Simulation, Estimation and Applications of Hawkes Processes

A PROJECT
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Katarzyna Obral

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF SCIENCE

Barry James, Kang James

June, 2016

Acknowledgements

There are a lot of people that I am grateful for their contribution to my time in graduate school. First of all, I would like to express my thankfulness to prof. Dalibor Froncek for encouraging me to apply for the Applied and Computational Mathematics program at the University of Minnesota Duluth and for all the help I received from him. I would also like to thank Michal Hrabia, for answering all my questions regarding the program and life in the USA before my arrival. I would also like to express my gratitude to Elsa Erickson for her friendship and support throughout these two years I have spent in Duluth. Last, but not least, I would like to thank my advisers, prof. Barry James, who proposed and supported this project and prof. Kang James, whose questions and advice during the Statistics Seminar helped me stay on track with my work. I am grateful for all their help and time I needed while working on this project.

Dedication

To Konrad.

Abstract

Hawkes processes are a particularly interesting class of stochastic processes that were introduced in the early seventies by A. G. Hawkes, notably to model the occurrence of seismic events. Since then they have been applied in diverse areas, from earthquake modeling to financial analysis. The processes themselves are characterized by a stochastic intensity vector, which represents the conditional probability density of the occurrence of an event in the immediate future. They are point processes whose defining characteristic is that they *self-excite*, meaning that each arrival increases the rate of future arrivals for some period of time.

In this project, we present background and all major aspects of Hawkes processes. Before introducing the univariate Hawkes process, we recall the theory regarding counting processes and Poisson processes. Then we provide two equivalent definitions of Hawkes process and the theory regarding stochastic time change, which is necessary in the analysis of Hawkes processes. We show two simulation algorithms, using two different approaches, which let us generate Hawkes processes. Then we define what a multivariate Hawkes process is. We also present a real data example, which is the analysis of the discharge of one antennal lobe neuron of a cockroach during a spontaneous activity. We provide a plot of this data and we estimate the intensity process, using the methods developed by Christophe Pouzat. Finally, we discuss possible applications of Hawkes processes as well as possibilities for future research in the area of self-exciting processes.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Events that are observed in time frequently cluster naturally. An earthquake typically increases the geological tension in the region where it occurs, and aftershocks likely follow [25]. A fight between rival gangs can ignite a spate of criminal retaliations [22]. Selling a significant quantity of a stock could precipitate a trading flurry or, on a larger scale, the collapse of a Wall Street investment bank could send shock-waves through the world's financial centers, as we read in [18].

Hawkes processes were introduced for the first time by Alan Hawkes in 1971. They are a mathematical model for these self–exciting processes, in which the occurrence of an event increases the probability of occurrence of another event. We will specifically study the case in which this increase is instantaneous and then exponentially decaying over time.

Chapter 2 introduces definitions of counting processes and their properties, Poisson process, nonhomogeneous Poisson process and intensity process. We give a general definition of Hawkes process and then we specify the exact model we are going to use in simulation algorithms. We introduce the definition of integrated intensity and we state a stochastic time change theorem. We also provide proofs for some of the results.

In Chapter 3 we show two simulation algorithms, using two different approaches, the cluster-based approach to Hawkes processes and the intensity-based approach. We provide proofs and examples for both of the algorithms.

In Chapter 4 we define the multivariate Hawkes process and we show an example of bivariate Hawkes process.

In Chapter 5 we provide some insight into methods developed by Christophe Pouzat for estimation of intensity, based on a real data example. We provide detailed explanations and a few graphs to help better understanding of the method. Then we comment on what conclusions can be drawn from our result.

Finally, in Chapter 6 we discuss possible applications of Hawkes processes. We talk about how seismic events, crime data, insurance company surplus, trades-through in a limit order book as well as fluctuations in social systems can be modeled. We state what work has been so far and we discuss the possibilities for future work in the area of self-exciting processes.

Chapter 2

Basic Definitions and Results

Let us now introduce some definitions and results we are going to need and use throughout our discussion.

Since Hawkes process is a special type of counting process, we will begin by defining what a counting process is. However; to be able to fully understand what a counting process is, a point process must be defined. We will discuss the properties and types of counting processes, which will lead into the definition of Hawkes process.

2.1 Counting Process

Definition 2.1.1 (Point Process). Let $\{T_i, i \in \mathbb{N}\}$ be a sequence of non–negative random variables such that $\forall i \in \mathbb{N}$, $T_i < T_{i+1}$. We call $\{T_i, i \in \mathbb{N}\}$ a simple *point process* on $\mathbb{R}_+$. We assume that $T_0 = 0$. Now we can move on to the definition of a counting process.

Definition 2.1.2 (Counting Process). A stochastic process $\{N(t), t \geq 0\}$ is said to be a *counting process* if $N(t)$ represents the total number of events that occur by time t.

A counting process can be also defined a little bit more formally, with the help of the notion of point processes, which we have introduced in Definition 2.1.1.

Definition 2.1.3 (Counting Process). Let $\{T_i, i \in \mathbb{N}\}$ be a point process. The right-continuous process

$$N(t) = \sum_{i \in \mathbb{N}} I_{\{T_i \leq t\}}$$

is called the *counting process* associated with $\{T_i, i \in \mathbb{N}\}$.

Let us have a look at a couple of real–life counting processes examples. If we let $N(t)$ equal the number of persons who enter a particular store at or prior to time t, then $\{N(t), t \geq 0\}$ is a counting process in which an event corresponds to a person entering the store. If we say that an event occurs whenever a child is born, then $\{N(t), t \geq 0\}$ is a counting process when $N(t)$ equals the total number of people who were born by time t. If $N(t)$ equals the number of goals that a given hockey player scores by time t, then $\{N(t), t \geq 0\}$ is a counting process. An event of this process will occur whenever the hockey player scores a goal.

Corollary 2.1.1. From Definition 2.1.2 a counting process $N(t)$ must satisfy

1. $N(t) \geq 0$,
2. $N(t)$ is integer valued,
3. If $s < t$, then $N(s) \leq N(t)$,
4. For $s < t$, $N(t) - N(s)$ equals the number of events that occur in the interval $(s, t]$.

If we consider the interval $(0, \mathcal{T}]$, let

$$0 < T_1 < T_2 < \ldots < T_{N(t)} \leq t \leq \mathcal{T}$$

be a set of event (spike) times from a point process. For $t \in (0, \mathcal{T}]$, $N(t)$ is the sample path of the associated counting process. The sample path is a right-continuous function, that jumps 1 at the event times and is constant otherwise, as defined in [34]. In this way, $N(t)$ counts the number and location of spikes in the interval $(0, t]$. Therefore, it contains all the information in the sequence of events and spike times.
Now let us look at three important properties of counting processes.

Definition 2.1.4 (Independence). A counting process is said to possess *independent increments* if the numbers of events that occur in disjoint time intervals are independent.

Definition 2.1.5 (Stationarity). A counting process is said to possess *stationary increments* if the distribution of the number of events that occur in any interval of time depends only on the length of the time interval.

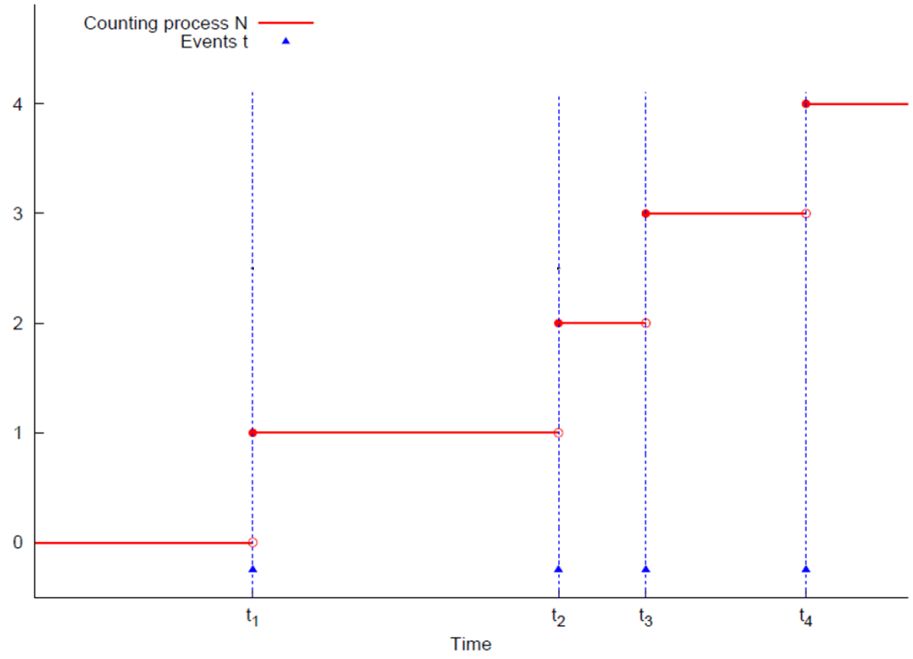


Figure 2.1: Illustration of a simple point process and a corresponding counting process.

In other words, the process has stationary increments if the number of events in the interval $(s, s+t]$ has the same distribution for all s.

Definition 2.1.6 (Homogeneity). A stochastic process is said to be *homogeneous* in space if the transition probability between any two state values at two given times depends only on the difference between those state values.
The process is *homogeneous* in time if the transition probability between two given state values at any two times depends only on the difference between those times.

2.2 Poisson Process

We are going to have a closer look at one of the most important type of counting processes. Beforehand we recall the concept of a function $f(\cdot)$ being $o(g)$ and $O(g)$.

2.2.1 Landau Symbols

We will introduce *Big O* and *Little o* notation, sometimes also called *Landau symbols* after German mathematician Edmund Landau.

Definition 2.2.1. Let $f : \mathbb{R} \longrightarrow \mathbb{R}$ and $g : \mathbb{R} \longrightarrow \mathbb{R}$, be two functions and let g be positive in a punctured neighborhood of $a \in \mathbb{R}$, that is for $x \in (a-\xi, a+\xi)\backslash\{a\}$ for some $\xi > 0$.

We say f is 'Big O' of g in a neighbourhood of a and we write $f(x) = O(g(x))$ as $x \to a$ if

$$\limsup_{x\to a} \frac{\mid f(x) \mid}{g(x)} < \infty.$$

We say f is 'Little o' of g in a neighbourhood of a and we write $f(x) = o(g(x))$ as $x \to a$ if

$$\lim_{x\to a} \frac{f(x)}{g(x)} = 0.$$

Note that $O(f(x))$ represents the set of all functions with a smaller or the same order of growth as $f(x)$ when approaching a, whereas $o(f(x))$ is the set of all functions with a strictly smaller rate of growth than $f(x)$.

Since a symbol h is commonly used, to represent a number approaching 0. In this paper we will say that the function $f(\cdot)$ is $o(h)$ if

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

Note, that for the function $f(\cdot)$ to be $o(h)$ it is necessary that $f(h)$ goes to zero faster than h does.

2.2.2 Poisson Process

Definition 2.2.2 (Poisson Process). The counting process $\{N(t), t \geq 0\}$ is said to be a *Poisson process* with rate $\lambda > 0$ if the following axioms hold:

1. $N(0) = 0$,
2. $\{N(t), t \geq 0\}$ has independent increments,
3. $\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h)$,
4. $\mathbb{P}(N(t+h) - N(t) = 1) = \lambda h + o(h)$.

Theorem 2.2.1 (Ross). If $\{N(t), t \geq 0\}$ is a Poisson process with rate $\lambda > 0$, then for all $s > 0$, $t > 0$, $N(s+t) - N(s)$ is a Poisson random variable with mean λt.

That is, the number of events in any interval of length t is a Poisson random variable with mean λt. To be able to prove Theorem 2.2.1 we need to define the Moment Generating Function.

Definition 2.2.3 (Moment Generating Function). If X is a random variable, then the *moment generating function* $\phi(u)$ of X is defined for all values of u by

$$\phi(u) = \mathbb{E}[e^{ux}].$$

We call $\phi(u)$ the moment generating function because all of the moments of X can be obtained by successively differentiating $\phi(u)$. Note that, replacing u by $(-t)$ gives the Laplace transform of X.
An important property of moment generating functions is that it uniquely determines the distribution. That is, there exists a one–to–one correspondence between the moment generating function and the distribution function of a random variable.

