Nonlinear Hawkes Processes

by

Lingjiong Zhu

A dissertation submitted in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Department of Mathematics

New York University

May 2013

_____

Professor S. R. S. Varadhan

Dedication

To the memory of my grandpa

Zhixuan Zhu (1923-2001)

Acknowledgements

It is difficult to overstate my gratitude to my adviser Professor Varadhan. Working with Professor Varadhan has been an absolutely amazing experience for me. I thank him for always keeping his door open and patiently answering my questions. I thank him for his superb guidance, understanding, and generosity. I thank him for suggesting the topic for my thesis, which would not be possible without his deep wisdom and sharing of many new ideas. He has been everything that one can reasonably ask for in an advisor and more, and I am truly grateful to him.

I want to thank the Courant community for guiding me through this process and for putting up with me in general. Tamar Arnon does her job exceptionally well and her efforts are much appreciated. I want to thank the faculty for many well taught and interesting classes. I am indebted to Gérard Ben Arous, Sourav Chatterjee and Raghu Varadhan for writing me recommendations for my first academic job. I also want to thank Peter Carr for his interest in my thesis.

I remember a joke told by Jalal Shatah that the most important thing as an undergraduate student is to go to a top graduate program. But once you are already at a graduate school, the most important thing is to get out of it! This would not be possible without the final step, i.e. thesis defense! I am grateful to have Henry McKean, Chuck Newman and Raghu Varadhan as the three readers and Gérard Ben Arous and Lai-Sang Young as the two non-readers on my thesis committee.

Most importantly, I want to thank my fellow colleagues for all the fun memories that I take with me from Courant. New York City, without good friends, can be the most populated lonely place in the world, but thankfully the constant friend-

Abstract

The Hawkes process is a simple point process that has long memory, clustering effect, self-exciting property and is in general non-Markovian. The future evolution of a self-exciting point process is influenced by the timing of the past events. There are applications in finance, neuroscience, genome analysis, seismology, sociology, criminology and many other fields. We first survey the known results about the theory and applications of both linear and nonlinear Hawkes processes. Then, we obtain the central limit theorem and process-level, i.e. level-3 large deviations for nonlinear Hawkes processes. The level-1 large deviation principle holds as a result of the contraction principle. We also provide an alternative variational formula for the rate function of the level-1 large deviations in the Markovian case. Next, we drop the usual assumptions on the nonlinear Hawkes process and categorize it into different regimes, i.e. sublinear, sub-critical, critical, super-critical and explosive regimes. We show the different time asymptotics in different regimes and obtain other properties as well. Finally, we study the limit theorems of linear Hawkes processes with random marks.

Contents

List of Figures

Introduction

This thesis is about the nonlinear Hawkes process, a simple point processes, that has long memory, the clustering effect, the self-exciting property and is in general non-Markovian. The future evolution of a self-exciting point process is influenced by the timing of the past events. There are applications in finance, neuroscience, genome analysis, sociology, criminology, seismology, and many other fields.

Chapter 1 includes the introduction of the model and the survey of the results already known in the literature about Hawkes processes. That includes the stability results, limit theorems, power spectra of linear Hawkes processes and stability results of nonlinear Hawkes processes.

Chapter 2 is about the functional central limit theorem of nonlinear Hawkes processes. A Strassen's invariance holds under the same assumptions. The work in Chapter 2 is based on Zhu [114].

Chapter 3 is dedicated to the process-level large deviations, i.e. level-3 large deviations, of the nonlinear Hawkes processes. The proofs consist of the proofs of the lower bound, the upper bound and the superexponential estimates. The level-1 large deviation principle is derived as a result of the contraction principle. This chapter is based on Zhu [113].

Chapter 4 is dedicated to the study of level-1 large deviation principle for nonlinear Hawkes processes when the exciting functions are exponential or sums of exponentials. It is based on the observation that when the exciting functions are exponential or sums of exponentials, the process is Markovian and a combination of Feynman-Kac formula for the upper bound of large deviations of Markov processes and tilting of the intensity function of Hawkes processes for the lower bound will

establish a level-1 large deviation principle with the rate function expressed in terms of some variational formula. This chapter is based on Zhu [112].

Chapter 5 is about the asymptotics for nonlinear Hawkes processes. In this chapter, we drop the usual assumptions on nonlinear Hawkes processes, and study the phase transitions in different regimes. We categorize nonlinear Hawkes processes into the following regimes: sublinear regime, sub-critical regime, critical regime, super-critical regime and explosive regime. Different time asymptotics and various properties are obtained in different regimes. This chapter is based on Zhu [117].

Chapter 6 is about the limit theorems for linear Hawkes processes with random marks. The Central limit theorem and the large deviation principle are derived. We end this chapter with a simple application to a risk model. This is based on the joint work with my colleague Dmytro Karabash, see [62].

During my time as a PhD student at Courant Institute, I have the joy to work on some other problems either by myself or with my colleagues. For example, I studied the large deviations of self-correcting point processes with Sanchayan Sen, see [100] and also did some work on biased random walks on Galton-Watson trees without leaves with Behzad Mehrdad and Sanchayan Sen, see [75]. But since they are not closely related to the topics of my thesis, I do not include them here.

Chapter 1

Hawkes Processes

1.1 Introduction

Hawkes process is a self-exciting simple point process first introduced by Hawkes [51]. The future evolution of a self-exciting point process is influenced by the timing of past events. The process is non-Markovian except for some very special cases. In other words, Hawkes process depends on the entire past history and has a long memory. Hawkes process has wide applications in neuroscience, see e.g. Johnson [59], Chornoboy et al. [25], Pernice et al. [93], Pernice et al. [94], Reynaud et al. [98]; seismology, see e.g. Hawkes and Adamopoulos [53], Ogata [87], Ogata [88], Ogata et al. [90]; genome analysis, see e.g. Gusto and Schbath [46], Reynaud-Bouret and Schbath [96]; psycology, see e.g. Halpin and De Boeck [48]; spread of infectious disease, see e.g. Meyer et al. [76]; finance, see e.g. Bauwens and Hautsch [7], Bowsher [13], Hewlett [56], Large [67], Cartea et al. [22], Chavez-Demoulin et al. [23], Errais et al. [36]. Embrechts et al. [35], Muni Toke and Pomponio [83], Bacry et al. [3], [4], [1]; and in many other fields.

Let N be a simple point process on $\mathbb{R}$ and $\mathcal{F}_t^{-\infty} := \sigma(N(C), C \in \mathcal{B}(\mathbb{R}), C \subset (-\infty, t])$ be an increasing family of σ-algebras. Any nonnegative $\mathcal{F}_t^{-\infty}$-progressively measurable process λ_t with

$$(1.1) \qquad \mathbb{E}\left[N(a,b)|\mathcal{F}_a^{-\infty}\right] = \mathbb{E}\left[\int_a^b \lambda_s ds \Big| \mathcal{F}_a^{-\infty}\right],$$

a.s. for all intervals $(a, b]$ is called the $\mathcal{F}_t^{-\infty}$-intensity of N. We use the notation $N_t := N(0, t]$ to denote the number of points in the interval $(0, t]$.

A nonlinear Hawkes process is a simple point process N admitting an $\mathcal{F}_t^{-\infty}$-intensity

$$(1.2) \qquad \lambda_t := \lambda\left(\int_{-\infty}^t h(t-s)N(ds)\right),$$

where $\lambda(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ is locally integrable and left continuous, $h(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$. We always assume that $\|h\|_{L^1} = \int_0^\infty h(t)dt < \infty$ unless otherwise specified. Here $\int_{-\infty}^t h(t-s)N(ds)$ stands for $\int_{(-\infty,t)} h(t-s)N(ds)$, which is important for $\mathcal{F}_t^{-\infty}$-predictability. The local integrability assumption of $\lambda(\cdot)$ is to avoid explosion and the left continuity assumption of $\lambda(\cdot)$ is to ensure that the process is $\mathcal{F}_t^{-\infty}$-predictable.

In the literature, $h(\cdot)$ and $\lambda(\cdot)$ are usually referred to as exciting function and rate function respectively.

A Hawkes process is said to be linear if $\lambda(\cdot)$ is linear and it is nonlinear otherwise. For a linear Hawkes process, we can assume that the intensity is

$$(1.3) \qquad \lambda_t := \nu + \int_{(-\infty,t)} h(t-s)N(ds).$$

4

In this thesis, unless otherwise specified, we assume the following.

- $\lambda(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous and non-decreasing.

- $h(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ is continuous and non-increasing.

- $N(-\infty, 0] = 0$, i.e. the Hawkes process has empty past history.

Throughout, we define Z_t as $Z_t := \int_0^t h(t-s)N(ds)$. Thus, $\lambda_t = \lambda(Z_t)$.

The first assumption says that the occurence of the past and present events have positive impact on the occurence of the future events. The second assumption says that as time evolves, the impact of the past events is decreasing. For most of the results in this paper, these two assumptions may not be necessary. We nevertheless make them to avoid some technical difficulties.

If one looks at (1.2), it is clear that if you witness some events occuring, λ_t increases since $\lambda(\cdot)$ is increasing and you would expect even more events occuring. This is called the self-exciting property. Because of this, you would expect to see some clustering effects.

