1. *(Transient, decay curves and permanent effect)* For all $t \geq t_0$, the market impact is given by:

$$\eta(t) = \int_{t_0}^{\infty} f(r(s))\left(G(t-s) - (\kappa * G)(t-s)\right) ds$$

where $G(t) = \int_0^t (g_1(u) - g_2(u))du$ and $\kappa = \sum_{n=1}^{\infty}(-1)^{n-1}\mu^{*n}$. In the particular case of a Hawkes impact model, with the corresponding kernels given before, this simplifies to:

$$\eta(t) = \int_{t_0}^{t} f(r(s))\left(1 - \left(1 + \frac{C}{||\mu||}\right)\int_0^{t-s}\kappa(u)du\right) ds \quad \forall t \geq t_0$$

And in the particular situation of having a constant rate strategy, that is, $r(t) = r \ \forall t \in [t_0, t_0 + \tau]$, then, the permanent effect of the metaorder on the price is:

$$\eta(\infty) = \lim_{t \to +\infty} \eta(t) = f(r)\tau \frac{1 - C}{1 + ||\mu||}$$

Proof. Direct consequence of Theorem 2.26 and by considering that

$$\sum_{i=0}^{\infty} (-1)^n x^n = \frac{-x}{1+x}$$

For more details see Appendix A of [4]. □

If now we consider that the trade rate is constant and that the excitation function has the form of a power law, then the market impact curve will decay asymptotically as a power-law, so that [4]:

Corollary 3.2. Suppose a constant rate strategy in a Hawkes impact model. Assuming that $\mu \geq \mu^{*2}$ and that $\exists K > 0$ such that $\lim_{t \to \infty} \mu(t) t^{-b} = K$ for $b \in (-2, -1)$, then the market impact curve decays to $\eta(\infty)$ asymptotically as a power-law with exponent $b + 1$, that is:

$$\inf \left(p : \int_1^{\infty} (\eta(t) - \eta(\infty)) t^{-p-1} dt < \infty \right) = b + 1$$

Proof. See Appendix A of [4]. □

Thus, in the case of a unique market maker[2] with noise traders, and assuming an idealized market, the impulsive Hawkes impact model is seen as the modelling of the market maker inventory, that is, $\lambda_1^* - \lambda_2^*$. The market maker adjusts the interaction level for which they accept the metaorder by backtestings and their experience for a given risk budget. With this, they supply attractive prices with a small risk until the inventory exceeds a threshold. A market maker with a low C does not yield a lot of resistance to the metaorder pressure and does not supply a large herding effect when they close out their inventory. Instead, a market maker with a large C generates a high opposite pressure to metaorder and, when the threshold is exceeded, a large herding effect is produced (which is compensated by slower market maker with large inventory).
Note that this model could be extended by considering a continuum market maker, where C follows a distribution taking into account the distribution of metaorders

[2] **Definition:** a *market maker* is a company/individual that provides quotes for selling and buying trades which accepts to provide liquidity under their own risks.

in a given market.
Suppose a duration $T(C)$ of a metaorder. Initially, when the order starts, the metaorder interacts with market makers with low C. A time after, as the buyers consider they are adversely selected, they stop ordering and they provide their inventory to market makers with the highest C. When the metaorder has finished, the interaction continues with market makers with the best C, and then, they have to unwind their inventory slowly to see their benefits.

3.1.2 Tetravariate Hawkes process for price modeling

As described in [7], another possibility to model the market impact is by taking into account a four-dimensional point process, which considers the order arrival self-excitation, the price changes mean reversion, the impact of an order arrival on the price and the feedback of price variations on the trading activity. Here, all market trades and all mid-price change events are consider. However, the trade volumes are not taken into account so as to make easy the notation (but they could be considered by adding a mark variable).
For this modeling, jumps of the market caused by orders are described by a bivariate Hawkes process which counts the orders until time t:

$$T(t) = \begin{pmatrix} T^-(t) \\ T^+(t) \end{pmatrix}$$

where $T^-(t)$ gives the number of orders at the best ask[3] and $T^+(t)$, the number of orders at the best bid[4]. Thus, when a new market order arrives, $T^+(t)$ or $T^-(t)$ is increased by one. The trade process T is completely characterised by the conditional intensity vector $\lambda_T^*(t)$.
Similarly, the mid-price[5] is described by a two-dimensional Hawkes process, particularly $X(t) = N_1(t) - N_2(t)$, which $N^+(t)$ defines the number of upward price jumps at time t and $N^-(t)$, the number of downward price jumps at time t.

$$N(t) = \begin{pmatrix} N^-(t) \\ N^+(t) \end{pmatrix}$$

When the price varies, $N^-(t)$ or $N^+(t)$ increase by 1. The mid-price is characterised by the conditional intensity vector, $\lambda_N^*(t)$.

[3] **Definition:** The *best ask* is the lowest quoted offer price from competing market makers for a particular trading instrument (see `https://www.investopedia.com/terms/b/bestask.asp` for more details).

[4] **Definition:** The *best bid* is the highest quoted for a particular security among all bids offered by competing market makers; that is, the highest price that an investor is willing to pay for an asset (see `https://www.investopedia.com/terms/b/bestbid.asp` for more details).