Proposition 2.2.1. If X is a Poisson random variable with mean λ, then its moment generating function $\phi(u)$ has the following form

$$\phi(u) = \mathbb{E}[e^{uX}] = e^{\lambda(e^u - 1)} \tag{2.2.1}$$

Proof.

$$\begin{aligned}
\mathbb{E}[e^{uX}] &= \sum_i e^{ui} e^{-\lambda} \frac{(\lambda)^i}{i!} \\
&= e^{-\lambda} \sum_i \frac{(\lambda e^u)^i}{i!} \\
&= e^{-\lambda} \cdot e^{\lambda e^u} = e^{\lambda(e^u - 1)}.
\end{aligned}$$

Now, we are ready to prove Theorem 2.2.1. □

Proof of Theorem 2.2.1. Similarly to the steps shown in [32], we begin by deriving

the moment generating function of $N(t)$. Let us fix $u > 0$ and define

$$g(t) = \mathbb{E}[e^{uN(t)}]$$

We obtain $g(t)$ by deriving a differential equation in the following way.

$$\begin{aligned} g(t+h) &= \mathbb{E}[e^{uN(t+h)}] \\ &= \mathbb{E}[e^{u(N(t)+N(t+h)-N(t))}] \\ &= \mathbb{E}[e^{uN(t)}e^{u(N(t+h)-N(t))}] \end{aligned} \tag{2.2.2}$$

Since the increments of the Poisson process are independent, from equation (2.2.2) we obtain that

$$\begin{aligned} g(t+h) &= \mathbb{E}[e^{uN(t)}]\mathbb{E}[e^{u(N(t+h)-N(t))}] \\ &= g(t)\mathbb{E}[e^{u(N(t+h)-N(t))}] \end{aligned} \tag{2.2.3}$$

From axioms 3 and 4 from Definition 2.2.2 we get that

$$\begin{aligned} &\mathbb{P}(N(t+h) - N(t) = 0) = 1 - \lambda h + o(h), \\ &\mathbb{P}(N(t+h) - N(t) = 1) = \lambda h + o(h), \\ &\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h). \end{aligned}$$

We obtain the expected value $\mathbb{E}[e^{u(N(t+h)-N(t))}]$ by conditioning on which of the three possibilities presented above occurs.

$$\begin{aligned} \mathbb{E}[e^{u(N(t+h)-N(t))}] &= e^{u\cdot 0} \cdot \big(1 - \lambda h + o(h)\big) + e^{u\cdot 1} \cdot \big(\lambda h + o(h)\big) + o(h) \\ &= 1 \cdot (1 - \lambda h + o(h)) + e^{u} \cdot (\lambda h + o(h)) + o(h) \\ &= 1 - \lambda h + e^{u}\lambda h + o(h) \\ &= 1 + \lambda h(e^{u} - 1) + o(h) \end{aligned} \tag{2.2.4}$$

From equations (2.2.3) and (2.2.4) we obtain

$$g(t+h) = g(t)(1 + \lambda h(e^{u} - 1) + o(h)).$$

It can be rewritten as

$$\begin{aligned} g(t+h) - g(t) &= g(t)\lambda h(e^u - 1) + o(h) \\ \frac{g(t+h) - g(t)}{h} &= g(t)\lambda(e^u - 1) + \frac{o(h)}{h} \end{aligned} \tag{2.2.5}$$

As h goes to 0, we get

$$g'(t) = g(t)\lambda(e^u - 1),$$

which can be simplified to

$$\frac{g'(t)}{g(t)} = \lambda(e^u - 1).$$

After integrating on both sides from 0 to t we obtain

$$\log(g(t)) = \lambda(e^u - 1)t + C.$$

To get the value of constant C we need to evaluate $g(0)$.

$$g(0) = \mathbb{E}[e^{u(N(0)}] = \mathbb{E}[e^0] = 1 \tag{2.2.6}$$

From which follows that $C = 0$ and

$$g(t) = \mathbb{E}[e^{uN(t)}] = e^{\lambda t(e^u - 1)}. \tag{2.2.7}$$

By comparing equation (2.2.1) and (2.2.7), we see that the obtained value of $g(t)$ is actually the moment generating function $\phi(u)$ of a Poisson random variable with mean λt. Because the moment generating function uniquely determines the distribution, we conclude that $N(t)$ is Poisson with mean λt.

Now, let us fix s and let $N_s(t) = N(s+t) - N(s)$ be equal to the number of events happening in the first t time units after we start our count at time s. The counting process $\{N_s(t), t \geq 0\}$ satisfies then all the axioms for being Poisson process with rate λ. In consequence, we conclude that $N_s(t) = N(s+t) - N(s)$ is Poisson with mean λt. $\square$

Now, let us move on to more complicated counting processes.

2.3 Nonhomogeneous Poisson Process

Definition 2.3.1 (Nonhomogeneous Poisson Process). The counting process $\{N(t), t \geq 0\}$ is said to be a *nonhomogeneous* Poisson process with intensity function $\lambda(t)$, $t \geq 0$, if

1. $N(0) = 0$,
2. $\{N(t), t \geq 0\}$ has independent increments,
3. $\mathbb{P}(N(t+h) - N(t) \geq 2) = o(h)$,
4. $\mathbb{P}(N(t+h) - N(t) = 1) = \lambda(t)h + o(h)$.

The intensity function $\lambda(t)$ of a nonhomogeneous Poisson process is a deterministic function.

Definition 2.3.2 (Mean Value Function). The function $m(t)$ defined by

$$m(t) = \int_0^t \lambda(y)\mathrm{dy}$$

is called the *mean value function* of the nonhomogeneous Poisson process.

Theorem 2.3.1 (Ross). If $\{N(t), t \geq 0\}$ is a nonstationary Poisson process with intensity function $\lambda(t)$, $t \geq 0$, then $N(t+s) - N(s)$ is a Poisson random variable with mean $m(t+s) - m(s) = \int_s^{s+t} \lambda(y)\mathrm{dy}$.

Proof. The proof of Theorem 2.3.1 is analogous to the proof of Theorem 2.2.1. We begin by deriving the moment generating function of $N(t)$. Let $g(t) = \mathbb{E}[e^{uN(t)}]$. Then

$$\begin{aligned} g(t+h) &= g(t)\mathbb{E}[e^{u(N(t+h)-N(t))}] \\ &= g(t)\mathbb{E}[e^{uN_t(h)}] \end{aligned} \tag{2.3.1}$$

where $N_t(h) = N(t+h) - N(t)$. From axioms 3 and 4 from Definition 2.2.2 we get that

$$\begin{aligned} &\mathbb{P}(N_t(h) = 0) = \mathbb{P}(N(t+h) - N(t) = 0) = 1 - \lambda(t)h + o(h), \\ &\mathbb{P}(N_t(h) = 1) = \mathbb{P}(N(t+h) - N(t) = 1) = \lambda(t)h + o(h), \\ &\mathbb{P}(N_t(h) \geq 2) = \mathbb{P}(N(t+h) - N(t) \geq 2) = o(h). \end{aligned}$$

We obtain the expected value $\mathbb{E}[e^{uN_t(h)}]$ by conditioning on whether $N_t(h)$ is 0, 1 or ≥ 2.

$$\begin{aligned} g(t+h) &= g(t)\big(e^{u\cdot 0}\cdot(1-\lambda(t)h+o(h)) + e^{u\cdot 1}\cdot(\lambda(t)h+o(h)) + o(h)\big) \\ &= g(t)\big(1\cdot(1-\lambda(t)h+o(h)) + e^{u}\cdot(\lambda(t)h+o(h)) + o(h)\big) \\ &= g(t)(1-\lambda(t)h+e^{u}\lambda(t)h+o(h)) \\ &= g(t)-g(t)\lambda(t)h+g(t)e^{u}\lambda(t)h+o(h) \end{aligned} \tag{2.3.2}$$

Therefore:

$$g(t+h)-g(t) = g(t)\lambda(t)h(e^{u}-1)+o(h). \tag{2.3.3}$$

Dividing by h yields

$$\frac{g(t+h)-g(t)}{h} = g(t)\lambda(t)(e^{u}-1)+\frac{o(h)}{h}. \tag{2.3.4}$$

As $h \to 0$, we get

$$g'(t) = g(t)\lambda(t)(e^{u}-1),$$

which can be simplified to

$$\frac{g'(t)}{g(t)} = \lambda(t)(e^{u}-1).$$

Integrating on both sides from 0 to t results in

$$\log(g(t)) - \log(g(0)) = (e^{u}-1)\int_0^t \lambda(t)\mathrm{dt}$$

Since $g(0) = 1$ (analogous as in equation 2.2.6) and $\int_0^t \lambda(t)\mathrm{dt} = m(t)$ from Definition 2.3.2, therefore

$$g(t) = \mathbb{E}[e^{uN(t)}] = e^{m(t)(e^{u}-1)}. \tag{2.3.5}$$

After comparing equation (2.3.5) with Proposition 2.2.1, it follows that obtained value of $g(t)$ is actually the moment generating function of a Poisson random variable with mean $m(t)$. We conclude that $N(t)$ is Poisson with mean $m(t)$.

Note that the counting process $\{N_s(t), t \geq 0\}$, where $N_s(t) = N(s+t) - N(s)$

is a nonhomogeneous Poisson process with intensity function $\lambda_s(t) = \lambda(s+t)$, $t > 0$. In consequence, $N_s(t) = N(s+t) - N(s)$ is Poisson random variable with mean

$$\int_0^t \lambda_s(y)\mathrm{dy} = \int_0^t \lambda(s+y)\mathrm{dy} = \int_s^{s+t} \lambda(x)\mathrm{dx}$$

which ends the proof. □

2.4 Hawkes Process

We are interested in counting processes $\{N(t), t \geq 0\}$ whose intensity function value at time t is not fixed, but depends on some random inputs, including the history of the process. Let then $\mathcal{H}_t$ denote the *history* of the process up to time t, which is supposed to summarize everything that goes into setting the intensity. Then $\lambda(t|\mathcal{H}_t)$, the intensity rate at time t, is a random variable, whose value is determined by $\mathcal{H}_t$ and which is such that

$$\mathbb{P}(N(t+h) - N(t) = 1|\mathcal{H}_t) = \lambda(t|\mathcal{H}_t)h + o(h),$$

and

$$\mathbb{P}(N(t+h) - N(t) \geq 2|\mathcal{H}_t) = o(h).$$

Note that $\lambda(t|\mathcal{H}_t) = \lambda(t|N(s), s \leq t)$.

A counting process having a random intensity function is called a Hawkes process. It assumes that there is a base intensity function $\lambda_0(t) > 0$, and that associated with each event is a nonnegative random variable, called *a mark*, whose value is independent of all that has previously occurred and has distribution given by the cumulative distribution function G. Whenever an event occurs, it is supposed that the current value of the random intensity function increases by the amount of the event's mark, with this increase decreasing over time at an exponential rate δ.

If there have been a total of $N(t)$ events by time t, with

$$0 < T_1 < T_2 < \ldots < T_{N(t)} \leq t$$

being the event times and Y_i being the mark of event i, for $i = 1, \ldots, N(t)$, then

$$\lambda(t|\mathcal{H}_t) = \lambda_0(t) + \sum_{i=1}^{N(t)} Y_i e^{-\delta(t-T_i)}.$$

Intuitively a Hawkes process is a counting process in which

1. $\lambda(0|\mathcal{H}_0) = \lambda_0(0)$;
2. whenever an event occurs, the random intensity increases by the value of the event's mark;
3. if there are no events between time s and time $s + t$ then $\lambda(s + t|\mathcal{H}_{s+t}) = \lambda_0(s + t) + (\lambda(s|\mathcal{H}_s) - \lambda_0(s))e^{-\delta t}$.

Because the intensity increases each time an event occurs, the Hawkes process is said to be a *self-exciting* process. Now let us present formal definition of the intensity process.

Definition 2.4.1 (Intensity Process). Let N be a point process and let $\{\mathcal{H}_t, t \geq 0\}$ be a history of the process N. The right-continuous *intensity process* is defined as

$$\lambda(t|\mathcal{H}_t) = \lim_{h \downarrow 0} \mathbb{E}[\frac{N(t+h) - N(t)}{h}|\mathcal{H}_t],$$

or equivalently

$$\lambda(t|\mathcal{H}_t) = \lim_{h \downarrow 0} \frac{1}{h}\mathbb{P}[N(t+h) - N(t) > 0|\mathcal{H}_t].$$

Now, let us have a look at the general definition of a self–exciting process.

Definition 2.4.2 (Self-exciting Process). $N = \{N(t), t \geq 0\}$ is a *self-exciting process* if

$$\begin{aligned}\lambda(t|\mathcal{H}_t) &= \lambda_0(t) + \int_0^t \nu(t-s)\mathrm{dN(s)}, \\ &= \lambda_0(t) + \sum_{T_i \leq t} \nu(t - T_i),\end{aligned}$$

where $\lambda_0 : \mathbb{R} \to \mathbb{R}_+$ is a deterministic base intensity and function, $\nu : \mathbb{R}_+ \to \mathbb{R}_+$ is called a *kernel* and it expresses the positive influence of the past events T_i on the current value of the intensity process.

A Hawkes process is a self-exciting process with exponential kernel.

Definition 2.4.3 (General Hawkes Process). A Hawkes process is a self–exciting process of order P with exponential kernel $\nu(t) = \sum_{j=1}^{P} Y_j e^{-\delta_j t}$, so that the intensity of the model becomes:

$$\begin{aligned}\lambda(t|\mathcal{H}_t) &= \lambda_0(t) + \int_0^t \sum_{j=1}^{P} Y_j e^{-\delta_j (t-s)} \mathrm{dN(s)} \\ &= \lambda_0(t) + \sum_{T_i \leq t} \sum_{j=1}^{P} Y_j e^{-\delta_j (t-T_i)}.\end{aligned}$$

Observe that for $P = 1$ and $\lambda_0(t)$ constant we obtain

$$\lambda(t|\mathcal{H}_t) = \lambda_0 + \int_0^t Y e^{-\delta(t-s)} \mathrm{dN(s)} = \lambda_0 + \sum_{T_i \leq t} Y e^{-\delta(t-T_i)}.$$

In such case both Y and δ are constant.

Alan Hawkes introduced this general type of Hawkes process, where the current intensity rate of the arrival of points is determined by points in the past with associated weights [14]. In this project, we focus on a slightly modified version, where the sizes of self-excited jumps in the intensity process are allowed to be either fixed or random. The randomness in jump sizes could provide higher flexibility for measuring the self-exciting effect.

2.4.1 Intensity-based Hawkes Process Model

To avoid any confusion we are going to give the definition of intensity-based Hawkes process, which we are going to use later on, for simulation purposes.

Definition 2.4.4 (Intensity-based Hawkes process). Let

$$\lambda_0(t) = a + (\lambda_0 - a)e^{-\delta t},$$

where $a \geq 0$ is called the constant reversion level, $\lambda_0 > 0$ is the initial intensity at time $t = 0$, and $\delta > 0$ is the constant rate of exponential decay, then the intensity

of the Hawkes process is:

$$\lambda(t|\mathcal{H}_t) = a + (\lambda_0 - a)e^{-\delta t} + \sum_{T_k \leq t} Y_k e^{-\delta(t-T_k)}, \quad t \geq 0$$

where $\{Y_k, k = 1, 2, \ldots\}$ are sizes of self-excited jumps, a sequence of i.i.d. positive random variables with distribution function $G(y)$, $y > 0$, and the processes $\{T_k, k = 1, 2, \ldots\}$ and $\{Y_k, k = 1, 2, \ldots\}$ are assumed to be independent of each other.

Note that, in the long run, if there are no new events happening the intensity becomes equal to a, the constant reversion level. We also generally assume that $\lambda_0 > a$, which makes the intensity exponentially decaying when no events occur. If $\lambda_0 < a$, then the underlying intensity process $\lambda_0(t)$ is exponentially growing.