Figure 1.1 shows the histograms of a Hawkes process and a usual Poisson process. A Poisson process is stationary with independent increments. On the contrary, the Hawkes process has dependent increments and has clustering effects. As a result, in the picture, the Poisson process is more or less flat whilst the Hawkes process has peaks when it gets "excited" and has valleys when it "cools down". Figure 1.2 shows the plot of the intensity λ_t of a Hawkes process. Unlike the usual Poisson process for which the intensity is a positive constant, the intensity of Hawkes process increases when you witness arrivals of points and it decays when there are no arrivals of points.

The self-exciting and clutstering properties of the Hawkes process make it ideal to characterize the correlations in some complex systems, including the default clustering effect in finance.

One generalization of classical linear Hawkes process is the so-called multivariate Hawkes process. We will define the multivariate Hawkes process and discuss some basic results in Section 1.6 of Chapter 1. The multivariate Hawkes process has been well studied in the literature and we would like to point out that if you have the result for the univariate Hawkes process, mathematically, it is not too difficult to generalize your result to multivariate Hawkes process.

Unlike the univariate Hawkes process, which only has the self-exciting property, the multivariate Hawkes process also has the mutually-exciting property. In the context of industry, consider that you have a large portfolio of companies, then the failure of one company can have impact on the performance of other companies. In other words, multivariate Hawkes process captures the cross-sectional clustering effect. That is why in most applications of Hawkes processes in finance, people usually consider multivariate Hawkes processes. We will review some basic results about multivariate linear Hawkes process in Chapter 1.

Another possible generalization to Hawkes process is the marked Hawkes process, i.e. Hawkes process with random marks. Just like univariate Hawkes process vesus multivariate Hawkes process, if you have the results in unmarked Hawkes process, usually it can be generalized to marked Hawkes process without much difficulty. For instance, the large deviations for linear Hawkes process is proved in Bordenave and Torrisi [11] and the large deviations for linear marked Hawkes process is then proved in Karabash and Zhu [62]. We will discuss the details of limit theorems of linear marked Hawkes process in Chapter 6.

Most of the literature on Hawkes processes studies only the linear case, which has an immigration-birth representation (see Hawkes and Oakes [54]). The stability, law of large numbers, central limit theorem, large deviations, Bartlett spectrum etc. have all been studied and understood very well. Almost all of the applications of Hawkes processes in the literature consider exclusively the linear case. Daley and Vere-Jones [27] and Liniger [71] provide nice surveys about the theory and applications of Hawkes processes.

One special case of the Hawkes process is when the exciting function $h(\cdot)$ is exponential. In this case, the Hawkes process is a continuous time Markov process. If $\lambda(\cdot)$ is linear, the process is a special case of affine jump-diffusion process and is analytically tractable. This special case was for example studied in Oakes [85] and Errais et al. [36].

Because of the lack of computational tractability and immigration-birth representation, nonlinear Hawkes process is much less studied. However, some efforts have already been made in this direction. For instance, see Brémaud and Massoulié [14] for stability results, and Bremaud et al. [15] for the rate of convergence to stationarity. Karabash [63] recently proved the stability results for a wider class of nonlinear Hawkes processes.

As to the limit theorems, Bacry et al. [2] proved the central limit theorem for linear Hawkes process and Bordenave and Torrisi [11] proved the large deviation principle for linear Hawkes process.

For nonlinear Hawkes process, there is no explicit expression for the variance in the central limit theorem or the rate function for the large deviation principle. The method is more abstract and much more involved. Zhu [114] proved a central limit theorem for ergodic nonlinear Hawkes processes. Zhu [112] studied the
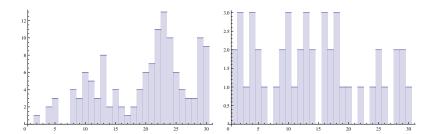
Figure 1.1: This is a comparison of a Hawkes process with a Poisson process. The figure on the left shows the histogram of a Hawkes process with $h(t) = \frac{1}{(t+1)^2}$ and $\lambda(z) = 1 + \frac{9}{10}z$ and the figure on the right the histogram of a Poisson process with constant intensity $\lambda \equiv \frac{3}{2}$. In the figure, each column represents the number of points that arrived in that unit time subinterval.

large deviations in the Markovian case, i.e. when $h(\cdot)$ is exponential or sum of exponentials. And Zhu [113] proved the large deviation principle for more general nonlinear Hawkes processes at the process-level, i.e. level-3.
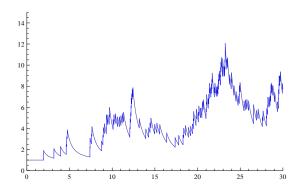
Figure 1.2: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = \frac{1}{(t+1)^2}$ and $\lambda(z) = 1 + 0.9z$.

8

1.2 Applications of Hawkes Processes

1.2.1 Applications in Finance

The applications of Hawkes processes in finance include market orders modelling, see e.g. Bauwens and Hautsch [7], Bowsher [13], Hewlett [56], Large [67] and Cartea et al. [22]; value-at-risk, see e.g. Chavez-Demoulin et al. [23]; and credit risk, see e.g. Errais et al. [36]. Embrechts et al. [35] applied Hawkes processes to model the financial data. Muni Toke and Pomponio [83] applied Hawkes processes to model the trade-through. Bacry et al. [3] used Hawkes processes to reproduce empirically microstructure noise and discussed the Epps effect and lead-lag. The self-exciting and clustering properties of Hawkes processes are especially appealing in financial applications.

Currently, most of the applications of Hawkes process in the finance literature are about market orders modelling, see e.g. Bauwens and Hautsch [7], Bowsher [13] and Large [67].

Recently, Chavez-Demoulin and McGill [24] used Hawkes processes to study the extremal returns in high-frequency trading. The Hawkes process captures the volatility clustering behavior of the intraday extremal returns. and provides a suitable estimation of high-quantile based risk measures (e.g. VaR, ES) for financial time series.

Filimonov and Sornette [39] used Hawkes process to model market events, with the aim of quantifying precisely endogeneity and exogeneity in market activity. By using Hawkes process, Filimonov and Sornette [39] analyzed E-mini S&P futures contract over the period 1998-2010 and discovered that the degree of self-reflexivity has increased steadily in the last decade, an effect they attribute to the increased

deployment of high-frequency and algorithmic trading. When they calibrated over much shorter time intervals (10 minutes), the Hawkes process analysis is found to detect precursors of the flash-crash that happened on May 6th, 2010. An early detection can benefit market regulators.

Very recently, Hardiman et al. [50] used (linear) Hawkes process to model the arrival of mid-price changes in the E-Mini S&P futures contract. Using several estimation methods, they found that the exciting function $h(\cdot)$ has a power-law decay and $\|h\|_{L^1}$ is close to 1. They pointed out that markets are and have always been close to criticality, challenging the studies of Filimonov and Sornette [39] which indicates that self-reflexivity (endogeneity) has increased in recent years as a result of increased automation of trading.

Egami et al. [33] studied the credit default swap (CDS) markets in both Japan and U.S. They made a dynamic analysis of the bid-ask spreads in both countries, which surged dramatically during the 2008-2009 financial crisis and they used the Hawkes process to predict the bid-ask spreads.

As pointed out in Errais et al. [36], "The collapse of Lehman Brothers brought the financial system to the brink of a breakdown. The dramatic repercussions point to the exisence of feedback phenomena that are channeled through the complex web of informational and contractual relationships in the economy... This and related episodes motivate the design of models of correlated default timing that incorporate the feedback phenomena that plague credit markets." According to Peng and Kou [92], "We need better models to incorporate the default clustering effect, i.e., one default event tends to trigger more default events both across time and cross-sectionally." The Hawkes process provides a model to characterize default events across time and if one uses a multivariate Hawkes process, that

would describe the cross-sectional clustering effect as well.

Hawkes processes have been proposed as models for the arrival of company defaults in a bond portfolio, starting with the papers Giesecke and Tomecek [42] and Giesecke et al. [41]. It is not hard to see that when the exciting function $h(\cdot)$ is exponential, the linear Hawkes processes are affine jump-diffusion processes, see for instance Errais et al. [36]. With the help of the theory of affine jump-diffusions, one can then analyze price processes related to certain credit derivatives analytically.

1.2.2 Applications in Sociology

The Hawkes process has also been applied to the study of social interactions. Crane and Sornette [26] analysed the viewing of YouTube videos as an example of a nonlinear social system. They identified peaks in the time series of viewing figures for around half a million videos and studied the subsequent decay of the peak to a background viewing level. In Crane and Sornette [26], the Hawkes process was proposed as a model of the video-watching dynamics, and a plausible link made to the social interactions that create strong correlations between the viewing actions of different people. Individual viewing is not random but influenced by various channels of communication about what to watch next. Mitchell and Cates [77] used computer simulation to test the the claims in Crane and Sornette [26] that robust identification is possible for classes of dynamic response following activity bursts. They also pointed out some limitations of the analysis based on the Hawkes process.