[5] **Definition:** The *mid-price* corresponds to the average of the best ask and best bid prices.

Hence, the process that is being considered is:

$$P(t) = \begin{pmatrix} T(t) \\ N(t) \end{pmatrix} = \begin{pmatrix} T^-(t) \\ T^+(t) \\ N^-(t) \\ N^+(t) \end{pmatrix}$$

which allows to define the influence of the different components and its conditional intensity vector is given by:

$$\lambda^*(t) = \begin{pmatrix} \lambda^*_T(t) \\ \lambda^*_N(t) \end{pmatrix} = \begin{pmatrix} \lambda^*_{T^-}(t) \\ \lambda^*_{T^+}(t) \\ \lambda^*_{N^-}(t) \\ \lambda^*_{N^+}(t) \end{pmatrix}$$

The interest in using Hawkes processes relies on the issue that it takes into account the influence of any component of $P(t)$ because of the structure of the conditional intensity function, which is $\lambda^*(t) = \lambda + dP(t) * \mu$. In this particular case, the background intensity, which takes into account the exogeneous intensity of trades, is:

$$\lambda = \begin{pmatrix} \lambda_0 \\ \lambda_0 \\ 0 \\ 0 \end{pmatrix}$$

because the best bid and best ask are assumed to be equal and that the mid-price jumps are only caused by endogeneous variables[6]. The kernel which gives the excitation function is then:

$$\mu(t) = \begin{pmatrix} \mu^T(t) & \mu^F(t) \\ \mu^I(t) & \mu^N(t) \end{pmatrix}$$

where

μ^T (influence of T on λ^*_T) takes into account the trade correlations.

μ^I (influence of T on λ^*_N) takes into account the impact of a trade on the price.

μ^F (influence of N on λ^*_T) is the response influence of the price moves on the trades.

μ^N (influence of N on λ^*_N) is the influence of past and future changes in price.

[6]**Definition:** An *endogeneous variable* is a variable that depends on other variables of the system

Symmetries between the best bid and the best ask for trades and symmetries between the up and down directions for price lead to an excitation function of the form:

$$\mu^i(t) = \begin{pmatrix} \mu_i^s(t) & \mu_i^c(t) \\ \mu_i^c(t) & \mu_i^s(t) \end{pmatrix}$$

for $i = I, F, T, N$ and where the superindex s indicates the self influence and c, the cross influence. As these matrices can be diagonalised using the same basis, they commute because their eigenvalues are the sum and the difference of the self and cross terms.

Note that in the situation of having a buying order that takes all the volume at the best ask, this results in an instantaneous jump for the mid-price. Although instantaneous jumps are not possible, this can be modelled by considered that the changes in the mid-price are a small time interval after the order because of the causality relation. With this, $\mu^I(t)$ will result in a function close to the Dirac distribution arount $t = 0$. This distribution is obtained by considering a positive L^1-function with norm I and with domain Δt of a few milliseconds. Then, the price increases after Δt milliseconds of the order and thus, the price change follows a Poisson distribution with parameter I. With this, $\mu^I(t) = \delta(t)\mathbb{I}$.

In this situation, the stability condition is given by:

$$(||\mu_F^s|| + ||\mu_F^c||)(||\mu_I^s|| + ||\mu_I^c||) < (1 - ||\mu_T^s|| - ||\mu_T^c||)(1 - ||\mu_N^s|| - ||\mu_N^c||)$$

and

$$||\mu_T^s|| + ||\mu_T^c|| \leq 1 \quad ||\mu_N^s|| + ||\mu_N^c|| < 1$$

The proof can be seen in Annex 9.1 of [7]. In the particular case that $\mu^F = 0$, that is, the price changes do not influence the orders; then the stability condition is reduced to the second and third conditions.

For this model, which is to model orders which can be influenced by different agents, the impact of an order can only be obtained by the response function [7]:

Definition 3.3. The *response function* is the price variation between 0 and t, knowing that on time 0 there has been a trade:

$$R(t) = \begin{cases} E[N^+(t) - N^-(t) | dT^+(0) = 1] & \text{if} \quad t > 0 \\ 0 & \text{otherwise} \end{cases}$$

Notice that the response function considers the impact of all market orders that are in the same metaorder as the market order that is being consider. Otherwise, the market impact given in Proposition 3.2 takes into account the market orders of a single agent. An analytical expression and how it is obtained can be found in [7].

Conclusions

Now, it is time to conclude this dissertation by providing an overview of it and going through as a whole.
First of all, Hawkes processes permit to model situations where self-excitation and contagion are two properties quite important. Moreover, the fact that they are a type of point processes, particularly an extension of homogeneous Poisson processes, allows to take into account some of their characterizations, such as the conditional intensity function and the compensator.
Secondly, as there are different representations of these processes and they are quite easy to compute, their used has been extended during last years in many fields: not only on the ones that have been encompassed here, that is earthquakes and finances, but also criminality and neuronal signal transmissions.
Finally, we have seen that the application in a specific theme, in this case the market impact, can be modelled by using different types of Hawkes processes, which justifies that these processes are adaptable; not only because of their extension to different dimensions, but also because of the flexible structure that the conditional intensity function can present.
In conclusion, although some improvements and applications could have been done, this project allows to have a fulfilling overview not only of Hawkes processes, but also on point processes; together with an introduction in the finance field.