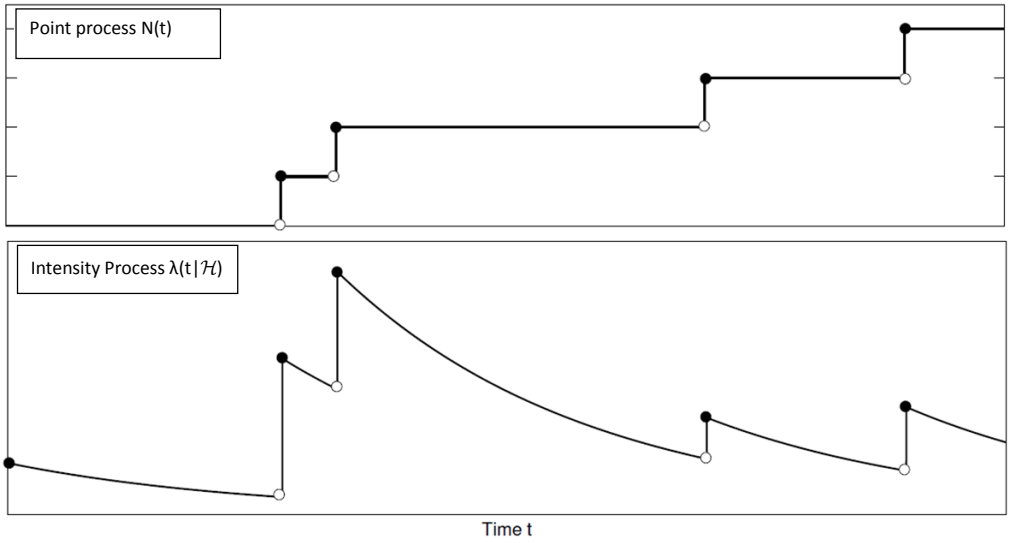


Figure 2.2: A Hawkes process with exponentially decaying intensity.

2.4.2 Cluster-based Hawkes Process Model

On the other hand, a Hawkes process $\{N(t), t \geq 0\}$ can also be alternatively constructed as a marked Poisson cluster process, with the clusters following a recursive branching structure, see [15].

This definition can lead to very accurate and easy to understand simulation algorithm (presented in Chapter 3).

Definition 2.4.5 (Cluster-based Hawkes Process). A Hawkes process with exponentially decaying intensity is a marked Poisson cluster process $C = \{(T_i, Y_i), i = 1, 2, \ldots\}$ with times $T_i \in \mathbb{R}_+$ and marks Y_i. The number of points in $(0, t]$ is defined by $N(t) = N_{C(0,t]}$, where $N_{C(0,t]}$ is the number of points in C up to the time t. The cluster centers of C are the particular points called immigrants, the rest of the points are called offspring, and they have the following structure:

1. The immigrants $\mathcal{I} = \{V_m, m = 1, 2, \ldots\}$ on $\mathbb{R}_+$ are distributed as an **nonhomogeneous** Poisson process of rate $\lambda_0(t) = a + (\lambda_0 - a)e^{-\delta t}, \quad t \geq 0$.

2. The marks $\{Y_m, m = 1, 2, \ldots\}$ associated to immigrants $\mathcal{I}$ are i.i.d. with $Y_m \sim G$, and are independent of the immigrants.

3. Each immigrant V_m generates one cluster C_m, and these clusters are independent.

4. Each cluster C_m is a random set formed by marked points of generations of order $n = 0, 1, \ldots$ with the following branching structure:

   - The immigrant and its mark (V_m, Y_m) is said to be of generation 0.
   - Recursively, given generations $0, 1, \ldots, n$ in C_m, each $(R_j, Y_j) \in C_m$ of generation n generates a Poisson process of offspring of generation $n + 1$ on (R_j, ∞) with intensity $Y_j e^{-\delta(t-R_j)}, \quad t > R_j$, where mark $Y_j \sim G$, independent of generations $0, 1, \ldots, n$.

5. C consists of the union of all clusters, i.e. $C = \bigcup_{m=1,2,..} C_m$.

The definition via marked Poisson cluster representation offers a nice economic interpretation: the immigrants and offspring could be considered as primary shocks and

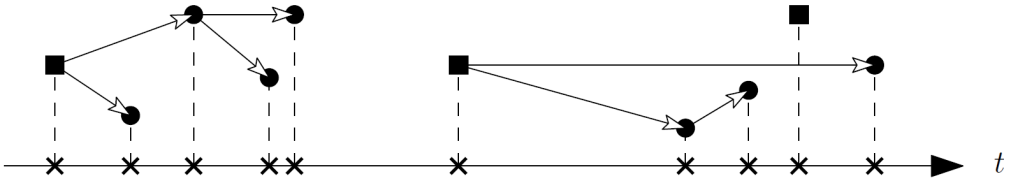


Figure 2.3: Hawkes process represented as a collection of family trees of immigrants. Squares indicate immigrants, circles are offspring, and the crosses denote the generated point process. Figure by [18]

associated aftershocks respectively. Note that, as this point process is defined on $\mathbb{R}_+$, there are no immigrants before time 0.

Note that Definition 2.4.4 and Definition 2.4.5 are equivalent. The stationarity condition for this Hawkes process is $\delta > \int_0^\infty y \mathrm{dF(y)}$ [6], although this is not required in the simulation algorithm we develop later.

2.5 Stochastic Time Change

There are some results about transforming one stochastic process into another by stretching and shrinking the time–scale, sometimes in a deterministic manner, more often by means of a random change of time-scale which depends on the realized trajectory of the process you started with. The simplest point process is the homogeneous Poisson process with unit intensity: there is a constant probability per unit time of an event happening, which we can normalize to 1. The consequence is that the distance between successive events is exponentially distributed, with mean 1.

Suppose we have a homogeneous Poisson process which does not have unit intensity, but one whose probability per unit time of an event is λ. To make a realization of this process look like a realization of the standard Poisson process, we take the time–axis and rescale it by λ. We stretch out the separation between points if the process has high intensity, or compress them if the process has low intensity. Symbolically, we define a new time,

$$\tau = t\lambda,$$

and then $X(\tau)$ looks like a realization of the standard Poisson process, even though $X(t)$ does not.

If the Poisson process is nonhomogeneous, the idea is similar. We want the new time to run slow and stretch out when the intensity is high, and we want it to run fast and compress events when the intensity is low. The right definition is

$$\tau(t) = \int_0^t \lambda(s)\mathrm{ds}.$$

More specifically, if $T_1, T_2, \ldots$ are the times of the events, then the transformed times $\tau(T_1), \tau(T_2), \ldots$ come from a standard Poisson process, and $\tau(T_{i+1}) - \tau(T_i)$ are iid and exponentially distributed with mean 1.

Now, we will move on to our case when the intensity function is not fixed, but depend on some random inputs. First let's define the integrated intensity function.

Definition 2.5.1 (Integrated Intensity). The integrated intensity function Λ is defined as

$$\forall i \in \mathbb{N}, \quad \Lambda(T_{i-1}, T_i) = \int_{T_{i-1}}^{T_i} \lambda(s|\mathcal{H}_s)\mathrm{ds}.$$

Theorem 2.5.1 (Time Change Theorem). Let N be point process on $\mathbb{R}_+$ such that $\int_0^\infty \lambda(s|\mathcal{H}_s)\mathrm{ds} = \infty$. Let t_τ be the stopping time defined by

$$\int_0^{t_\tau} \lambda(s|\mathcal{H}_s)\mathrm{ds} = \tau.$$

Then the process $\tilde{N}(\tau) = N(t_\tau)$ is a homogeneous Poisson process with constant intensity $\lambda = 1$.

It is still the case that $\tau(T_1), \tau(T_2), \ldots$ looks exactly like a realization of a standard Poisson process. But the way we have rescaled time is random, and possibly different from one realization of the original point process to another, because the times at which events happened can be included in the information represented by $\mathcal{H}_t$.

Idea of proof. The whole proof is presented in [2] on page 329.

We are trying to fit a model to a point process, so we have a set of possible guesses about the conditional intensity function, call it $\phi(t|\mathcal{H}_t;\theta)$, where θ is a parameter indexing the different functions. For one parameter value, call it θ_0, this is actually

right:

$$\phi(t|\mathcal{H}_t;\theta_0) = \lambda(t|\mathcal{H}_t).$$

If our guess at θ_0, call it $\hat{\theta}$, is right, then we can use $\phi(t|\mathcal{H}_t;\hat{\theta})$ to transform the original point process into a realization of a standard Poisson process. To test whether something is a realization of such a process we need to answer the question whether the inter-event times are exponentially distributed with mean 1 and whether they are independent. Since there are no free parameters in the hypothesis, one ends up testing. We can estimate the parameter however we like, and then do a back-end test on the goodness of fit. □

Chapter 3

Simulation Algorithms

There are two major simulation approaches in the literature: so called *intensity-based* and *cluster-based*, since a Hawkes process can either be defined via a conditional stochastic intensity representation (Definition 2.4.4) or via a Poisson cluster representation (Definition 2.4.5).

3.1 Simulation Algorithm - Cluster Based

This computational method is able to exactly generate the point process and intensity process, by sampling inter-arrival times directly via the underlying analytic distribution functions without numerical inverse, and hence avoids simulating intensity paths and introducing discretization bias. Moreover, it is flexible to generate points with either stationary or non-stationary intensity, starting from any arbitrary time with any arbitrary initial intensity. It is also straightforward to implement, and can easily be extended to multi-dimensional versions, for applications in modeling contagion risk or clustering arrival of events in finance, insurance, economics and many other fields.

In this algorithm $\lambda(t) = \lambda(t|\mathcal{H}_t)$.

3.1.1 Simulation Algorithm

Algorithm 3.1.1 (Dassios, Zhao). The simulation algorithm for one sample path of a one-dimensional Hawkes process with exponentially decaying intensity $\{(N(t), \lambda(t)), t \geq 0\}$

conditional on λ_0 and $N(0) = 0$, with jump-size distribution $Y \sim G$ and $\bar{K}$ jump-times $\{T_1, T_2, ..., T_{\bar{K}}\}$:

1. Set the initial conditions $T_0 = 0, \quad \lambda(T_0) = \lambda_0 > a, \quad N(0) = 0$, and $k \in \{0, 1, 2, ..., \bar{K} - 1\}$.

2. Simulate the $(k+1)^{th}$ inter-arrival time S_{k+1} by

$$S_{k+1} = \begin{cases} S_{k+1}^{(1)} \wedge S_{k+1}^{(2)}, & \text{if} \quad D_{k+1} > 0, \\ S_{k+1}^{(2)}, & \text{if} \quad D_{k+1} < 0, \end{cases}$$

   where

$$D_{k+1} = 1 + \frac{\delta \ln U_1}{\lambda(T_k^+) - a}, \quad U_1 \sim \text{UNIF}[0, 1],$$

   and

$$S_{k+1}^{(1)} = -\frac{1}{\delta} \ln D_{k+1}, \qquad S_{k+1}^{(2)} = -\frac{1}{a} \ln U_2, \quad U_2 \sim \text{UNIF}[0, 1].$$

3. Record the $(k+1)^{th}$ jump-time T_{k+1} by

$$T_{k+1} = T_k + S_{k+1}.$$

4. Record the change at the jump-time T_{k+1} in the process $\lambda(t)$ by

$$\lambda(T_{k+1}) = \lambda(T_{k+1}^-) + Y_{k+1}, \quad Y_{k+1} \sim G, \tag{3.1.1}$$

   where

$$\lambda(T_{k+1}^-) = (\lambda(T_k) - a) e^{-\delta(T_{k+1} - T_k)} + a.$$

5. Record the change at the jump-time T_{k+1} in the point process $N(t)$ by

$$N(T_{k+1}) = N(T_{k+1}^-) + 1 \tag{3.1.2}$$

Proof. Given the k^{th} jump-time T_k, the point process has the intensity process $\{\lambda(t), \quad T_k \leq t < T_k + S_{k+1}\}$ following the ordinary differential equation (between

the times of k^{th} and $(k+1)^{th}$ jump)

$$\frac{\mathrm{d}\lambda(t)}{\mathrm{d}t} = -\delta(\lambda(t) - a), \tag{3.1.3}$$

with the initial condition

$$\lambda(t)\Big|_{t=T_k} = \lambda(T_k). \tag{3.1.4}$$

To solve the equation (3.1.3), we divide it on both sides by $(\lambda(t) - a)$ and multiply by dt. We obtain

$$\frac{\mathrm{d}\lambda(t)}{\lambda(t) - a} = -\delta \mathrm{dt}. \tag{3.1.5}$$

Then we integrate equation (3.1.5) from 0 to t, to obtain

$$\ln(\lambda(t) - a) = -\delta t + C_1.$$

Therefore,

$$\lambda(t) - a = e^{-\delta t} \cdot C_2. \tag{3.1.6}$$

To evaluate the value of C_2 we need to use information provided by the initial condition (3.1.4). For $t = T_k$, equation (3.1.6) takes the following form:

$$\lambda(T_k) - a = e^{-\delta T_k} \cdot C_2.$$

Hence,

$$C_2 = \frac{\lambda(T_k) - a}{e^{-\delta T_k}} = (\lambda(T_k) - a)e^{\delta T_k}. \tag{3.1.7}$$

From (3.1.6) and (3.1.7) we get that

$$\begin{aligned} \lambda(t) - a &= e^{-\delta t} \cdot (\lambda(T_k) - a)e^{\delta T_k} \\ &= (\lambda(T_k) - a)e^{-\delta(t - T_k)}. \end{aligned}$$

As a result

$$\lambda(t) = (\lambda(T_k) - a)e^{-\delta(t-T_k)} + a, \quad \text{for} \quad T_k \leq t < T_k + S_{k+1}. \tag{3.1.8}$$

Let us evaluate the cumulative distribution function F of $(k+1)^{th}$ inter-arrival time S_{k+1}.

$$\begin{aligned} F_{S_{k+1}}(s) &= \mathbb{P}\{S_{k+1} \leq s\} \\ &= 1 - \mathbb{P}\{S_{k+1} > s\} \\ &= 1 - \mathbb{P}\{N(T_k + s) - N(T_k) = 0\} \end{aligned} \tag{3.1.9}$$