In sociology, Hawkes process has also been used by Blundell et al. [10] to study the reciprocating relationships. Reciprocity is a common social norm, where one person's actions towards another increases the probability of the same type of

action being returned, e.g., if Bob emails Alice, it increases the probability that Alice will email Bob in the near future. The mutually-exciting processes, e.g. multivariate Hawkes processes, are able to capture the causal nature of reciprocal interactions.

1.2.3 Applications in Seismology

Ogata [87] used a particular case of the Hawkes process to predict earthquakes and the Hawkes process appears to be superior to other models by residual analysis. The specific model used by Ogata [87] is now known as ETAS (Epidemic Type Aftershock-Sequences) model. The discussions of ETAS model can be found in Daley and Vere-Jones [27].

1.2.4 Applications in Genome Analysis

Gusto and Schbath [46] used the Hawkes process to model the occurences along the genome and studied how the occurences of a given process along a genome, genes or motifs for instance, may be influenced by the occurrences of a second process. More precisely, the aim is to detect avoided and/or favored distances between two motifs, for instance, suggesting possible interactions at a molecular level. The statistical method proposed by Gusto and Schbath [46] is useful for functional motif detection or to improve knowledge of some biological mechanisms.

Reynaud-Bouret and Schbath [96] provided a new method for the detection of either favored or avoided distances between genomic events along DNA sequences. These events are modeled by the Hawkes process. The biological problem is actually complex enough to need a non-asymptotic penalized model selection approach and Reynaud-Bouret and Schbath [96] provided a theoretical penalty that satisfies

an oracle inequality even for quite complex families of models.

1.2.5 Applications in Neuroscience

Chornoboy et al. [25] used the Hawkes process to detect and model the functional relationships between the neurons. The estimates are based on the maximum likelihood principle.

In most neural systems, neurons communicate via sequences of action potentials. Johnson [59] used various point processes, including Poisson process, renewal process and the Hawkes process and showed that neural discharges patterns convey time-varying information intermingled with the neuron's response characteristics. By applying information theory and estimation theory to point processes, Johnson [59] described the fundamental limits on how well information can be extracted from neural discharges.

More recently, Pernice et al. [93] and Pernice et al. [94] have used Hawkes process to model the spike train dynamics in the studies of neuronal networks. As pointed out in Pernice et al. [93], "Hawkes' point process theory allows the treatment of correlations on the level of spike trains as well as the understanding of the relation of complex connectivity patterns to the statistics of pairwise correlations." Reynaud et al. [98] proposed new non-parametric adaptive estimation methods and adapted other recent similar results to the setting of spike trains analysis in neuroscience. They tested homogeneuous Poisson process, inhomogeneous Poisson process and the Hawkes process. A complete analysis was performed on single unit activity recorded on a monkey during a sensory-motor task. Reynaud et al. [98] showed that the homogeneous Poisson process hypothesis is always rejected and that the inhomogeneous Poisson process hypothesis is rarely accepted.

The Hawkes model seems to fit most of the data.

The application of the Hawkes process in neuroscience has also been mentioned in Brémaud and Massoulié [14].

1.2.6 Applications in Criminology

Hawkes processes have also been used in criminology. Violence among gangs exhibits retaliatory behavior, i.e. given that an event has happened between two gangs, the likelihood that another event will happen shortly afterwards is increased. A problem like this can be modeled naturally by a self-exciting point process. Mohler et al. [78] and Egesdal et al. [34] have successfully modeled the pairwise gang violence as a Hawkes process. As pointed out in Hegemann et al. [55], in real-life situations, data is incomplete and law-enforcement agencies may not know which gang is involved. However, even when gang activity is highly stochastic, localized excitations in parts of the known dataset can help identify gangs responsible for unsolved crimes. The works before Hegemann et al. [55] incorporated the observed clustering in time of the data to identify gangs responsible for unsolved crimes by assuming that the parameters of the model are known, when in reality they have to be estimated from the data itself. Hegemann et al. [55] proposed an iterative method that simultaneously estimates the parameters in the underlying point process and assigns weights to the unknown events with a directly calculable score function.

Hawkes processses have also been used in the studies of terrorist activities. For example, Porter and White [95] used Hawkes process to examine the daily number of terrorist attacks in Indonesia from 1994 through 2007. Their model explains the self-exciting nature of the terrorist activities. It estimates the probability of future

attacks as a function of the times since the past attacks.

Lewis et al. [69] used Hawkes process to model the temporal dynamics of violence and civilian deaths in Iraq.

1.3 Related Models

There are other generalizations or variations of the Hawkes processes in the literature. For example, Bormetti et al. [12] introduced a one factor model where both the factor and the idiosyncratic jump components are described by a Hawkes process. Their model is a better candidate than classical Poisson or Hawkes models to describe the dynamics of jumps in a multi-asset framework. Another example is a multivariate Hawkes process with constraints on its conditional density introduced by Zheng et al. [111]. Their study is mainly motivated by the stochastic modelling of a limit order book for high frequency financial data analysis. Dassios and Zhao [28] proposed a dynamic contagion process. It is basically a combination of a marked Hawkes process with exponential exciting function and an external shot noise process. Their model is Markovian. They also applied their model to insurance, see e.g. Dassios and Zhao [29].

In [116], Zhu incorporated Hawkes jumps into the classical Cox-Ingersoll-Ross model and obtained limit theorems and various other properties.

In seismology, Wang et al. [104] proposed a new model, i.e. the Markov-modulated Hawkes process with stepwise decay (MMHPSD), to investigate the variation in seismicity rate during a series of earthquakes sequence including multiple main shocks. The MMHPSD is a self-exciting process which switches among different states, in each of which the process has distinguishable background seis-

micity and decay rates. Stress release models are often used in seismology. In Brémaud and Foss [17], they created a new earthquake model combining the classical stress release model for primary shocks with the Hawkes model for aftershocks and studied the ergodicity of this new model.

In addition to the classical Hawkes process, one can also study the spatial Hawkes process, see e.g. Møller and Torrisi [81], Møller and Torrisi [82], Bordenave and Torrisi [11]. In addition, the space-time Hawkes process has been used, see e.g. Musmeci and Vere-Jones [84] and Ogata [88].

1.4 Linear Hawkes Processes

In this section, let us review some known results about linear Hawkes process. Unlike the nonlinear Hawkes process, the linear Hawkes process has been very well studied in the literature. Hawkes and Oakes [54] introduced an immigration-birth representation of the linear Hawkes process, which can be viewed as a special case of the Poisson cluster process. The stability results of the linear Hawkes process, i.e. existence and uniqueness of a stationary linear Hawkes process have been summarised in Chapter 12 of Daley and Vere-Jones [27]. The rate of convergence to equilibrium has been stuided by Brémaud et al. [15]. The second-order analysis, i.e. the Bartlett spectrum etc. have been studied in Hawkes [51] and Hawkes [52]. Reynaud-Bouret and Roy [97] considered the linear Hawkes process as a special case of Poisson cluster process and studied the non-asymptotic tail estimates of the extinction time, the length of a cluster, and the number of points in an interval. Reynaud-Bouret and Roy [97] also obtained some so-called non-asymptotic ergodic theorems. The limit theorems have also been studied for linear Hawkes

process. The central limit theorem was considered in Bacry et al. [2], the large deviation principle was obtained in Bordenave and Torrisi [11], and very recently the moderate deviation principle was proved in Zhu [115]. The simulations and calibrations of linear Hawkes process have been studied in Ogata [89], Møller and Rasmussen [80], [79], Vere-Jones [110], Ozaki [91] and many others.

1.4.1 Immigration-Birth Representation

Consider the linear Hawkes process N with empty history, i.e. $N(-\infty, 0] = 0$ and intensity

$$(1.4) \qquad \lambda_t = \nu + \int_0^t h(t-u) N(du), \quad \nu > 0,$$

where $\int_0^\infty h(t) dt < 1$. It is well known that it has the following immigration-birth representation; see for example Hawkes and Oakes [54]. The immigrant arrives according to a homogeneous Poisson process with constant rate ν. Each immigrant reproduces children and the number of children has a Poisson distribution with parameter $\|h\|_{L^1}$. Conditional on the number of the children of an immigrant, the time that a child was born has probability density function $\frac{h(t)}{\|h\|_{L^1}}$. Each child produces children according to the same laws, independent of other children. All the immigrants produce children independently. Now, $N(0, t]$ is the same as the total number of immigrants and children in the time interval $(0, t]$.

1.4.2 Stability Results

Consider the linear Hawkes process N with empty history, i.e. $N(-\infty, 0] = 0$ and intensity

$$(1.5) \qquad \lambda_t = \nu + \int_0^t h(t - u)N(du),$$

where $\int_0^\infty h(t)dt < 1$. We review here the known results of existence and uniqueness of a stationary version of the process. We follow the arguments of Chapter 12 of Daley and Vere-Jones [27].