Bibliography

[1] Aït-Shalia, Y.; Cacho-Diaz, J.; Laeve, R.J.A. *Modeling financial contagion using mutually exciting jump processes*, Journal of Financial Economics. 2014.

[2] Asmussen, S. *Applied Probability and Queues*, Springer. 2003.

[3] Athreya, K.B.; Ney, P.E. *Branching processes*, Springer. 1972.

[4] Bacry, E.; Iuga, A.; Lasnier, M.; Lehalle, C.A. *Market impacts and the life cycle of investors orders*. 2014.

[5] Bacry, E.; Mastromatteo I.; Muzy J.F. *Hawkes processes in finance*. 2015.

[6] Bacry, E.; Delattre, S.; Hoffmann, M.; Muzy, J.F. *Some limit theorems for Hawkes processes and application to financial statistics*. 2013.

[7] Bacry, E.; Muzy, J.F. *Hawkes model for price and trades high-frequency dynamics*. 2013.

[8] Bacry, E.; Muzy, J.F. *Second order statistics characterization of Hawkes processes and non-parametric estimation*. 2014.

[9] Bacry, E.; Delattre, S.; Hoffmann, M.; Muzy, J.F. *Modeling microstructure noise with mutually exciting point processes*. 2011.

[10] Bauwens, L.; Hautsch, N. *Modelling Financial High Frequency Data Using Point Processes*. 2007.

[11] Bielecki, T.R.; Rutkowski, M. *Credit Risk: Modeling, Valuation and Hedging*, Springer. 2001

[12] Bowsher, C.G. *Modelling Security Market Events in Continuous Time: Intensity Based, Multivariate Point Process Models*. 2003.

[13] Brémaud, P.; Massoulié, L. *Stability of non linear Hawkes process*, Annals of Probability. 1996.

[14] Brémaud, P.; Masoulié, L. *Hawkes branching point processes without ancestors*. Journal Applied Probability. 2001.

[15] Chavez-Demoulin, V.; McGill, J.A. *High-frequency financial data modeling using Hawkes processes*, Journal of Banking & Finance. 2012.

[16] Connecticut Universitat *Chapter 3: Stochastic calculus* [`http://bass.math.uconn.edu/math5016f14/post3.pdf`]

[17] Cox, D.R.; Isham, V. *Point Processes*, Chapman and Hall. 1980.

[18] Daley, D.J.; Vere-Jones, D. *An Introduction to the theory of Point Processes*, Springer-Verlag. 1988.

[19] Engelke, S.; Hautphenne, S.; Hurth T. Notes of *Applied Stochastic Processes*. EPFL. 2018.

[20] Errais, E.; Giesecke, K.; Goldberg, L.R. *Affine Point Processes and Portfolio Credit Risk*, Journal Financial Mathematics. 2010.

[21] Fathi-Vajargah, B.; Khoshkar-Foshtomi, H. *Simulating of point process using conditional intensity function (Hazard function)*, International Journal of Advanced Statistics and Probability. 2014.

[22] Gao, X.; Zhu, L. *Functional central limit theorems for stationary Hawkes processes and application to infinite-server queues*. 2018.

[23] Gulddahl, J.; *Lecture Notes: Temporal Point Processes and the Conditional Intensity Function*. 2018.

[24] Hall, P.; Heyde, C.C. *Martingale limit theory and its application*, Probability and Mathematical Statistics. 1980.

[25] Hawkes, A. *Point Spectra of Some Mutually Exciting Point Processes*. 1971.

[26] Hawkes, A.; Oakes, D. *A cluster process representation of a self-exciting process*. 1974.

[27] Karatzas, I.; Shreve, S.E. *Brownian Motion and Stochastic Calculus*, Springer. 1998.

[28] Laub, P.J.; Taimre, T.; Pollett, P.K. *Hawkes Processes*. 2015.

[29] Loether, G. *Compensators of Counting Processes*. 2011. [`https://almostsure.wordpress.com/2011/12/27/compensators-of-counting-processes/`]

[30] Ozaki, T. *Maximum likelihood estimation of Hawkes self-exciting point processes*. 1979.

[31] Pinto, J.A.; Viana, P. *Bayesian Variable Selection Methods for Log-Gaussian Cox Processes*, Springer. 2018.

[32] Reiss, R.D. *A Course on Point Processes*. Springer-Verlag. 1993.

[33] Roueff, F.; von Sachs, R.; Sansonnet, L. *Locally stationary Hawkes processes*, Science Direct. 2015.

[34] Spiegel, M.R.; Lipschutz, S.; Liu, J. *Schaum: Fórmulas y tablas de matemática aplicada*, Mc Graw Hill. 2005.

[35] van Neerven, J. *The Doob-Meyer decomposition theorem*. 2004.