$\{N(s), s \geq 0\}$ is a nonhomogeneous Poisson process between the jump that occurs at time T_k and T_{k+1}, with the intensity function $\lambda(s)$. Then by Theorem 2.3.1 $\{N(T_k + s) - N(T_k), \quad s \geq 0\}$ is a Poisson random variable with mean $m(T_k + s) - m(T_k) = \int_{T_k}^{T_k+s} \lambda(y)\mathrm{dy}$. Therefore

$$\begin{aligned} \mathbb{P}\{N(T_k + s) - N(T_k) = 0\} &= e^{-(m(T_k+s)-m(T_k))} \cdot \frac{(m(T_k + s) - m(T_k))^0}{0!} \\ &= e^{-\int_{T_k}^{T_k+s} \lambda(u)\mathrm{du}}. \end{aligned}$$

As a result

$$F_{S_{k+1}}(s) = 1 - e^{-\int_{T_k}^{T_k+s} \lambda(u)\mathrm{du}}.$$

Now we change the integrating variable in the following way: $v = u - T_k$, and $u = T_k + v$, to obtain

$$F_{S_{k+1}}(s) = 1 - e^{-\int_0^s \lambda(T_k+v)\mathrm{d}v}.$$

We substitute $\lambda(T_k + v)$ with our result from (3.1.8) for $t = T_k + v$.

$$F_{S_{k+1}}(s) = 1 - e^{-\int_0^s [(\lambda(T_k)-a)e^{-(\delta T_k+v-T_k)}+a]\mathrm{d}v}$$

$$
\begin{aligned}
F_{S_{k+1}}(s) &= 1 - e^{-\int_0^s [(\lambda(T_k)-a)e^{-(\delta v)}+a]\mathrm{d}v} \\
&= 1 - e^{-(\lambda(T_k)-a)\cdot\frac{e^{-(\delta s)}+1}{-\delta} - as} \\
&= 1 - e^{-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta s)}}{\delta} - as} \qquad (3.1.10)
\end{aligned}
$$

By the inverse transformation method, we have

$$
S_{k+1} = F^{-1}_{S_{k+1}}(U), \qquad U \sim \mathrm{UNIF}[0,1].
$$

However, we can avoid inverting the function $F_{S_{k+1}}(\cdot)$, presented in equation (3.1.10), by decomposing S_{k+1} into two simpler and independent random variables $S^{(1)}_{k+1}$ and $S^{(2)}_{k+1}$ via

$$
S_{k+1} = S^{(1)}_{k+1} \wedge S^{(2)}_{k+1},
$$

where

$$
\begin{aligned}
\mathbb{P}\{S^{(1)}_{k+1} > s\} &= e^{-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta s)}}{\delta}}, \\
\mathbb{P}\{S^{(2)}_{k+1} > s\} &= e^{-as},
\end{aligned}
$$

since from (3.1.10)

$$
\begin{aligned}
\mathbb{P}\{S_{k+1} > s\} &= e^{-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta s)}}{\delta}} \cdot e^{-as} \\
&= \mathbb{P}\{S^{(1)}_{k+1} > s\} \cdot \mathbb{P}\{S^{(2)}_{k+1} > s\} \\
&= \mathbb{P}\{S^{(1)}_{k+1} > s, S^{(2)}_{k+1} > s\} \\
&= \mathbb{P}\{S^{(1)}_{k+1} \wedge S^{(2)}_{k+1} > s\} \qquad (3.1.11)
\end{aligned}
$$

Therefore we need to simulate $S^{(1)}_{k+1}$ and $S^{(2)}_{k+1}$.

We start by simulating $S^{(1)}_{k+1}$. Since

$$
F_{S^{(1)}_{k+1}}(s) = \mathbb{P}\{S^{(1)}_{k+1} \le s\} = 1 - e^{-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta s)}}{\delta}},
$$

we set

$$e^{-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta S^{(1)}_{k+1})}}{\delta}} = U_1.$$

We can invert this function explicitly in the following way:

$$\begin{aligned}
e^{-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta S^{(1)}_{k+1})}}{\delta}} &= U_1 \\
-(\lambda(T_k)-a)\cdot\frac{1-e^{-(\delta S^{(1)}_{k+1})}}{\delta} &= \ln(U_1) \\
-\frac{1-e^{-(\delta S^{(1)}_{k+1})}}{\delta} &= \frac{\ln(U_1)}{(\lambda(T_k)-a)} \\
-1+e^{-(\delta S^{(1)}_{k+1})}) &= \frac{\delta\ln(U_1)}{(\lambda(T_k)-a)} \\
e^{-(\delta S^{(1)}_{k+1})} &= 1+\frac{\delta\ln(U_1)}{(\lambda(T_k)-a)} \\
-(\delta S^{(1)}_{k+1}) &= \ln(1+\frac{\delta\ln(U_1)}{(\lambda(T_k)-a)}) \\
S^{(1)}_{k+1} &= -\frac{1}{\delta}\ln(1+\frac{\delta\ln(U_1)}{(\lambda(T_k)-a)})
\end{aligned} \tag{3.1.12}$$

Note that $S^{(1)}_{k+1}$ is a *defective* random variable, which means that it can take value ∞.

$$\lim_{s\to\infty} F_{S^{(1)}_{k+1}}(s) = \mathbb{P}\{S^{(1)}_{k+1} < \infty\} = 1 - e^{-\frac{\lambda(T_k)-a}{\delta}} < 1.$$

The condition for simulating a valid $S^{(1)}_{k+1}$ is

$$1+\frac{\delta\ln(U_1)}{\lambda(T_k)-a} > 0, \tag{3.1.13}$$

otherwise the natural logarithm could not be computed. That is why the variable D_{k+1} was introduced.

$$D_{k+1} = 1+\frac{\delta\ln(U_1)}{(\lambda(T_k)-a)}.$$

Then the condition (3.1.13) takes form $D_{k+1} > 0$.

Let us move on to the simulation of $S^{(2)}_{k+1}$.

Since $S_{k+1}^{(2)} \sim EXP(a)$, then using the standard simulation method we have

$$S_{k+1}^{(2)} = -\frac{1}{a}\ln U_2. \tag{3.1.14}$$

Hence, for the simulation of S_{k+1}, we have

$$S_{k+1} = \begin{cases} S_{k+1}^{(1)} \wedge S_{k+1}^{(2)}, & \text{if} \quad D_{k+1} > 0, \\ S_{k+1}^{(2)}, & \text{if} \quad D_{k+1} < 0, \end{cases}$$

where $S_{k+1}^{(1)}$ is given by (3.1.12) and $S_{k+1}^{(2)}$ is given by (3.1.14). Therefore, the $(k+1)^{th}$ jump-time T_{k+1} in the Hawkes process is given by

$$T_{k+1} = T_k + S_{k+1},$$

and the change in $\lambda(t)$ and $N(t)$ at time T_{k+1} then can be easily derived as given by (3.1.1) and (3.1.2), respectively.

□

Note that Algorithm 3.1.1 applies to any arbitrary distribution assumption G for self-excited jump-sizes $\{Y_k, k = 1, 2, \ldots\}$, and also to the case when jump-sizes are fixed.

3.1.2 Simulation Example

Let us simulate a Hawkes process, using algorithm 3.1.1, if jump sizes follow an exponential distribution of parameter β, with the probability density function:

$$g(y) = \beta e^{-\beta y}, \quad for \quad y, \beta > 0.$$

For $a = 0.89$, $\delta = 1$, $\lambda_0 = 0.9$, and $\beta = 1.2$, using the simulation Algorithm 3.1.1 we obtain the result presented in Figure 3.1.

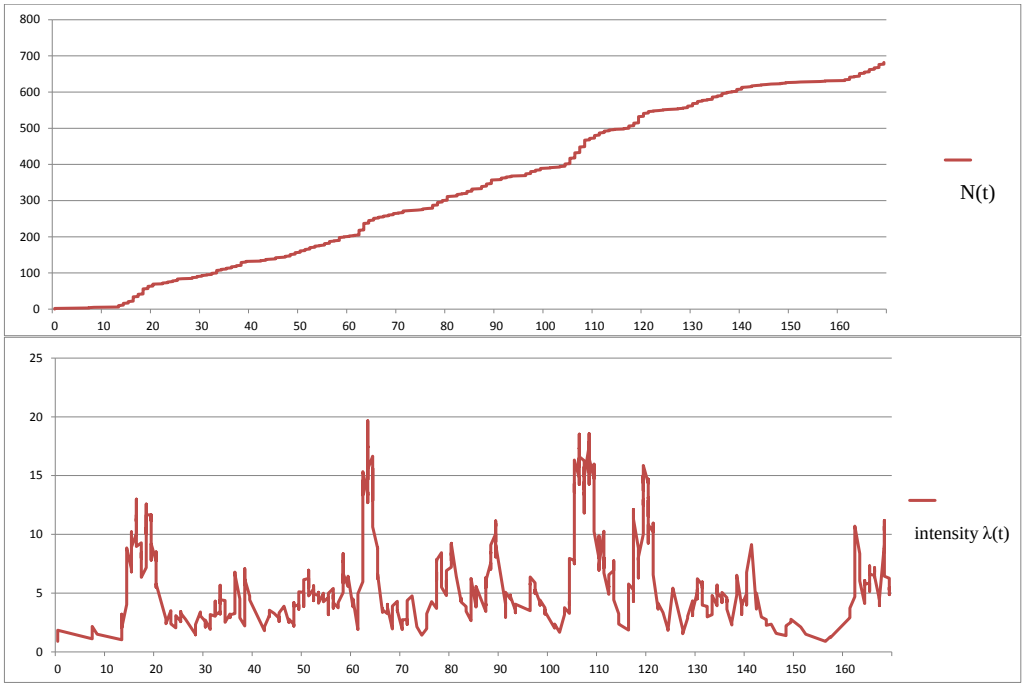


Figure 3.1: One simulated sample path of Hawkes process for $a = 0.89$, $\delta = 1$, $\lambda_0 = 0.9$, $\beta = 1.2$.

3.2 Simulation Algorithm - Intensity-based

3.2.1 Thinning Procedure

Lewis & Shedler (1979) propose a *thinning procedure* that allows the simulation of a point process with bounded intensity.

Theorem 3.2.1 (Basic Thinning Theorem). (Lewis, Shedler) Consider a one-dimensional non-homogeneous Poisson process $\{N^*(t), t \geq 0\}$ with rate function $\lambda^*(t)$, so that the number of points $N^*(T_0)$ in a fixed interval $(0, T_0]$ has a Poisson distribution with parameter $\mu_0^* = \Lambda^*(0, T_0) = \int_0^{T_0} \lambda^*(s)\text{ds}$. Let $T_1^*, T_2^*, \ldots, T_{N^*(T_0)}^*$ be the points of the process in the interval $(0, T_0]$. Suppose that for $0 \leq t \leq T_0$, $\lambda(t) \leq \lambda^*(t)$.
For $i = 1, 2, \ldots, N^*(T_0)$, delete the points T_i^* with probability $1 - \frac{\lambda(T_i^*)}{\lambda^*(T_i^*)}$. Then the remaining points form a non-homogeneous Poisson process $\{N(t), t \geq 0\}$ with rate function $\lambda(t)$ in the interval $(0, T_0]$.

Proof. Since $\{N^*(t), t \geq 0\}$ is nonhomogeneous Poisson process, and points are deleted independently, the number of points $\{N(t) : t \geq 0\}$ in any set of nonoverlapping intervals are mutually independent random variables. Thus it is sufficient to show that the number of points in $\{N(t) : t \geq 0\}$ in any arbitrary interval $(a, b]$ with $0 \leq a < b \leq T_0$, denoted by $N(a, b)$, has a Poisson distribution with parameter

$$\Lambda(0, b) - \Lambda(0, a) = \int_a^b \lambda(s)\mathrm{ds}.$$

Observe that with

$$p(a, b) = \frac{\Lambda(0, b) - \Lambda(0, a)}{\Lambda * (0, b) - \Lambda * (0, a)} = \frac{\int_a^b \lambda(s)\mathrm{ds}}{\int_a^b \lambda^*(s)\mathrm{ds}}$$

we have conditional probability

$$\mathbb{P}\{N(a, b) = n | N^*(a, b) = k\} = \begin{cases} 1 & \text{if} \quad n = k = 0 \\ (p(a, b))^n (1 - p(a, b))^{k-n} & \text{if} \quad k \geq n \geq 0 \\ & \quad \text{and} \quad k \geq 1, \\ 0 & \text{if} \quad n \geq 1, \\ & \quad \text{and} \quad k < n. \end{cases} \tag{3.2.1}$$

Equation (3.2.1) is a consequence of the fact that, conditional on n points in the interval $(a, b]$, the joint density of the n points in the process $\{N^*(t), t \geq 0\}$ is

$$\frac{\lambda^*(T_1) \cdots \lambda^*(T_n)}{(\Lambda * (0, b) - \Lambda * (0, a))^n}.$$

The desired result is obtained from equation (3.2.1) by removing the condition. □

Theorem 3.2.1 is the basis for the method of simulating nonhomogeneous Poisson process and Hawkes process.

3.2.2 Simulation Algorithm

Ogata (1981) proposes an algorithm for the simulation of Hawkes processes.

Algorithm 3.2.1. (Ogata) Let us denote by UNIF$[0,1]$ the uniform distribution on the interval $[0,1]$ and $[0,\mathcal{T}]$ the time interval on which the process is to be simulated. We assume here that $P = 1$.

1. **Initialization:** Set $\lambda^* \leftarrow \lambda_0(0), \quad n \leftarrow 1$.

2. **First event**: Generate $U \sim$ UNIF$[0,1]$ and set $s \leftarrow -\frac{1}{\lambda^*}\ln U$.
   If $s \leq \mathcal{T}$,
   Then $T_1 \leftarrow s$,
   Else go to last step.

3. **General routine**: Set $n \leftarrow n+1$.

   (a) **Update maximum intensity:** Set $\lambda^* \leftarrow \lambda(T_{n-1}^-) + Y$.
   λ^* exhibits a jump of size Y as an event has just occurred.

   (b) **New event**: Generate $U \sim$ UNIF$[0,1]$ and set $s \leftarrow s - \frac{1}{\lambda^*}\ln U$.
   If $s \geq \mathcal{T}$,
   Then go to last step.