The existence of a stationary version of the process can be seen from the immigration-birth representation of the linear Hawkes process. To show uniqueness, let us do the following. Let $N^\dagger$ be a stationary version with intensity

$$(1.6) \qquad \lambda_t^\dagger = \nu + \int_{-\infty}^t h(t - u)N^\dagger(du),$$

and mean intensity $\mu := \mathbb{E}[\lambda_t^\dagger] = \frac{\nu}{1 - \|h\|_{L^1}}$. For both N and $N^\dagger$, we consider the shifted versions $\theta_s N$ and $\theta_s N^\dagger$ that bring the origin back to zero. $\theta_s N^\dagger$ can be split into two components, the one with the same structure as $\theta_s N$, being generated from the clusters initiated by immigrants arriving after time $-s$ and the component $N_{-s}^\dagger$ that counts the children of the immigrants that arrived before time $-s$. On $\mathbb{R}^+$, the contribution from the latter form a Poisson process with intensity

$$(1.7) \qquad \lambda_{-s}^\dagger(t) = \int_{-\infty}^{-s} h(t - u)N_{-s}^\dagger(du).$$

18

For any $T < \infty$,

$$(1.8) \qquad \mathbb{P}(N_{-s}^\dagger(0,T) > 0) = \mathbb{E}\left[1 - e^{-\int_0^T \lambda_{-s}^\dagger(t)dt}\right]$$

$$\leq \mathbb{E}\left[\int_0^T \lambda_{-s}^\dagger(t)dt\right] \leq \mu T \int_s^\infty h(u)du \to 0,$$

as $s \to \infty$. Let $\mathcal{P}$ and $\mathcal{P}^\dagger$ represent the probability measures corresponding to N and $N^\dagger$. For any $T > 0$, we have

$$(1.9) \qquad \|\theta_{-s}\mathcal{P} - \mathcal{P}^\dagger\|_{[0,T]} \leq \mathbb{P}(N_{-s}^\dagger(0,T) > 0) \to 0,$$

as $s \to \infty$, where $\|\cdot\|$ denotes the variation norm. This implies the weak convergence and thus the weak asymptotic stationarity of N.

Under a stronger assumption $\int_0^\infty th(t)dt < \infty$, i.e. the mean time to the appearance of a child is finite. Since the mean number of offspring is also finite (because $\|h\|_{L^1} < 1$), the random time T from the appearance of an ancestor to the last of its descendants has finite mean, i.e. $\mathbb{E}[T] < \infty$. Thus, we have

$$(1.10) \qquad \mathbb{P}(N_{-s}^\dagger[0,\infty) > 0) = 1 - e^{-\nu \int_s^\infty \mathbb{P}(T>u)du} \to 0,$$

as $s \to \infty$ and $\|\theta_{-s}\mathcal{P} - \mathcal{P}^\dagger\|_{[0,\infty]} \to 0$ as $s \to \infty$, which implies that the process starting from empty history is strongly asymptotically stationary.

Brémaud et al. [15] studied the rate of convergence to the equilibrium in a more general setting, i.e. Hawkes process with random marks. Here, we only consider the unmarked case. Assume $N(-\infty, 0] = 0$ and let $N^\dagger$ denote the unique stationary Hawkes process. The convergence in variation is seen via coupling, namely, N and $N^\dagger$ are constructed on the same space and there exists a finite random time T such

that

(1.11)
$$\mathbb{P}(N(t,\infty) = N^\dagger(t,\infty) \text{ for all } t \geq T) = 1.$$

In the exponential case, there exists some $\beta > 0$ such that $\int_0^\infty e^{\beta t} h(t) dt = 1$.
Let us define

(1.12)
$$H(t) := \frac{\nu}{1 - \|h\|_{L^1}} \int_t^\infty h(s) ds.$$

If $e^{\beta t} H(t)$ is directly Riemann integrable on $\mathbb{R}^+$, then for any

(1.13)
$$K > \frac{\int_0^\infty e^{\beta t} H(t) dt}{\beta \int_0^\infty t e^{\beta t} h(t) dt},$$

there exists $t_0(K)$ such that $\mathbb{P}(T > t) \leq K e^{-\beta t}$ for any $t \geq t_0(K)$.

In the subexponential case, the distribution function G with density $g(t) = \frac{h(t)}{\|h\|_{L^1}}$ is subexponential, in the sense that,

(1.14)
$$\lim_{t \to \infty} \frac{1 - G^{*n}(t)}{1 - G(t)} = n, \quad \text{for any } n \in \mathbb{N}.$$

Further assume that $\int_0^\infty t h(t) dt < \infty$. Then, for any

(1.15)
$$K > \frac{\nu \|h\|_{L^1}}{(1 - \|h\|_{L^1})^2},$$

there exists some $t_0(K)$ such that for any $t \geq t_0(K)$, we have

(1.16)
$$\mathbb{P}(T > t) \leq K \int_t^\infty \overline{G}(u) du,$$

20

where $\overline{G} = 1 - G$.

1.4.3 Bartlett Spectrum for Linear Hawkes Processes

The methods of analysis for point processes by spectrum were introduced by Bartlett [5] and [6]. We refer to Chapter 8 of Daley and Vere-Jones [27] for a detailed discussion.

Let N be a second-order stationary point process on $\mathbb{R}$. (For the definition of second-order stationary point process, we refer to Daley and Vere-Jones [27].) Define the set $\mathcal{S}$ as the space of functions of rapid decay, i.e. $\phi \in \mathcal{S}$ if

$$(1.17) \qquad \left| \frac{d^k \phi(x)}{dx^k} \right| \leq \frac{C(k,r)}{(1 + |x|)^r},$$

for some constants $C(k,r) < \infty$ and all positive integers r and k.

For bounded measurable ϕ with bounded support and also $\phi \in \mathcal{S}$, there exists a measure Γ on $\mathcal{B}$ such that

$$(1.18) \qquad \mathrm{Var}\left(\int_{\mathbb{R}} \phi(x)N(dx) \right) = \int_{\mathbb{R}} |\hat{\phi}(\omega)|\Gamma(d\omega),$$

where $\hat{\phi}(\omega) = \int_{\mathbb{R}} e^{i\omega u}\phi(u)du$ is the Fourier transform of ϕ. Γ is refered to as the Bartlett spectrum. We also have

$$(1.19) \qquad \mathrm{Cov}\left(\int_{\mathbb{R}} \phi(x)N(dx), \int_{\mathbb{R}} \psi(x)N(dx) \right) = \int_{\mathbb{R}} \hat{\phi}(\omega)\hat{\psi}(\omega)\Gamma(d\omega).$$

Hawkes [52] proved that for the linear stationary Hawkes process with

$$(1.20) \qquad \lambda_t = \nu + \int_{-\infty}^{t} h(t-s)N(ds),$$

$\nu > 0$ and $\|h\|_{L^1} < 1$, the Bartlett spectrum is given by

$$(1.21) \qquad \Gamma(d\omega) = \frac{\nu}{2\pi(1 - \|h\|_{L^1})|1 - \hat{h}(\omega)|^2} d\omega.$$

Moreover, if $\mu(\tau) := \mathbb{E}[dN(t + \tau)dN(t)]/(dt)^2 - \mu^2$ is the covariance density, where $\mu := \frac{\nu}{1 - \|h\|_{L^1}}$, then Hawkes [51] proved that $\mu(\tau) = \mu(-\tau)$, $\tau > 0$, satisfies the equation

$$(1.22) \qquad \mu(\tau) = \mu h(\tau) + \int_{-\infty}^{\tau} h(t - v)\mu(v)dv.$$

Since $\mu(\tau) = \mu(-\tau)$, we have

$$(1.23) \qquad \mu(\tau) = \mu h(\tau) + \int_0^{\infty} h(\tau + v)\mu(v)dv + \int_0^{\tau} h(\tau - v)\mu(v)dv, \quad \tau > 0.$$

In general, $\mu(\tau)$ may not have an analytical form. However, when $h(\cdot)$ is exponential, say $h(t) = \alpha e^{-\beta t}$, Hawkes [51] showed that

$$(1.24) \qquad \mu(\tau) = \frac{\nu \alpha \beta(2\beta - \alpha)}{2(\beta - \alpha)^2} e^{-(\beta - \alpha)\tau}, \quad \tau > 0.$$

The Bartlett spectrum analysis has later been generalized to marked linear Hawkes processes and some more general models. We refer to Brémaud and Massoulié [18] and Brémaud and Massoulié [19].

1.4.4   Limit Theorems for Linear Hawkes Processes

When $\lambda(\cdot)$ is linear, say $\lambda(z) = \nu + z$, for some $\nu > 0$ and $\|h\|_{L^1} < 1$, the Hawkes process has a very nice immigration-birth representation, see for example

Hawkes and Oakes [54]. For such a linear Hawkes process, the limit theorems are very well understood. Consider a stationary Hawkes process $N^\dagger$ with intensity

$$\lambda_t^\dagger = \nu + \int_{-\infty}^t h(t-s)N^\dagger(ds). \tag{1.25}$$

Taking expecatations on the both sides of the above equation and using stationarity, we get

$$\mu := \mathbb{E}[\lambda_t^\dagger] = \nu + \int_{-\infty}^t h(t-s)\mathbb{E}[\lambda_s^\dagger]ds = \nu + \mu\|h\|_{L^1}, \tag{1.26}$$

which implies that $\mu = \frac{\nu}{1-\|h\|_{L^1}}$. By ergodic theorem, we have

$$\frac{N_t}{t} \to \frac{\nu}{1-\|h\|_{L^1}}, \quad \text{as } t \to \infty \text{ a.s.} \tag{1.27}$$

Moreover, Bordenave and Torrisi [11] proved a large deviation principle for $(\frac{N_t}{t} \in \cdot)$.