   (c) **Rejection test**: Generate $D \sim$ UNIF$[0,1]$
   If $D \leq \frac{\lambda(s)}{\lambda^*}$,
   Then $T_n \leftarrow s$, and go through general routine again
   Else update $\lambda^* \leftarrow \lambda(s)$ and try a new date at step (b) of general routine.

4. **Output**: Retrieve the simulated process $\{T_n\}$ on $[0,\mathcal{T}]$.

3.2.3 Simulation Example

Let us simulate Hawkes process, using Algorithm 3.2.1. For set values of λ_0, Y, δ and $\mathcal{T}$ we obtained a Hawkes (self-exciting) process; a graph of its intensity, along with the event times, is presented in Figure 3.2. We see that as defined an arrival (an event) causes the conditional intensity function to increase. This behavior causes temporal clustering.

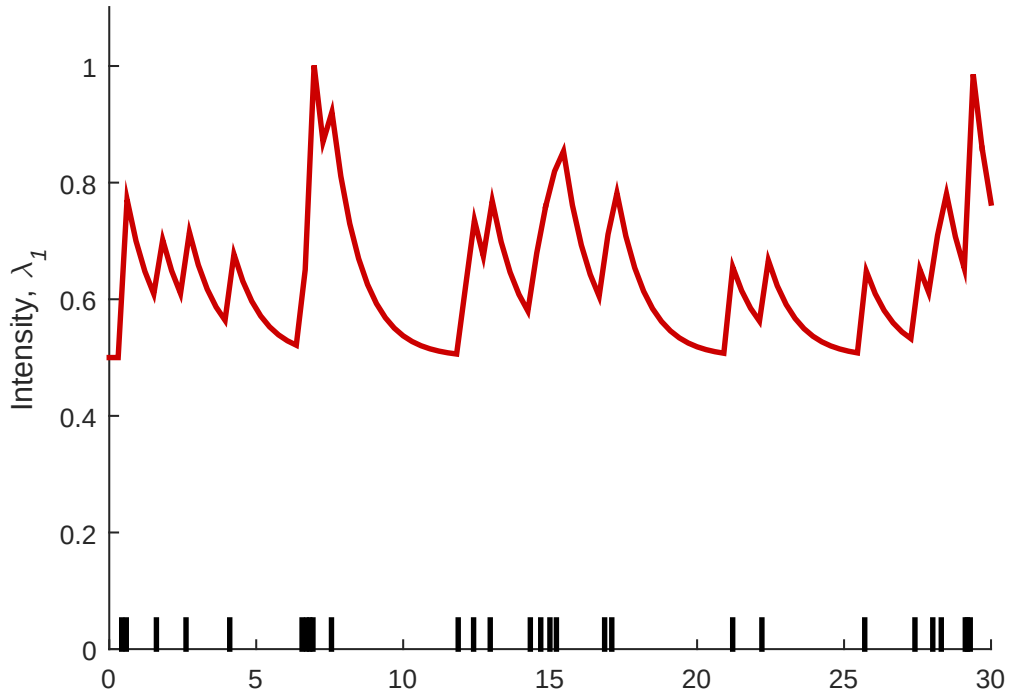


Figure 3.2: A simulation of self-exciting Hawkes process for $\lambda_0(t) \equiv 0.5$, $Y \equiv 0.15$, $\delta = 1$ and $\mathcal{T} = 30$. The graph of intensity $\lambda(t|\mathcal{H}_t)$ is presented, with the times of events marked on the time-axis.

In this project we focus our attention on self-exciting processes. However, self-regulating (sometimes also called self-damping) process can also be defined by using the conditional intensity function: if the conditional intensity function drops after an arrival (meaning that we assign a negative mark to the event) the process is called self-regulating and the arrival times appear quite temporally regular. Such processes are not examined hereafter, though an illustrative example would be the arrival of speeding tickets to a driver over time (assuming each arrival causes a period of heightened caution when driving). A self-regulating process example can be seen on the Figure 3.3.

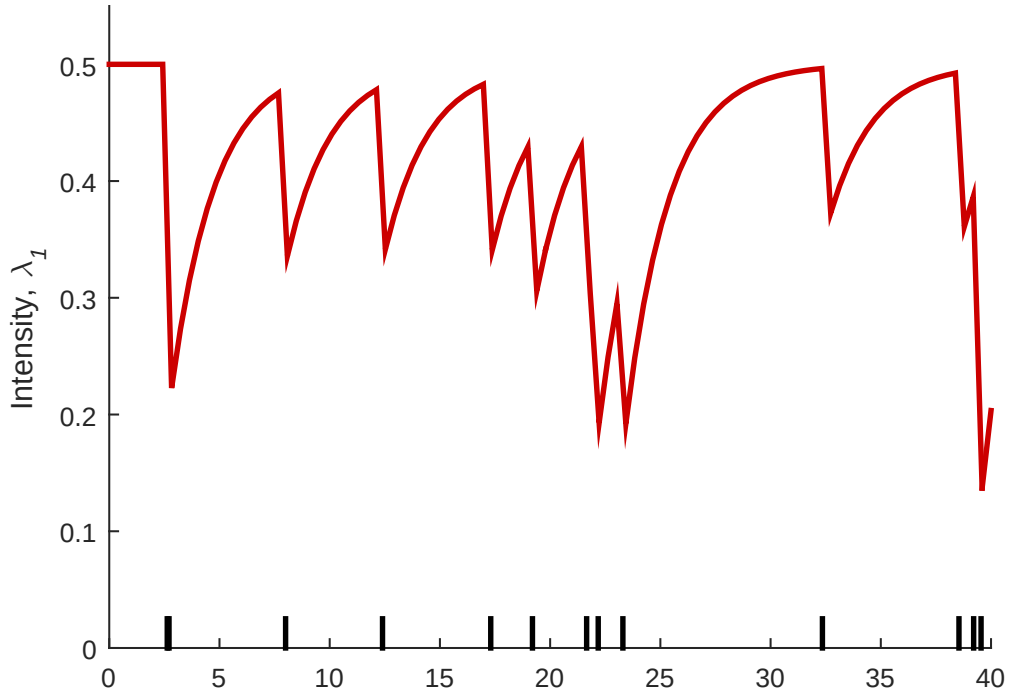


Figure 3.3: A simulation of self-regulating Hawkes process for $\lambda_0(t) \equiv 0.5$, $Y \equiv -0.15$, $\delta = 1$ and $\mathcal{T} = 40$. The graph of intensity $\lambda(t|\mathcal{H}_t)$ is presented, with the times of events marked on the time-axis.

Chapter 4

Multidimensional Hawkes Processes

Definition 4.0.1. A *multidimensional Hawkes process* is defined with intensities λ_m, $m = 1, \ldots, M$ given by

$$\lambda_m(t) = \lambda_{0,m}(t) + \sum_{n=1}^{M} \int_0^t \sum_{j=1}^{P} Y_{j,m,n} e^{-\delta_{j,m,n}(t-s)} \mathrm{dN_n(s)}.$$

Observe that for $P = 1$ and $\lambda_{0,m}(t)$ constant,

$$\begin{aligned} \lambda_m(t) &= \lambda_{0,m} + \sum_{n=1}^{M} \int_0^t Y_{m,n} e^{-\delta_{m,n}(t-s)} \mathrm{dN_n(s)} \\ &= \lambda_{0,m} + \sum_{n=1}^{M} \sum_{T_{i,n} \leq t} Y_{m,n} e^{-\delta^{m,n}(t-T_{i,n})}. \end{aligned}$$

Using vectorial notation, we have:

$$\boldsymbol{\lambda}(t) = \boldsymbol{\lambda_0} \int_0^t \boldsymbol{G}(t-s) \mathrm{d}\boldsymbol{N}(s)$$

$$\boldsymbol{G}(t) = \big(Y_{m,n} e^{-\delta_{m,n}(t-s)}, \quad m, n = 1, \ldots, M\big).$$

Assuming stationarity gives $\mathbb{E}[\boldsymbol{\lambda}(t)] = \boldsymbol{\mu}$ constant vector, and thus stationary intensities

must satisfy:

$$\boldsymbol{\mu} = \left(\boldsymbol{I} - \int_0^\infty \boldsymbol{G}(u)\mathrm{du}\right)^{-1} \boldsymbol{\lambda}_0.$$

An example of bivariate Hawkes process is presented in the Figure 4.1.

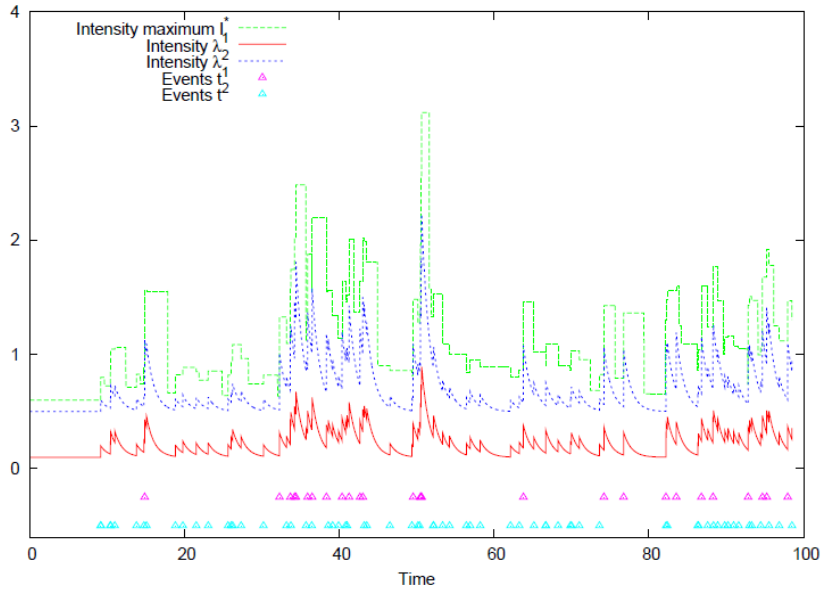


Figure 4.1: Example of bivariate Hawkes process, by [23].

Chapter 5

Intensity Estimation

In this chapter we will use a real data example to show how an intensity process of a Hawkes process can be estimated with the help of the R package STAR, created by Christophe Pouzat.

Spike trains of several cockroach antennal lobe neurons were recorded from six cockroaches (Periplaneta Americana). Antennal lobe neurons (putative projection neurons) were recorded simultaneously and extracellularly during spontaneous activity and odors (vanilin, citral, citronellal, terpineol, beta-ionon) responses from six different animals. The data sets contained the sorted spike trains of the neurons. Recording and spike sorting was performed by Antoine Chaffiol at the Cerebral Physiology Lab, at CNRS (*Centre national de la recherche scientifique*, the French National Center for Scientific Research), UMR (*unités mixtes de recherche*, mixed organization) 8118.

We analyze the discharge of one neuron of one animal, it will be neuron 1 of data set e060824spont, which we call just neuron 1.

We start by getting a quick summary of neuron 1 spike train. It is a spike train with 505 events, starting at time 0.594 seconds and ending at: time 58.585 seconds. The mean inter-spike interval, which we will denote later as (ISI) is: 0.115 seconds and its standard deviation is: 0.36 seconds. The shortest interval is: 0.008 seconds and the longest interval is: 3.811 seconds.

So far eyeballing is still considered a relatively good method in checking whether a stochastic process resembles a Hawkes process. That is why we start by obtaining a graph of the point process (for which events occur at discharge time of neuron) and

the associated counting process. Such type of graph is called by Christophe Pouzat in [27] a *spike train plot*.

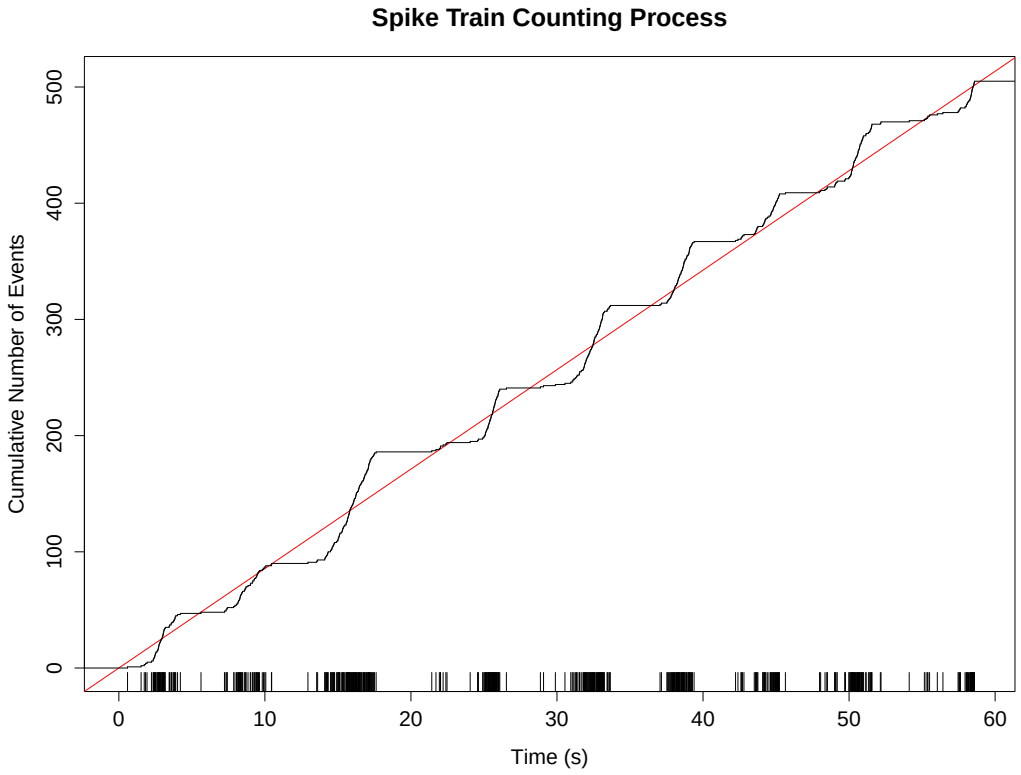


Figure 5.1: The *spike train plot* of neuron 1.

A staircase pattern can be seen on the realization of the counting process in Figure 5.1. Also the non-uniform distribution of the ticks on the raster plot shown at the bottom of the graph rules out a model based on a homogeneous Poisson process for this spike train. We observe the clustering, which is typical for Hawkes processes. Furthermore, biophysics teaches us that every neuron exhibits a refractory period following a spike [27] (ruling out the homogenous Poisson process as a true model) and that will lead us to always include the elapsed time since the last spike in our models.

We have to include then the occurrence times of the other previous spikes, or equivalently, the duration of the previous inter-spike intervals (ISIs) in creating the intensity model. How many previous inter-spike intervals should we then include in our variables

list? The answer to this question will be provided by the *partial autocorrelation function* of the ISIs [17]. A graph of this function is shown on Figure 5.2.

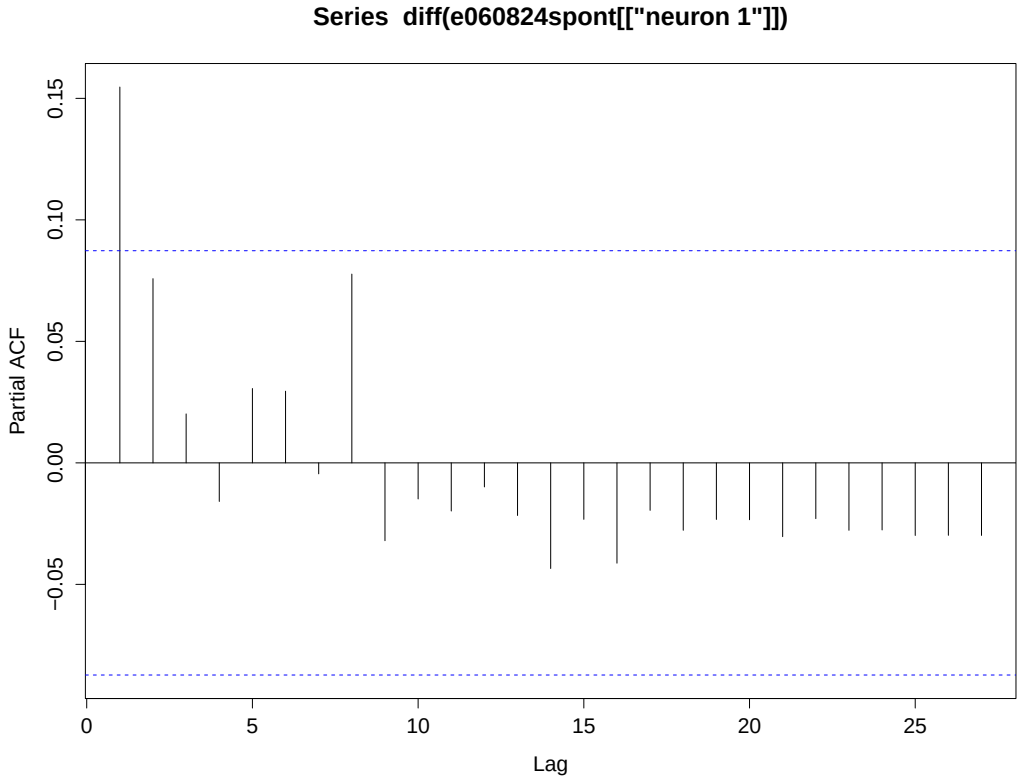


Figure 5.2: The *partial autocorrelation function* of neuron 1.

The dashed lines show the 95% confidence intervals. We observe that for the lag 1 the function is out of this 95% CI. This analysis would lead us to a model suggested in [27]:

$$\lambda(t|\mathcal{H}_t) = f(t - T_l, i_1)$$

where T_l is the occurrence time of the last spike before t, ($T_l = \max\{T_j : Tj < t\}$. In other words, $t - T_l$ is the elapsed time since the last spike, and i_1 is the duration of the last ISI (inter-spike interval).

We follow the approach of Brillinger [4], where a discretization of the spike train is performed. We go from the actual train (point process), $\{T_1, \ldots, T_n\}$ to a binary vector, E whose j_{th} element is the value of the sum over $\{T_1, \ldots, T_n\}$ of the indicator

function I_j defined by:

$$I_j(t) = \begin{cases} 1 & \text{if} \quad (j-1)\varepsilon < t \leq j\varepsilon \\ 0 & \text{otherwise,} \end{cases}$$

where the bin width ε is chosen small enough to have at most one spike per bin. We have then:

$$E_j = \sum_{k=1}^{n} = I_j(T_k).$$

So we do not estimate $f(t - T_l, i_1)$ directly but

$$f_\varepsilon(t - T_l, i_1) \equiv f((j - j_l)\varepsilon, (j_l - j_{l-1})\varepsilon)\varepsilon$$

where j is index of the bin containing time t, j_l is the index of the bin of the previous spike and j_{l-1} is the index of the second previous spike. f_ε that is the number between 0 and 1. We will use a nonparametric approach where $f_\varepsilon(t - T_l, i_1)$ will be estimated with a penalized likelihood method, described in [28]. As we mentioned before, all the operations in R can be easily performed with the help of the library STAR, created by Christophe Pouzat.

Now we will look at the evolution of the variables E and i_1 over time.

Figure 5.3 does not exhibit any clear trend in the graphed variables, therefore the process seems to be stationary.

We were using nonparametric model estimation procedures, following [27]. Since we wanted to have meaningful goodness of fit tests we have been fitting given model to one half of the data and then testing it on the other half before switching the fit and test data parts and repeating the procedure.

We were assessing the quality of the chosen model by evaluating the intensity process of the part of the data that we had not use for model estimation. This intensity process was then used to perform a time transformation (as in Theorem 2.5.1) after which a new counting process was obtained (this method was proposed in [25]). If our model was good this process was the realization of a homogeneous Poisson process with rate 1. The latter process was then the null hypothesis against which we were doing tests (test of uniformity on the time axis, Wiener process test, Berman test, renewal test and

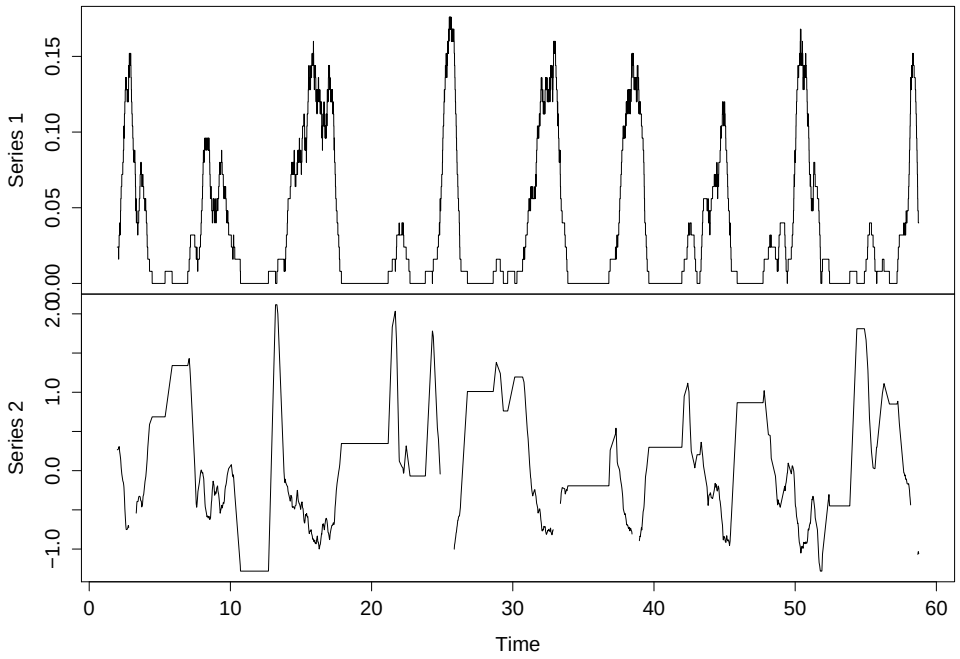


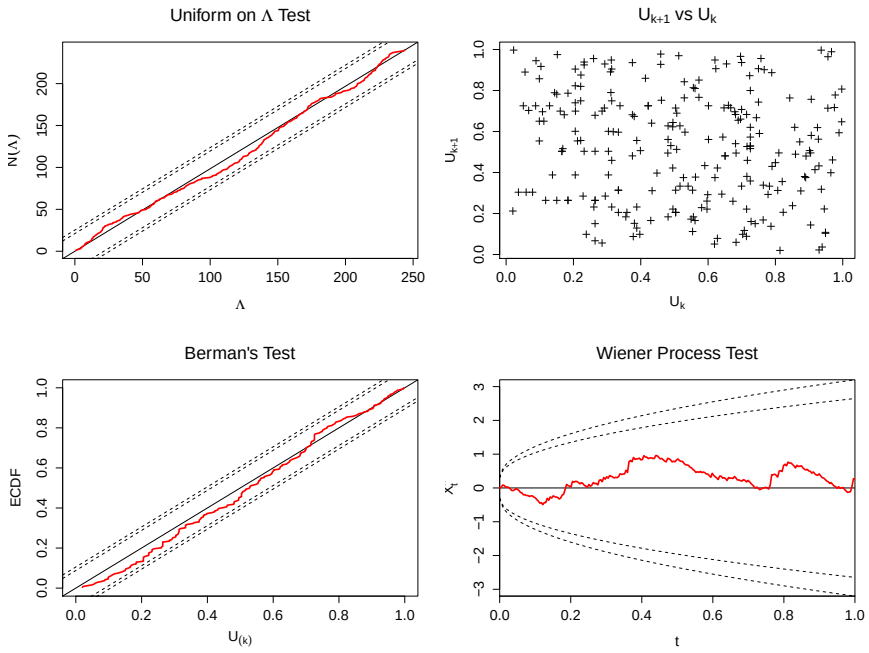
Figure 5.3: Time evolution of E_j and i_1

variance vs time test with 5 time windows) all those tests are thoroughly explained in [29]. All the functions for time transformation were already defined in package STAR.

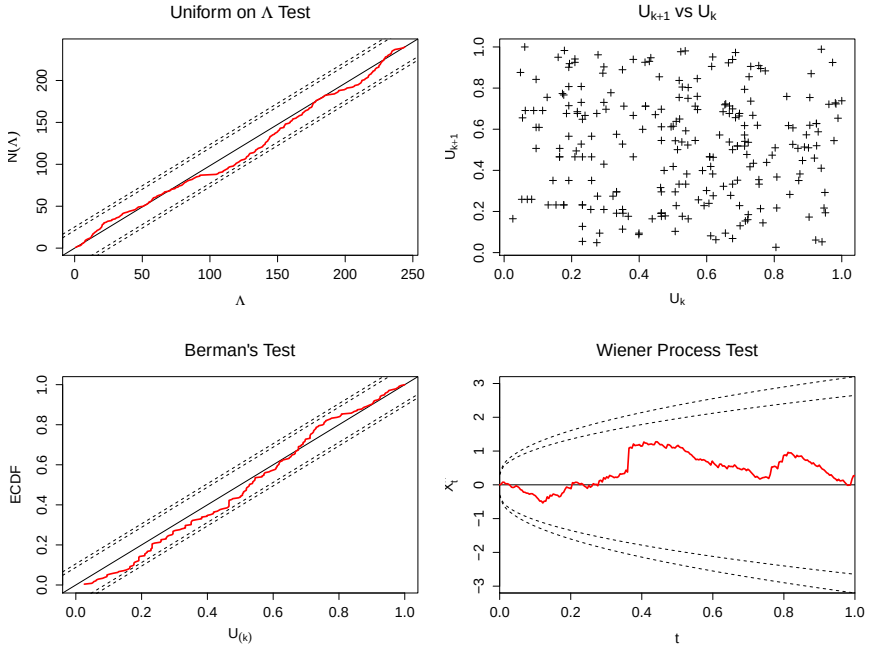
We have first explored a model containing an interaction term between variable E and the variable i_1. It was interesting to try simplifying it to see if a model without interaction would not give as good results. Therefore, we tested also the model without the interaction. We ran the same tests as for the model with interaction and then we compared which model was a better fit.

The diagnostic plots are shown on Figure 5.4. The simpler model looks as good as the first one. Which model should we choose then?

Following the idea shown in [27] we look at the probability the models give to data which were not the data used to fit them. This can be done with a function already defined in the STAR package, which returns the log probability of some data under some model. The log probability of our data using the simpler model is $-1775.767.$, which is smaller than -1762.754 which the log probability of more complex model (with



(a) Results for model with interaction.



(b) Results for model without interaction.

Figure 5.4: Tests applied to the time transformed neuron 1 spike train.

interaction). Therefore we chose the model with interaction.

We conclude the analysis of the spike train of neuron 1 by looking at the intensity process obtained with our two models (with and without interaction) on a small part of the spike train. We first get the predicted value of f_ε of on the logit scale for the second half of the second half of the data set using the fit obtained from the first half.

The plot showing the intensity process of the two models appears on Figure 5.5.

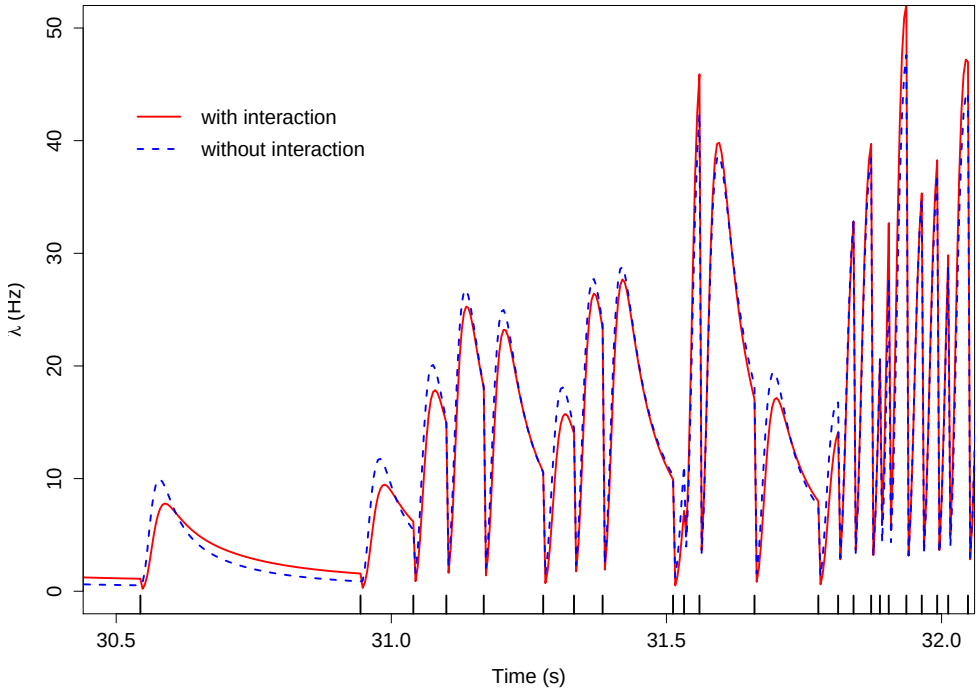


Figure 5.5: The intensity processes of the two considered models, with (continuous) and without (dashed) interactions between the elapsed time since the last spike and the last ISI (inter-spike interval).