Theorem 1 (Bordenave and Torrisi 2007). $(N_t/t \in \cdot)$ *satisfies a large deviation principle with the rate function*

$$I(x) = \begin{cases} x\log\left(\frac{x}{\nu+x\|h\|_{L^1}}\right) - x + x\|h\|_{L^1} + \nu & \text{if } x \in [0,\infty) \\ +\infty & \text{otherwise} \end{cases}. \tag{1.28}$$

Recently, Bacry et al. [2] proved a functional central limit theorem for linear multivariate Hawkes process under certain assumptions. That includes the linear Hawkes process as a special case and they proved that

Theorem 2 (Bacry et al. 2011)**.**

$$(1.29) \qquad \frac{N_{\cdot t} - \cdot \mu t}{\sqrt{t}} \to \sigma B(\cdot), \quad \text{as } t \to \infty,$$

where $B(\cdot)$ is a standard Brownian motion. The convergence is weak convergence on $D[0,1]$, the space of cádlág functions on $[0,1]$, equipped with Skorokhod topology. Here,

$$(1.30) \qquad \mu = \frac{\nu}{1 - \|h\|_{L^1}} \quad \text{and} \quad \sigma^2 = \frac{\nu}{(1 - \|h\|_{L^1})^3}.$$

Unlike the central limit theorem and the law of the iterated logarithm, there are not as many good crietria one can use to prove the moderate deviation principle for nonlinear Hawkes processes, which would fill in the gap between the central limit theorem and the large deviation principle. Nevertheless, due to the analytical tractability and birth-immigration representation of linear Hawkes process, Zhu [115] proved the moderate deviations for linear Hawkes processes.

Theorem 3. *Assume $\lambda(z) = \nu + z$, $\nu > 0$, $\|h\|_{L^1} < 1$ and $\sup_{t>0} t^{3/2} h(t) = C < \infty$. For any Borel set A and time sequence $a(t)$ such that $\sqrt{t} \ll a(t) \ll t$, we have the following moderate deviation principle.*

$$(1.31) \qquad -\inf_{x \in A^o} J(x) \leq \liminf_{t \to \infty} \frac{t}{a(t)^2} \log \mathbb{P}\left(\frac{N_t - \mu t}{a(t)} \in A\right)$$
$$\leq \limsup_{t \to \infty} \frac{t}{a(t)^2} \log \mathbb{P}\left(\frac{N_t - \mu t}{a(t)} \in A\right) \leq -\inf_{x \in \overline{A}} J(x),$$

where $J(x) = \frac{x^2 (1 - \|h\|_{L^1})^3}{2\nu}$.

The proof of Theorem 3 will be given in Section 1.4.5.

In a nutshell, linear Hawkes processes satisfy very nice limit theorems and the limits can be computed more or less explicitly.

1.4.5 Proof of Theorem 3

Since a Hawkes process has a long memory and is in general non-Markovian, there is no good criterion in the literature for moderate deviations that we can use directly. For example, Bacry et al. [2] used a central limit theorem for martingales to obtain a central limit theorem for linear Hawkes processes. But there is no criterion for moderate deviations for martingales that can fit into the context of Hawkes processes. Our strategy relies on the fact that for linear Hawkes processes there is a nice immigration-birth representation from which we can obtain a semi-explicit formula for the moment generating function of N_t in Lemma 1. A careful asymptotic analysis of this formula would lead to the proof of Theorem 3.

Proof of Theorem 3. Let us first prove that for any $\theta \in \mathbb{R}$,

$$(1.32) \qquad \lim_{t \to \infty} \frac{t}{a(t)^2} \log \mathbb{E}\left[e^{\frac{a(t)}{t}\theta(N_t - \mu t)} \right] = \frac{\nu\theta^2}{2(1 - \|h\|_{L^1})^3}.$$

By Lemma 1, for fixed $\theta \in \mathbb{R}$ and t sufficiently large, we have

$$(1.33) \qquad \mathbb{E}\left[e^{\frac{a(t)}{t}\theta N_t} \right] = e^{\nu \int_0^t G_t(s)ds},$$

where $G_t(s) = e^{\frac{a(t)}{t}\theta + \int_0^s h(u)G_t(s-u)du} - 1$, $0 \leq s \leq t$. Here, $G_t(s)$ is simply the $F(s) - 1$ in Lemma 1. Because $\frac{a(t)}{t}\theta$ depends on t, we write $G_t(s)$ instead of $G(s)$ to indicate its dependence on t. Clearly, $G_t(s)$ is increasing in s and $G_t(\infty)$ is the minimal solution to the equation $x_t = e^{\frac{a(t)}{t}\theta + \|h\|_{L^1}x_t} - 1$. (See the proof of Lemma

1 and the reference therein.) Since $\|h\|_{L^1} < 1$, it is easy to see that $x_t = O(a(t)/t)$. Since $x_t = O(a(t)/t)$, we have $G_t(s) = O(a(t)/t)$ uniformly in s. By Taylor's expansion,

$$(1.34) \quad G_t(s) = \frac{a(t)\theta}{t} + \int_0^s h(u)G_t(s-u)du$$
$$+ \frac{1}{2}\left(\frac{a(t)\theta}{t}\right)^2 + \frac{1}{2}\left(\int_0^s h(u)G_t(s-u)du\right)^2$$
$$+ \frac{a(t)\theta}{t}\int_0^s h(u)G_t(s-u)du + O\left((a(t)/t)^3\right).$$

Let $G_t(s) = \frac{a(t)\theta}{t}G_1(s) + \left(\frac{a(t)}{t}\right)^2 G_2(s) + \epsilon_t(s)$, where

$$(1.35) \qquad\qquad G_1(s) := 1 + \int_0^s h(u)G_1(s-u)du,$$

and

$$(1.36) \quad G_2(s) := \int_0^s h(u)G_2(s-u)du + \frac{\theta^2}{2} + \theta^2(G_1(s)-1) + \frac{\theta^2}{2}(G_1(s)-1)^2.$$

Substituting (1.35) and (1.36) back into (1.34) and using the fact $G_t(s) = O(a(t)/t)$ uniformly in s, we get $\epsilon_t(s) = O((a(t)/t)^3)$ uniformly in s. Moreover, we claim that

$$(1.37) \qquad\qquad \lim_{t\to\infty}\frac{1}{a(t)}\left[\theta\nu\int_0^t G_1(s)ds - \theta\mu t\right] = 0,$$

$$(1.38) \qquad\qquad \lim_{t\to\infty}\frac{1}{t}\int_0^t G_2(s)ds = \frac{\theta^2}{2(1-\|h\|_{L^1})^3}.$$

26

To prove (1.37), notice first that

$$(1.39) \quad \frac{1}{t} \int_0^t G_1(s)ds = 1 + \frac{1}{t} \int_0^t \int_0^s h(u)G_1(s-u)duds$$

$$= 1 + \frac{1}{t} \int_0^t h(u) \int_u^t G_1(s-u)dsdu$$

$$= 1 + \frac{1}{t} \int_0^t h(u) \int_0^{t-u} G_1(s)dsdu$$

$$= 1 + \frac{1}{t} \int_0^t h(u) \int_0^t G_1(s)dsdu - \frac{1}{t} \int_0^t h(u) \int_{t-u}^t G_1(s)dsdu.$$

Therefore,

$$(1.40) \qquad \frac{1}{t} \int_0^t G_1(s)ds = \frac{1 - \frac{1}{t} \int_0^t h(u) \int_{t-u}^t G_1(s)dsdu}{1 - \int_0^t h(u)du}.$$

Hence,

$$(1.41) \quad \frac{1}{a(t)} \left[\theta\nu \int_0^t G_1(s)ds - \theta\mu t \right]$$

$$= \frac{\theta\nu}{a(t)} \int_0^t \left(G_1(s) - \frac{1}{1 - \|h\|_{L^1}} \right) ds$$

$$= \frac{\theta\nu t}{a(t)} \left[\frac{1}{1 - \int_0^t h(u)du} - \frac{1}{1 - \int_0^\infty h(u)du} \right] - \frac{\theta\nu}{a(t)} \frac{\int_0^t h(u) \int_{t-u}^t G_1(s)dsdu}{1 - \int_0^t h(u)du}.$$

For the first term in (1.41), we have

$$(1.42) \qquad \left| \frac{\theta\nu t}{a(t)} \left[\frac{1}{1 - \int_0^t h(u)du} - \frac{1}{1 - \int_0^\infty h(u)du} \right] \right| \leq \frac{|\theta|\nu t}{a(t)} \frac{\int_t^\infty h(u)du}{(1 - \|h\|_{L^1})^2} \to 0,$$

as $t \to \infty$, since by our assumption, $\sup_{t>0} t^{3/2}h(t) \leq C < \infty$, which implies that $\frac{t}{a(t)} \int_t^\infty h(u)du \leq \frac{t}{a(t)} \int_t^\infty \frac{C}{u^{3/2}}du = \frac{2C\sqrt{t}}{a(t)} \to 0$ as $t \to \infty$.