Notice that the intensity process of the model with interaction (between the elapsed time since the last spike and the last ISI) is almost always larger than the one of the other model just before the spike while it tends to be smaller in between the spikes. In other words this model predicts a lower event probability when there is actually no event and a larger probability when there are events. Which confirms that it is a better

model for the intensity process.

Summing up, we managed to obtain the underlying random intensity process from the given Hawkes process. Since it is a real-life example, it does not have fixed mark values, we also do not know what is the the distribution of these marks. To obtain the graph of the intensity process we used the functions provided by Christophe Pouzat in package STAR, which are numerical methods. The explicit formula for the intensity process $\lambda(t|H_t)$ is almost impossible to obtain.

The fact that the discharge for neuron 1 of a cockroach did behave like a Hawkes process suggests that it might be also the case for other neurons. In fact, Patricia Reynaud-Bouret, from whom I did obtain the data used in this project, is doing a lot of research in neuroscience. She investigates Hawkes processes application in neurosciences and genomics (see for example [30] and [31]).

Chapter 6

Applications

6.1 Seismic Events

In the 1970s, professor Alan Hawkes introduced a family of probability models for the occurrence of series of events, to be used in seismology. As we have already established, they are called Hawkes processes. These processes were further studied in a seismology context by among others Y. Ogata [25] and L. Adamopoulos [1]. We can see in Figure 6.1 that the number of shocks in periods of three months for an area of the North Atlantic resembles the stochastic intensity function of a Hawkes process.

Therefore, the ETAS (epidemic type aftershock sequence) model was introduced. This model is a particular type of marked Hawkes process for modeling earthquake times and magnitudes. Here $\kappa_i \in [0, \infty)$ denotes the magnitude of an earthquake occurring at time T_i. In its simplest form the ETAS model can be defined by its intensity

$$\lambda(t) = \lambda_0 + \alpha \sum_{T_i<t} e^{\beta\kappa_i} e^{-\delta(t-T_i)},$$

where $\alpha, \beta, \delta > 0$ are parameters, with an exponential distribution as its mark density

$$f(\kappa|t) = \gamma e^{-\gamma\kappa}.$$

Equivalently we could define it by its conditional intensity function including both marks

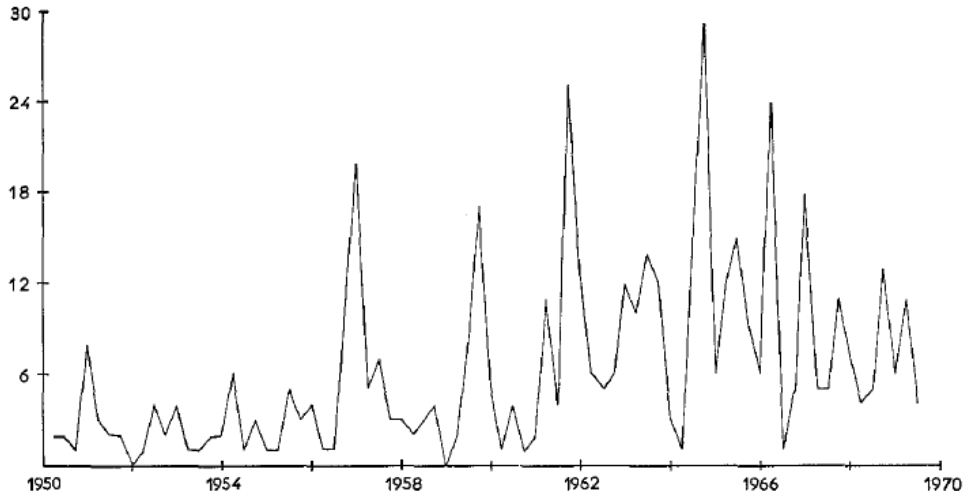


Figure 6.1: Number of shocks in periods of three months for area of North Atlantic. Figure by [1]

and times

$$\lambda(t,\kappa) = \big(\lambda_0 + \alpha \sum_{T_i<t} e^{\beta\kappa_i} e^{-\delta(t-T_i)}\big)\gamma e^{-\gamma\kappa}.$$

The idea behind using this model is that earthquakes cause aftershocks – this is reflected in the fact that every new earthquake increases the intensity by $\alpha e^{\beta\kappa_i}$. Note that large earthquakes increase the intensity more than small earthquakes. Note that this model does not have independent marks, since the intensity depends on the past marks. For more on the ETAS model, see [25].

6.2 Crime Data for Duluth Area

Applications to crime have been published by Egesdal, Fathauer, Louie & Neuman in Los Angeles [8]. Rivalry between two gangs can be modeled by considering two kinds of events: gang a attacks gang b, and gang b attacks gang a. Following an event of one type the next event might be another similar attack or a retaliation by the other gang. Following their work, Mohler et al. have published their paper on crime modeling, for urban crime and residential burglary [22].

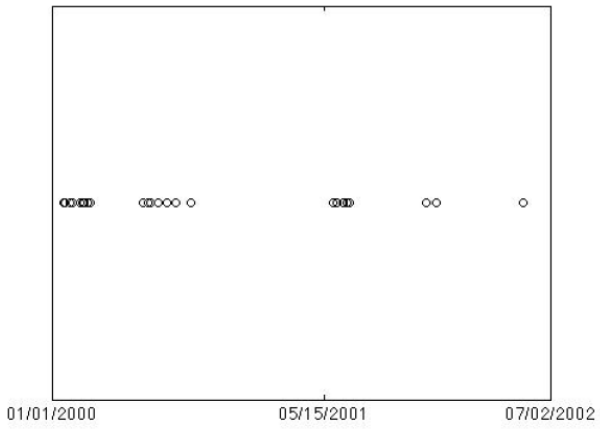


Figure 6.2: Times of violent crimes between two rival gangs in Los Angeles. Figure by [22]

In Figure 6.2, we see a plot of the times of all recorded violent crimes between the gang known as *Locke Street* and the rival gang known as *Lowell Street* occurring between 2000 and 2002 in the Los Angeles police district of Hollenbeck. We observe clear clustering patterns suggestive of self-excitation in the rate at which the two rival gangs attack each other.

Mohler et al. manage to show in [22] that the crime process is a Hawkes process. Therefore we wanted to do a similar analysis for the crime in Duluth. We did collect the crime data using website www.raidsonline.com. In our analysis we decided to only include the crimes which seemed to have some connection with possible gangs. Therefore, we used as our data points events of homicide, attempted homicide, death investigation, robbery, aggravated assault, burglary, arson, drug related crimes and weapons violations crime. We did not include events like shoplifting or traffic incidents etc. We created a plot of the intensity process, based on those events. Unfortunately the intensity process, presented in the Figure 6.3, does not look like an intensity process of a Hawkes process.

It seems to resemble rather an intensity process for a Poisson process. We need to admit that it does make sense that the crime data for Duluth area is not a Hawkes process. There are no gangs (at least no big gangs) in here, and for sure (we hope)

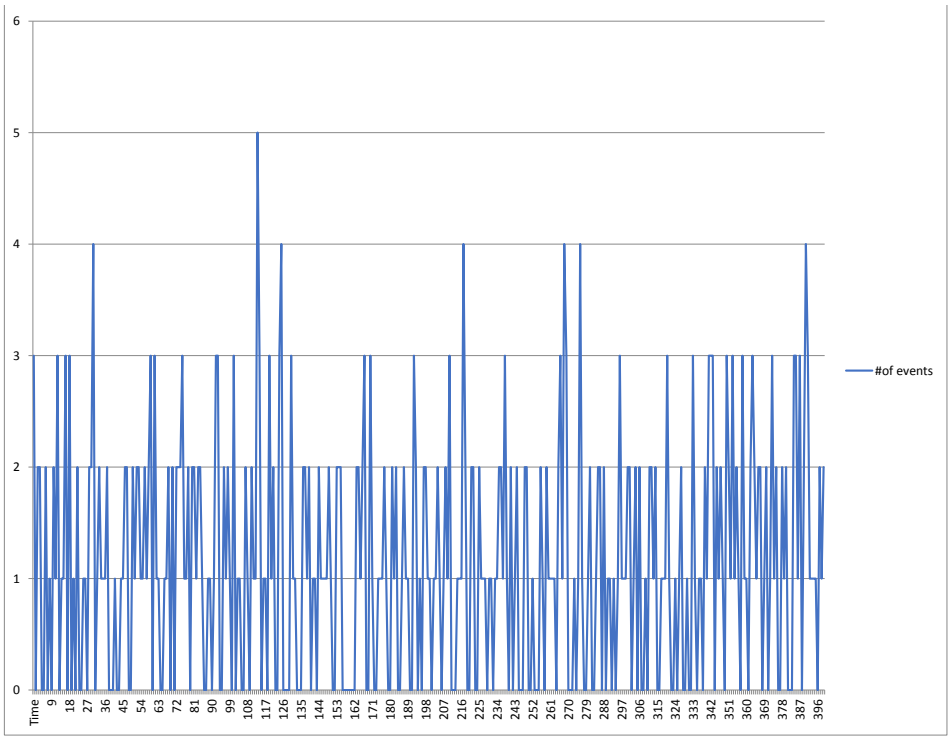


Figure 6.3: The intensity process of crime in Duluth.

there are no rival gangs, which would significantly influence the crime figures. We did look at some data from other areas (e.g. Oakland), without much success though, when it comes to discovering Hawkes processes in crime data. We decided that the purpose of this project, however noble and important, is not worth becoming a gang member to get better access to such data, and decided to analyze the neurons of cockroaches, as shown in Chapter 5, as the exemplary analysis of a Hawkes process.

6.3 Insurance Company Model

6.3.1 Classic Insurance Company Model

In 1903, F. Lundberg laid the foundation of risk theory in his thesis [20], where he introduced a risk model based on a homogeneous Poisson claim number process. Since

that time, this risk model has attracted much attention by mathematicians and actuaries. In particular, H. Cramér who incorporated F. Lundberg's ideas into the theory of stochastic processes [5]. The famous Cramér-Lundberg Theorem for small claims (presented in the following subsection) states that the infinite time ruin probability of an insurer can be bounded from above by an exponential function with an explicitly given exponent, if we assume a positive safety loading (defined later).

The risk process is supposed to model the time evolution of the reserves of an insurance company. In the following, we will first use the classic Cramér-Lundberg model to describe the risk process R: the process R is given by a Poisson process N with intensity λ and by a positive random variable X, independent of the process N, with distribution function F in the following way

$$R(t,x) = x + ct - \sum_{i=1}^{N(t)} X_i, \tag{6.3.1}$$

where $x > 0$ is the initial reserve of the insurance company, $c \in \mathbb{R}$ is the (constant) premium rate over time and $\{X_i, i = 1, 2, \ldots\}$ is an i.i.d. sequence, where X_i is modeling the i^{th} loss of the insurance company. The process $\{N(t), t \geq 0\}$ models the number of claims that occur in the time interval $[0, t]$. It has been a central topic of classical risk theory to obtain information about the *ruin probability* of an insurance company. The event *ruin* is defined as the first point of time, where the reserve of the insurance company drops below the level 0. The ruin probability is therefore defined as

$$\Psi(x) = \mathbb{P}[R(t,x) < 0 \text{ for } t \geq 0], \tag{6.3.2}$$

and the corresponding *time to ruin*

$$\tau(x) := \inf \{t \geq 0 : R(t,x) < 0\}. \tag{6.3.3}$$

Name	Tail $\overline{F}$ or density f	Parameters
Exponential	$\overline{F}(x) = e^{-\lambda x}$	$\lambda > 0$
Gamma	$f(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}$	$\alpha,\ \beta > 0$
Weibull	$\overline{F}(x) = e^{-cx^\tau}$	$c > 0,\ \tau \geq 1$
Truncated Normal	$f(x) = \sqrt{\frac{2}{\pi}} e^{-x^2/2}$	
Any distribution with bounded support		

Table 6.1: Light-tailed claim size distributions. All distributions have support $(0, \infty)$. Table by [10]

Observe that

$$\mathbb{E}[R(t,x)] = \mathbb{E}[x + ct - \sum_{i=1}^{N(t)} X_i] = x + ct - \lambda t \mathbb{E}[X_1].$$

Therefore

$$\lim_{t\to\infty} = \frac{\mathbb{E}[R(t,x)]}{t} = c - \lambda \mathbb{E}[X_1]$$

In order to exclude the case of almost sure ruin, it is reasonable to impose $c - \lambda\mathbb{E}[X_1]$. This leads to the definition of the *safety loading*,

$$\rho = \frac{c}{\lambda \mathbb{E}[X_1]} - 1$$

The condition $c - \lambda\mathbb{E}[X_1]$ is equivalent to a positive safety loading $\rho > 0$.

One can roughly classify the most popular classes of distributions which have been used to model the claim size X_i into two groups, light-tailed distributions and heavy-tailed distributions.

Light-tailed claim distributions are defined as distributions whose tail distribution function $\bar{F} = 1 - F(x)$ satisfies $\bar{F} = O(e^{-rx})$ for some $r > 0$, where O was introduced in Definition 2.2.1. Such light-tailed distributions are also called distributions with exponential moments. Examples for distributions that are light-tailed are given in Table 6.1.