27

For the second term in (1.41), we have

(1.43)
$$\limsup_{t\to\infty} \left| \frac{\theta\nu}{a(t)} \frac{\int_0^t h(u) \int_{t-u}^t G_1(s)dsdu}{1 - \int_0^t h(u)du} \right| \leq \lim_{t\to\infty} G_1(t) \limsup_{t\to\infty} \frac{|\theta|\nu}{a(t)} \frac{\int_0^t h(u)udu}{1 - \|h\|_{L^1}} = 0.$$

This is because (1.35) is a renewal equation and $\|h\|_{L^1} < 1$. By the application of the Tauberian theorem to the renewal equation, (see Chapters XIII and XIV of Feller [38]), $\lim_{t\to\infty} G_1(t) = \frac{1}{1-\|h\|_{L^1}}$. Moreover, our assumptions $\sup_{t>0} t^{3/2}h(t) \leq C < \infty$ and $\|h\|_{L^1} < \infty$ imply that

(1.44)
$$\frac{1}{a(t)} \int_0^t h(u)udu \leq \frac{1}{a(t)} \int_0^1 h(u)udu + \frac{1}{a(t)} \int_1^t \frac{C}{u^{1/2}}du \to 0,$$

as $t \to \infty$.

To prove (1.38), notice that $\lim_{t\to\infty} G_1(t) = \frac{1}{1-\|h\|_{L^1}}$ and again by the application of the Tauberian theorem to the renewal equation, (see Chapters XIII and XIV of Feller [38]), we have

(1.45)
$$\lim_{t\to\infty} \frac{1}{t} \int_0^t G_2(s)ds = \lim_{t\to\infty} G_2(t)$$
$$= \frac{\theta^2}{2} \frac{1 + 2\left(\frac{1}{1-\|h\|_{L^1}} - 1\right) + \left(\frac{1}{1-\|h\|_{L^1}} - 1\right)^2}{1 - \|h\|_{L^1}}$$
$$= \frac{\theta^2}{2(1 - \|h\|_{L^1})^3}.$$

Finally, from (1.33) and the definitions of $G_1(s)$, $G_2(s)$ and $\epsilon_t(s)$, we have

(1.46)
$$\frac{t}{a(t)^2} \log \mathbb{E}\left[e^{\frac{a(t)}{t}\theta(N_t - \mu t)}\right]$$

$$= \frac{t}{a(t)^2}\nu \int_0^t G_t(s)ds - \theta\mu\frac{t}{a(t)}$$

$$= \frac{1}{a(t)}\left[\nu\theta \int_0^t G_1(s)ds - \theta\mu t\right] + \frac{1}{t}\nu \int_0^t G_2(s)ds + \frac{t}{a(t)^2}\int_0^t \epsilon_t(s)ds.$$

Hence, by (1.37), (1.38) and the fact that $\epsilon_t(s) = O((a(t)/t)^3)$ uniformly in s, we conclude that, for any $\theta \in \mathbb{R}$,

(1.47)
$$\lim_{t\to\infty} \frac{t}{a(t)^2} \log \mathbb{E}\left[e^{\frac{a(t)}{t}\theta(N_t - \mu t)}\right] = \frac{\nu\theta^2}{2(1 - \|h\|_{L^1})^3}.$$

Applying the Gärtner-Ellis theorem (see for example [30]), we conclude that, for any Borel set A,

(1.48)
$$-\inf_{x\in A^o} J(x) \le \liminf_{t\to\infty} \frac{t}{a(t)^2} \log \mathbb{P}\left(\frac{N_t - \mu t}{a(t)} \in A\right)$$

$$\le \limsup_{t\to\infty} \frac{t}{a(t)^2} \log \mathbb{P}\left(\frac{N_t - \mu t}{a(t)} \in A\right) \le -\inf_{x\in\overline{A}} J(x),$$

where

(1.49)
$$J(x) = \sup_{\theta\in\mathbb{R}}\left\{\theta x - \frac{\nu\theta^2}{2(1 - \|h\|_{L^1})^3}\right\} = \frac{x^2(1 - \|h\|_{L^1})^3}{2\nu}.$$

□

Lemma 1. *For* $\theta \le \|h\|_{L^1} - 1 - \log\|h\|_{L^1}$,

(1.50)
$$\mathbb{E}[e^{\theta N_t}] = e^{\nu \int_0^t (F(s)-1)ds},$$

29

where $F(s) = e^{\theta + \int_0^s h(u)(F(s-u)-1)du}$ *for any* $0 \le s \le t$.

Proof. The Hawkes process has a very nice immigration-birth representation, see for example Hawkes and Oakes [54]. The immigrant arrives according to a homogeneous Poisson process with constant rate ν. Each immigrant produces a number of children, this being Poisson distributed with parameter $\|h\|_{L^1}$. Conditional on the number of the children of an immigrant, the time that a child is born has probability density function $\frac{h(t)}{\|h\|_{L^1}}$. Each child produces children according to the same laws independent of other children. All the immigrants produce children independently. Let $F(t) = \mathbb{E}[e^{\theta S(t)}]$, where $S(t)$ is the number of descendants an immigrant generates up to time t. Hence, we have

$$(1.51) \quad \mathbb{E}\left[e^{\theta N_t}\right] = \sum_{k=0}^{\infty} \frac{(\nu t)^k}{k!} e^{-\nu t} \frac{1}{t^k/k!} \int \cdots \int_{t_1 < t_2 < \cdots < t_k} F(t_1) \cdots F(t_k) dt_1 \cdots dt_k$$

$$= e^{\nu \int_0^t (F(s)-1)ds}.$$

By page 39 of Jagers [58], for all $\theta \in (-\infty, \|h\|_{L^1} - 1 - \log\|h\|_{L^1}]$, $\mathbb{E}[e^{\theta S(\infty)}]$ is the minimal positive solution of

$$(1.52) \qquad\qquad \mathbb{E}[e^{\theta S(\infty)}] = e^{\theta} \exp\left\{\mu(\mathbb{E}[e^{\theta S(\infty)}] - 1)\right\}.$$

Let K be the number of children of an immigrant and let $S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(K)}$ be the number of descendants of immigrant's kth child that were born before time t

(including the kth child if and only if it was born before time t). Then

$$(1.53) \qquad F(t) = \sum_{k=0}^{\infty} \mathbb{E}\left[e^{\theta S(t)} | K = k\right] \mathbb{P}(K = k)$$

$$= e^{\theta} \sum_{k=0}^{\infty} \mathbb{E}\left[e^{\theta S_t^{(1)}}\right]^k \mathbb{P}(K = k)$$

$$= e^{\theta} \sum_{k=0}^{\infty} \left(\int_0^t \frac{h(s)}{\|h\|_{L^1}} F(t-s)ds\right)^k e^{-\|h\|_{L^1}} \frac{\|h\|_{L^1}^k}{k!}$$

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

$\square$

1.4.6 Simulations and Calibrations

Assume the past of a Hawkes process is known up to present time zero, say the configuration of the history is ω^-. Let τ_1 be the first jump after time zero. Then, it is easy to see that

$$(1.54) \qquad\qquad \mathbb{P}(\tau_1 \geq t) = e^{-\int_0^t \lambda_s^{\omega^-} ds},$$

where $\lambda_s^{\omega^-} = \nu + \sum_{\tau \in \omega^-} h(s - \tau)$. This leads to a straight forward simulation method which is applicable for any simple point process. This algorithm and its theoretical foundation go back to a thinning procedure given Lewis and Shedler [70]. In the context of Hawkes processes, this simulation method was first used in Ogata [89]. It is sometimes called Ogata's modified thinning algorithm.

If we want to simulate the stationary version of the Hawkes process on a finite time interval, then the standard method for the simulation method described above does not work as the past of the process is not known and cannot be simulated, at

least not completely.

If one ignores the past of the process and simply starts to simulate the process at some given time, one speaks about an approximate simulation. In this case, one is actually simulating a transient version and not the stationary version of the process. But if one simulates for a long enough time interval, then the transient version converges to the stationary one. Such an approximate simulation method of Hawkes processes was discussed in Møller and Rasmussen [80]. A simulation method which directly simulates the stationary version without approximation is a so-called perfect simulation method. The idea is to incorporate somehow the effect of past observations without actually simulating the past of the process. For point processes, this type of simulation has first been described in Brix and Kendall [20]. In the context of Hawkes processes, the perfect simulation method was discussed in Møller and Rasmussen [79].

The calibrations, i.e. the estimation of the parameters of Hawkes processes, was first studied in Vere-Jones [110] and Ozaki [91], based on a maximum likelihood method for point processes introduced by Rubin [99]. The properties of the maximum likelihood estimator was analyzed in Ogata [86].

In Marsan and Lengline [73], an Expectation-Maximization (EM) algorithm, called "Model Independent Stochastic Declustering" (MISD), is introduced for the nonparametric estimation of self-exciting point processes with time-homogeneous background rate (For linear Hawkes process with intensity $\lambda_t = \nu_t + \sum_{\tau < t} h(t-\tau)$, ν_t is the background rate and $h(\cdot)$ is the exciting function).

The efficacy of the MISD algorithm was studied in Sornette and Utkin [101], where the authors found that the ability of MISD to recover key parameters such as $\|h\|_{L^1}$ depends on the values of the model parameters. In particular, they pointed

out that the accuracy of MISD improves as the timescale over which the exciting function $h(\cdot)$ decays shortens. In Lewis and Mohler [68], they introduced a Maximum Penalized Likelihood Estimation (MPLE) approach for the nonparametric estimation of Hawkes processes. The method is capable of estimating ν_t and $h(t)$ simultaneously, without prior knowledge of their form. Analogous to MPLE in the context of density estimation, the added regularity of the estimates allows for higher accuracy and/or lower sample sizes in comparison to MISD.

1.5 Nonlinear Hawkes Processes

Consider a simple point process with intensity

$$(1.55) \qquad \lambda_t = \lambda \left(\int_{-\infty}^t h(t-s) N(ds) \right),$$

where $\lambda(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ and $h(\cdot) : \mathbb{R}^+ \to [0, \infty)$. Brémaud and Massoulié [14] studied the existence and uniqueness of a stationary nonlinear Hawkes process that satisfies the dynamics (1.55) as well as its stability in distribution and in variation. They allow $h(\cdot)$ to take negative values as well. In this thesis, we always consider $h(\cdot)$ to be nonnegative.

The following result is about the existence of a stationary nonlinear Hawkes process satisfying the dynamics (1.55). We do not need $\lambda(\cdot)$ to be Lipschitz.

Theorem 4 (Brémaud and Massoulié [14]). *Let $\lambda(\cdot)$ be a nonnegative, nondecreasing and left-continuous function, satisfying $\lambda(z) \leq C + \alpha z$ for any $z \geq 0$, for some $C > 0$ and $\alpha \geq 0$ and let $h(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ be such that $\alpha \int_0^\infty h(t)dt < 1$. Then there exists a stationary point process N with dynamics (1.55).*

The following results concerns the uniqueness and stability in distribution and in variation of a nonlinear Hawkes process.

Theorem 5 (Brémaud and Massoulié [14]). *Let $\lambda(\cdot)$ be α-Lipschitz such that $\alpha\|h\|_{L^1} < 1$.*

(i) There exists a unique stationary distribution of N with finite average intensity $\mathbb{E}[N(O,1]]$ and with dynamics (1.55).

(ii) Let $\epsilon_a(t) := \int_{t-a}^{t} \int_{\mathbb{R}^-} h(s - u)N(du)ds$. The dynamics (1.55) are stable in distribution with respect to either the initial condition (1.56) or the condition (1.57) below,

$$\text{(1.56)} \qquad \sup_{t\geq 0} \epsilon_a(t) < \infty \ a.s. \ and \ \lim_{t\to\infty} \epsilon_a(t) = 0 \ a.e. \ for \ every \ a > 0,$$

$$\text{(1.57)} \qquad \sup_{t\geq 0} \mathbb{E}[\epsilon_a(t)] < \infty \ and \ \lim_{t\to\infty} \mathbb{E}[\epsilon_a(t)] = 0 \ for \ every \ a > 0.$$

(iii) The dynamics (1.55) are stable in variation with respect to the initial condition,

$$\text{(1.58)} \qquad \int_{\mathbb{R}^+} h(t)N[-t,0)dt = \int_{\mathbb{R}^+} \int_{-\infty}^{0} h(t-s)N(ds) < \infty, \ a.s.$$

if we assume further that $\int_0^\infty th(t)dt < \infty$.

Massoulié [74] extended the stability results to nonlinear Hawkes processes with random marks. He also considered the Markovian case and proved stability results without the Lipschitz condition for $\lambda(\cdot)$.

Very recently, Karabash [63] proved stability results for a much wider class of nonlinear Hawkes process, including the case when $\lambda(\cdot)$ is not Lipschitz.

Moreover, Brémaud et al. [15] considered the rate of extinction for nonlinear Hawkes process, that is the rate of convergence to the equilibrium when the stationary process is an empty process. Indeed, they considered a more general setting, i.e. Hawkes process with random marks. Let N be a nonlinear Hawkes process which is empty on $(-\infty, 0]$, i.e. $N(-\infty, 0] = 0$ which satisfies the dynamics

$$(1.59) \qquad \lambda_t := \nu(t) + \phi\left(\int_0^t h(t-s)N(ds)\right),$$

where $\nu : \mathbb{R}^+ \to \mathbb{R}^+$ is locally integrable, $\phi : \mathbb{R} \to [0,\infty)$, $\phi(0) = 0$, ϕ is 1-Lipschitz and $h : \mathbb{R}^+ \to \mathbb{R}$ is measurable and not necessarily nonnegative and $\int_0^\infty |h(t)|dt < 1$. The unique stationary process N^0 corresponding to the dynamics

$$(1.60) \qquad \phi\left(\int_0^t h(t-s)N^0(ds)\right),$$

is the empty process. Assume $\int_0^\infty \nu(t)dt < \infty$, $\int_0^\infty th(t)dt < \infty$ and $t \mapsto |h(t)|$ is locally bounded.

Then $\theta_t N$ converges in variation to the empty process. The convergence in variation takes place via coupling in the sense that there exists a finite random time T so that,

$$(1.61) \qquad \mathbb{P}(N(t,\infty) = 0 \text{ for any } t \geq T) = 1.$$

Depending on whether the tail of $|h(t)|$ is exponential or subexponentail, the following was obtained by Brémaud et al. [15].

In the exponential case, let $\beta > 0$ be such that $\int_0^\infty e^{\beta t}|h(t)|dt = 1$. Assume

$e^{\beta t}\nu(t)$ is directly Riemann integrable. Then, for any K with

$$(1.62) \qquad K > \frac{\int_0^\infty e^{\beta t}\nu(t)dt}{\beta \int_0^\infty te^{\beta t}|h(t)|dt},$$

there exists $t_0(K)$, for any $t \geq t_0(K)$,

$$(1.63) \qquad \mathbb{P}(T > t) \leq Ke^{-\beta t}.$$

In the subexponential case, assume that distribution functinon G with density $g(t) = \frac{|h(t)|}{\int_0^\infty |h(t)|dt}$ is subexponential, $\nu(\cdot)$ is bounded and that $B = \limsup_{t\to\infty} \frac{\nu(t)}{\overline{G}(t)} < \infty$, where $\overline{G} = 1 - G$. Then for any

$$(1.64) \qquad K > \frac{B}{1 - \int_0^\infty |h(t)|dt},$$

there exists $t_0(K)$ such that for any $t \geq t_0$,

$$(1.65) \qquad \mathbb{P}(T > t) \leq K \int_t^\infty \overline{G}(s)ds.$$

Kwieciński and Szekli [66] considered the nonlinear Hawkes process as a special case of self-exciting process. Let $\mathcal{N}(\mathbb{R}^+)$ be the space of point processes on $\mathbb{R}^+$, which can be regarded as an element of $\mathcal{D}(\mathbb{R}^+)$, the space of functions which are right-continuous with left limits, equipped with Skorohod topology. For any $\mu, \nu \in \mathcal{N}(\mathbb{R}^+)$, $\mu \prec_\mathcal{N} \nu$ if $\mu(B) \leq \nu(B)$ for any bounded set $B \in \mathcal{B}(\mathbb{R}^+)$. For any $\mu, \nu \in \mathcal{N}(\mathbb{R}^+)$, $\mu \prec_\mathcal{D} \nu$ if and only if $(\mu_t) \prec_\mathcal{D} (\nu_t)$ for the corresponding functions $\mu_t := \mu((0, t]), \nu_t := \nu((0, t]) \in \mathcal{D}(\mathbb{R}^+)$, i.e. $\mu_t \leq \nu_t$ for all $t > 0$.

Now, for a simple point process N with intensity $\lambda(t, N)$ and compensator

$\Lambda(t, N) := \int_0^t \lambda(s, N)ds$, we say that N is positively self-exciting w.r.t. $\prec_{\mathcal{N}}$ if for any $\mu, \nu \in \mathcal{N}(\mathbb{R}^+)$,

(1.66) $\mu \prec_{\mathcal{N}} \nu$ implies that for any $t > 0$, $\lambda(t, \mu) \leq \lambda(t, \nu)$,

and N is positively self-exciting w.r.t. $\prec_{\mathcal{D}}$ if for any $\mu, \nu \in \mathcal{N}(\mathbb{R}^+)$,

(1.67) $\mu \prec_{\mathcal{D}} \nu$ implies that for any $t > 0$, $\Lambda(t, \mu) \leq \Lambda(t, \nu)$.

Kwieciński and Szekli [66] pointed out that if $h(\cdot)$ is nonnegative and $\lambda(\cdot)$ nondecreasing, then N is positively self-exciting with respect to $\prec_{\mathcal{N}}$, and that if $h(\cdot)$ is nonnegative and nondecreasing and $\lambda(\cdot)$ is nondecreasing, then N is positively self-exciting with respect to $\prec_{\mathcal{D}}$.