6.3.2 Cramér-Lundberg Theorem

We are going to present classical results on ruin probabilities. In order to do so, for light-tailed claim distribution functions, we need to define the function

$$h(r) = \mathbb{E}[e^{rX}] - 1 = \int_0^\infty e^{rz} \mathrm{d}F(z) - 1, \quad r \geq 0. \tag{6.3.4}$$

We assume that there exists $r_\infty \in (0, \infty]$ such that $h(r) < \infty$ for $r < r_\infty$ and such that $h(r) \to \infty$, for $r \to r_\infty$. The function h has the following properties: $h(0) = 0$, h is increasing, convex, and continuous on [0,r_∞][13].

Theorem 6.3.1 (Cramér-Lundberg Theorem for Small Claims). Consider Cramér-Lundberg model, with the claim sizes coming form a light-tailed distribution. Assume a positive safety loading $\rho > 0$. Then there exist a number $0 < \nu < r_\infty$ such that that:

$$\lambda h(\nu) = c\nu, \tag{6.3.5}$$

and the following statement is true.

For every $x \geq 0$

$$\Psi(x) \leq e^{-\nu x}. \tag{6.3.6}$$

The classic proof of this result uses renewal theory (see [9]). In 1973, H. Gerber [12] gave an alternative proof, making use of martingale methods.

6.3.3 Risk Process with Non-stationary Hawkes Claims Arrivals

Stabile and Torrisi in [35] consider risk processes with non-stationary Hawkes claims arrivals, and they study the asymptotic behavior of infinite and finite horizon ruin probabilities under defined by us light-tailed conditions on the claims.

They introduce the following risk model for the surplus process (risk process)

$$R(t, x) = x + ct - \sum_{i=1}^{N^*(t)} X_i, \tag{6.3.7}$$

where $\{N^*(t), t \geq 0\}$ is the number of points of a non-stationary Hawkes process,

in the time interval $(0, t]$, $x, c > 0$ and $\{X_i, i = 1, 2, \ldots\}$ are the same as for the classic model.

For the sake of this model it is better to look at Hawkes process as a realization of the cluster-based Definition 2.4.5. The interpretation of the risk model is the following: the standard claims which occur according to the immigrant-points trigger claims according to the branching structure. Typically, the fertility rate of an immigrant is taken to be monotonic decreasing, meaning that the claim number process has a self -exciting structure in which recent events affect the intensity of claim occurrences more than distant ones.

The final result obtained by Stabile and Torrisi in [35] is analogous to Cramér–Lundberg result. The probability of ruin is also decreasing exponentially as the initial surplus value goes to infinity. The work leading to this result is however quite demanding and technical.

6.4 Fluctuations in Social Systems

The unpredictability and variability of social systems has achieved new levels of prominence, and possibly new extremes, with the rise of the Internet. Let us consider that for instance on the 26th of February 2015, a badly lit photograph of a $77 off-the-rack dress of indeterminate color created a battle on the Internet that garnered 16 million views within 6 hours on Buzzfeed alone, with more than 670,000 people viewing the post simultaneously at one point [33].

Another interesting phenomenon is that religious militant organization Islamic State of Iraq and Syria, ISIS, with quite some success exploits viral meme propagation to recruit volunteers and attract funding for its military campaigns by peer recommendation on Twitter [3].

Understanding how, and when, and why this kind of viral propagation takes place is crucial to understanding the function of modern society. Why did that particular dress photograph have such an impact? As we see from ISIS, understanding the dynamics of these systems is implicated in global life-and-death struggles and violent political upheaval. Learning to understand the dynamics of these systems is economically and

Figure 6.4: The dress that broke the Internet 16 million views in 6 hours. What color is it?

politically important. And, thanks to the quantification of communication on the Internet, potentially plausible.

In [21] Daniel MacKinlay explored the use of models of self-exciting systems for modeling the viral activity. He assumed that viral activity may be partly exogenous, triggered by influences from the outside world, and partly endogenous, triggered by their own past activity. In [21] we read: this stylized description is the kind of dynamic we observe in, for example, financial markets, where (exogenous) news about a certain company might trigger movement in the price of its stock, but also movement in the price of a companys stock could itself trigger further movements as traders attempt

to surf the tide. However; in social systems, the mysterious popularity of the photograph of a dress viewed 16 million times in a single day is a paradigmatic example of endogenous triggering; there is no plausible news content attached to it.

MacKinlay [21] worked on Youtube data, which seemed promising but end up being troubling, as far as revealing the secrets of endogenously triggered dynamics. Ultimately he was not able to recommend tenable estimates for the parameters for the model. As he explains, the data was sparsely observed, and the ad hoc interpolation scheme used to approximate the missing information destroyed some times of timing information, removing the ability to estimate some kernel parameters. Secondly, inhomogeneity in the data lead to extremely poor model identification for the estimator, and the distribution of the estimates so compiled is not a credible predictor of "true" parameters.

There remains much work to be done in this very interesting and fast developing area. It is definitely worth further studies.

6.5 Modeling Trades-through in a Limit Order Book

In [24] and in [23] Ioane Muni Toke presents how Hawkes process can be applied in finance. Let us explain some finance jargon [16].

An order book is an electronic list of buy and sell orders for a specific security or financial instrument, organized by price level. The order book lists the number of shares being bid or offered at each price point, or market depth. It also identifies the market participants behind the buy and sell orders, although some choose to remain anonymous. The order book is dynamic and constantly updated in real time throughout the day. Exchanges such as Nasdaq refer to this order book as the *continuous book.* Orders that specify execution only at market open or market close are maintained separately. These are known as the *opening (order) book* and *closing (order) book,* respectively.

A market order is an order that an investor makes through a broker or brokerage service to buy or sell an investment immediately at the best available current price. A market order is the default option and is likely to be executed because it does not contain restrictions on the buy/sell price or the time-frame in which the order can be executed.

A limit order is an order to buy or sell a stock at a specific price or better. A buy

limit order can only be executed at the limit price or lower, and a sell limit order can only be executed at the limit price or higher. A limit order is not guaranteed to execute.

What are trades-through is explained in [24]: Let us assume that a trader wants to trade a large order. He does not want to reveal its intentions to the markets, so that the price will not "move against him". If he were to submit one large market order, he would eat the whole liquidity in the order book, trading at the first limit, then the second, then the third, and so on. When "climbing the ladder" this way, the last shares would be bought (resp. sold) at a price much higher (resp. lower) than the first ones. This trader will thus split its large order in several smaller orders that he will submit one at a time, waiting between each submitted order for some limit orders to bring back liquidity in the order book. We say that the trader tries to minimize its market impact. In practice, this mechanism is widely used: traders constantly scan the limit order book and very often, if not always, restrict the size of their orders to the quantity available at the best limit. But sometimes speed of execution is more important than minimizing market impact. In this case, orders larger than the size of the first limit may be submitted: thus, trades-through are precisely the trades that stand outside the usual trading pattern, and as such are worth being thoroughly studied.

It was shown that the clustering properties of trades-through can be well modeled with such self-exciting processes, which could lead to future practical work on trades-through-based trading strategies and order book modeling.

6.6 Other Applications

The self-exciting process is also a possible epidemic model in large populations in so far as the occurrence of a number of cases increases the probability of further cases. Alan Hawkes himself suggested in [14] that the mutually exciting processes could provide models for epidemics in which different types of cases are considered (children, adults, animals) and for associated diseases such as shingles and chicken pox.

Dassios and Zhao [7] did some work in these area, by introducing a new point process, which they called the dynamic contagion process. This dynamic contagion process is a generalization of the Hawkes process and the Cox process with shot noise intensity, and it includes both self-excited and end externally excited jumps, which could be

used to model the dynamic contagion impact from endogenous and exogenous factors from underlying system. The process they introduced has the significant potential of being applicable to a variety of problems in insurance, economics, finance, such as managing portfolio of credit risk or pricing derivative. In [7] they only looked at possible implementation in credit risk. Other applications could be the object of further research work. Also one may look at other generalizations of Hawkes process.

Other applications might be for example, the computer and the human body, as a model of neuron firing. Another application might be to the emission of particles from a radiating body which is excited by impact of other particles.

It seems that the examples of possible applications of Hawkes process is unlimited.

References

[1] Adamopoulos, L., Hawkes, *Cluster Models for Earthquakes-regional Comparisons*, Bulletin of the International Statistical Institute, 45 (3), Proc. 39th Session of ISI 454C-461, 1973.

[2] Barbieri, R., Brown, E., Kass, R., Ventura, V., *The Time-Rescaling Theorem and Its Application to Neural Spike Train Data Analysis*, Neural Computation 14, 325–346, 2001.

[3] Berger, J., *How ISIS Games Twitter*, The Atlantic, June 16, 2014, url: http://www.theatlantic.com/international/archive/2014/06/isis-iraq-twitter-social-media-strategy/372856/ (visited on 06/09/2016).

[4] Brillinger, D., *Maximum Likelihood Analysis of Spike Trains of Interacting Nerve Cells*, Biological Cybernetics, 59(3): 189–200, 1988.

[5] Cramér, H., *Collected Works*, V. I & II, Springer, 1994, edited by Martin Loef.

[6] Dassios, A., Zhao, H., *Exact Simulation of Hawkes Process with Exponentially Decaying Intensity*, Electronic Communications in Probability 18, 2013.

[7] Dassios, A., Zhao, H., *A Dynamic Contagion Process*, Advances in Applied Probability 43, 814–846, 2011.

[8] Egesdal, M., Fathauer, C., Louie, K., Neuman, J., *Statistical and Stochastic Modeling of Gang Rivalries in Los Angeles*, SIAM Undergraduate Research Online, V. 3, 72–94, 2010.

[9] Feller, W., *An Introduction to Probability Theory and its Applications*, second ed., V. II, John Wiley & Sons, 1980.

[10] Gaier, J., *Asymptotic Ruin Probabilities and Optimal Investment for an Insurer*, Technischen Universitaet Wien, 2002.

[11] Gaier, J., Grandits, P., Schachermayer, W., *Asymptotic Ruin Probabilities and Optimal Investment*, Annals of Applied Probability V. 13, no. 3, 1054–1076,2003.

[12] Gerber, H., *Martingales in Risk Theory*, Mitteilungen Schweizerische Vereinigung der Versicherungsmathematike, no. 73, 205–216, 1973.

[13] Grandell, J., *Aspects of Risk Theory*, Springer-Verlag, 1991.

[14] Hawkes, A., *Spectra of Some Self–exciting and Mutually Exciting Point Processes*, Biometrika 58, no. 1, 83–90, 1971.

[15] Hawkes, A., Oakes, D., *A Cluster Process Representation of a Self-exciting Process*, Journal of Applied Probability 11, 493–503, 1974.

[16] Investopedia Dictionary, url: http://www.investopedia.com/terms/ (visited on 09/09/2015).

[17] Kuhnert, P., Venables, B., *An Introduction to R: Software for Statistical Modelling & Computing*, Technical report, CSIRO Mathematical and Information Sciences, 2005.

[18] Laub, P., Taimre, T., Pollett, P., *Hawkes Processes*, Technical Report, Discipline of Mathematics, The University of Queensland, 2015.

[19] Lewis, P., Shedler, G., *Simulation of Nonhomogeneous Hawkes Processes by Thinning*, Naval Research Logistics Quarterly 26(3), 403–413, 1979.

[20] Lundberg, F., *Approximerad Framstaellning av Sannolikhetsfunktionen*, Akademisk Afhandling, 1903.

[21] MacKinlay, D., *Estimating Self-excitation Effects for Social Media Using the Hawkes Process*, Departement Management, Technologie und Oekonomie, Eidgenoessische Technische Hochschule, Zurich, 2015.

[22] Mohler, G., Short, M., Bratingham, P., Schoenberg, F., Tita, G., *Self-exciting Point Process Modeling of Crime*, Journal of the American Statistical Association 106 (493), 2011.

[23] Muni Toke, I., *An Introduction to Hawkes Processes with Application to Finance*, BNP Paribas Chair Meeting, Paris, 2011.

[24] Muni Toke, I., Pomponio F., *Modeling Trades–Through in a Limit Order Book Using Hawkes Processes*, Economics: The Open-Assessment E-Journal, V. 6, 2012.

[25] Ogata, Y., *Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes*, Journal of the American Statistical Association, 9–27, 1988.

[26] Peszat, S., *Wstep do Teorii Procesow Stochastycznych*, manuscript, Krakow, 2010.

[27] Pouzat, C., *Intensity Estimation with STAR: short version*,UFR de mathematiques et informatique de l'universite Paris-Descartes, Paris, 2012.

[28] Pouzat, C., Chaffiol, A. *Automatic Spike Train Analysis and Report Generation. An Implementation with R, R2HTML and STAR*, Journal of Neuroscience Methods, 181:119–144, 2009.

[29] Pouzat, C., Chaffiol, A. *On Goodness of Fit Tests For Models of Neuronal Spike Trains Considered as Counting Processes*, Laboratoire de Physiologie Cérébrale, C-NRS UMR 8118, 2008.

[30] Reynaud-Bouret, P., Rivoirard, V., Tuleau-Malot, C., *Inference of Functional Connectivity in Neurosciences via Hawkes Processes*, 1st IEEE Global Conference on Signal and Information Processing, Austin, Texas 2013.

[31] Reynaud-Bouret, P., Schbath, S., *Adaptive Estimation for Hawkes Processes; Application to Genome Analysis*, Annals of Statististics, 38(5), 2781–2822, 2010.

[32] Ross, S., *Introduction to Probability Models*, 11th Edition, Academic Press, 2014.

[33] Shaffer, L., *The Dress that Broke the Internet 16 Million Views in 6 Hours*, In: (Feb. 27, 2015). url: http://www.cnbc.com/id/102461771 (visited on 06/09/2016).

[34] Snyder, D., Miller, M., *Random Point Processes in Time and Space*, 2nd Edition, New York: Springer Verlag, 1991.

[35] Stabile, G., Torrisi, G., *Risk Processes with Non-stationary Hawkes Claims Arrivals*, Methodology and Computing in Applied Probability, 2010.