Let $(\Omega, \mathcal{F})$ be a Polish space with a closed partical ordering $\prec$. A probability measure on $(\Omega, \mathcal{F})$ is associated $(\prec)$ if

(1.68) $$P(C_1 \cap C_2) \geq P(C_1)P(C_2),$$

for all increasing sets $C_1, C_2 \in \mathcal{F}$ (a set C is increasing if $x \in C$ and $x \prec y$ implies $y \in C$).

Kwieciński and Szekli [66] proved that if N is positively self-exciting point process w.r.t. $\prec_{\mathcal{N}}$ (resp. $\prec_{\mathcal{D}}$), then N is associated $(\prec_{\mathcal{N}})$ (resp. $(\prec_{\mathcal{D}})$). Therefore, it implies that for a nonlinear Hawkes process, if $h(\cdot)$ is nonnegative and $\lambda(\cdot)$ nondecreasing, then N is associated $(\prec_{\mathcal{N}})$ and if $h(\cdot)$ is nonnegative and nondecreasing and $\lambda(\cdot)$ is nondecreasing, then N is associated $(\prec_{\mathcal{D}})$.

Next, let us consider the limit theorems for nonlinear Hawkes process. When

$\lambda(\cdot)$ is nonlinear, the usual immigration-birth representation no longer works and you may have to use some abstract theory to obtain limit theorems. Some progress has already been made.

Brémaud and Massoulié [14]'s stability result implies that by the ergodic theorem,

$$\text{(1.69)} \qquad \frac{N_t}{t} \to \mu := \mathbb{E}[N[0,1]],$$

as $t \to \infty$, where $\mathbb{E}[N[0,1]]$ is the mean of $N[0,1]$ under the stationary and ergodic measure.

In this thesis, we will obtain a functional central limit theorem and a Strassen's invariance principle in Chapter 2 and a process-level, i.e. level-3 large deviation principle in Chapter 3 and thus a level-1 large deviation principle by contraction principle. We will also obtain an alternative expression for the rate function for level-1 large deviation principle of Markovian nonlinear Hawkes process as a variational formula in Chapter 4.

1.6 Multivariate Hawkes Processes

We say $N = (N_1, \ldots, N_d)$ is a multivariate Hawkes process if for any $1 \leq i \leq d$, N_i is a simple point process with intensity

$$\text{(1.70)} \qquad \lambda_{i,t} := \nu_i + \int_0^t \sum_{j=1}^d h_{ij}(t-s)dN_{j,s},$$

where $\nu_i \in \mathbb{R}^+$ and $h_{ij}(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$. Then, $\nu := (\nu_1, \ldots, \nu_d)$ is a vector and $\mathbf{h} := (h_{ij})_{1 \leq i,j \leq d}$ is a $d \times d$ matrix-valued function.

Let us assume that for any i, j, $\int_0^\infty h_{ij}(t)dt < \infty$ and that the spectral radius $\rho(\mathbf{K})$ of the matrix $\mathbf{K} = \int_0^\infty \mathbf{h}(t)dt$ satisfies $\rho(\mathbf{K}) < 1$. Then, Bacry et al. [2] proved a law of large numbers, i.e.

$$(1.71) \qquad \sup_{u \in [0,1]} \|T^{-1}N_{Tu} - u(\mathbf{I} - \mathbf{K})^{-1}\nu\| \to 0,$$

as $T \to \infty$ almost surely and also in $L^2(\mathbb{P})$. If we assume further that for any $1 \leq i, j \leq d$,

$$(1.72) \qquad \int_0^\infty h_{ij}(t)t^{1/2}dt < \infty.$$

Then, Bacry et al. [2] proved the following central limit theorem:

$$(1.73) \qquad \sqrt{T}\left(\frac{1}{T}N_{Tu} - u(\mathbf{I} - \mathbf{K})^{-1}\nu\right), \quad u \in [0,1]$$

converges in law as $T \to \infty$ under the Skorohod topology to

$$(1.74) \qquad (\mathbf{I} - \mathbf{K})^{-1}\Sigma^{1/2}W_u, \quad u \in [0,1],$$

where Σ is the diagonal matrix with $\Sigma_{ii} = ((\mathbf{I} - \mathbf{K})^{-1}\nu)_i$, $1 \leq i \leq d$.

It is well known that under the assumption that $\rho(\mathbf{K}) < 1$, there exists a unique stationary version of the multivariate Hawkes process satisfying the dynamics (1.70). The rate of convergence to the stationary version of the multivariate Hawkes process was obtained in Torrisi [103]. The Bartlett spectrum of the multivariate Hawkes process was derived in Hawkes [52]. Some non-asymptotics estimates for multivariate Hawkes processes were obtained in Hansen et al. [49].

A nice survey on multivariate linear Hawkes processes can be found in Liniger [71].

Chapter 2

Central Limit Theorem for Nonlinear Hawkes Processes

2.1 Main Results

In this chapter, we obtain a functional central limit theorem for the nonlinear Hawkes process under Assumption 1. Under the same assumption, a Strassen's invariance principle also holds. Let us recall that N is a nonlinear Hawkes process with intensity

$$(2.1) \qquad \lambda_t := \lambda \left(\int_{(-\infty,t)} h(t-s)N(ds) \right).$$

Assumption 1. *We assume that*

- $h(\cdot) : [0, \infty) \to \mathbb{R}^+$ *is a decreasing function and $\int_0^\infty th(t)dt < \infty$.*

- $\lambda(\cdot)$ *is positive, increasing and α-Lipschitz (i.e. $|\lambda(x) - \lambda(y)| \leq \alpha|x - y|$ for any x, y) and $\alpha\|h\|_{L^1} < 1$.*

Brémaud and Massoulié [14] proved that if $\lambda(\cdot)$ is α-Lipschitz with $\alpha\|h\|_{L^1} < 1$, there exists a unique stationary and ergodic Hawkes process satisfying the dynamics (1.2). Hence, under our Assumption 1 (which is slightly stronger than [14]), there exists a unique stationary and ergodic Hawkes process satisfying the dynamics (1.2).

Let $\mathbb{P}$ and $\mathbb{E}$ denote the probability measure and expectation for a stationary, ergodic Hawkes process, and let $\mathbb{P}(\cdot|\mathcal{F}_0^{-\infty})$ and $\mathbb{E}(\cdot|\mathcal{F}_0^{-\infty})$ denote the conditional probability measure and expectation given the past history.

The following are the main results of this chapter.

Theorem 6. *Under Assumption 1, let N be the stationary and ergodic nonlinear Hawkes process with dynamics (1.2). We have*

$$(2.2) \qquad \frac{N_{\cdot t} - \cdot \mu t}{\sqrt{t}} \to \sigma B(\cdot), \quad as\ t \to \infty,$$

where $B(\cdot)$ is a standard Brownian motion and $0 < \sigma < \infty$, where

$$(2.3) \qquad \sigma^2 := \mathbb{E}[(N[0,1] - \mu)^2] + 2\sum_{j=1}^{\infty} \mathbb{E}[(N[0,1] - \mu)(N[j, j+1] - \mu)].$$

The convergence in (2.2) is weak convergence on $D[0,1]$, the space of cádlág functions on $[0,1]$, equipped with Skorokhod topology.

Remark 1. *By a standard central limit theorem for martingales, i.e. Theorem 9, it is easy to see that*

$$(2.4) \qquad \frac{N_{\cdot t} - \int_0^{\cdot t} \lambda_s ds}{\sqrt{t}} \to \sqrt{\mu}B(\cdot), \quad as\ t \to \infty,$$

where $\mu = \mathbb{E}[N[0,1]]$. In the linear case, say $\lambda(z) = \nu + z$, Bacry et al. [2]

proved that σ^2 in (2.3) satisfies $\sigma^2 = \frac{\nu}{(1-\|h\|_{L^1})^3} > \mu = \frac{\nu}{1-\|h\|_{L^1}}$. That is not surprising because $N_{\cdot t} - \cdot \mu t$ "should" have more fluctuations than $N_{\cdot t} - \int_0^{\cdot t} \lambda_s ds$. Therefore, we guess that for nonlinear $\lambda(\cdot)$, σ^2 defined in (2.3) should also satisfy $\sigma^2 > \mu = \mathbb{E}[N[0,1]]$. However, it might not be very easy to compute and say something about σ^2 in such a case.

In the classical case for a sequence of i.i.d. random variables X_i with mean 0 and variance 1, we have the central limit theorem $\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \to N(0,1)$ as $n \to \infty$, and we also have $\frac{\sum_{i=1}^n X_i}{\sqrt{n \log \log n}} \to 0$ in probability as $n \to \infty$, but the convergence does not hold a.s. The law of the iterated logarithm says that $\limsup_{n\to\infty} \frac{\sum_{i=1}^n X_i}{\sqrt{n \log \log n}} = \sqrt{2}$ a.s. A functional version of the law of the iterated logarithm is called Strassen's invariance principle.

It turns out that we also have a Strassen's invariance principle for nonlinear Hawkes processes under Assumption 1.

Theorem 7. *Under Assumption 1, let N be the stationary and ergodic nonlinear Hawkes process with dynamics (1.2). Let $X_n := N[n-1,n] - \mu$, $S_n := \sum_{i=1}^n X_i$, $s_n^2 := \mathbb{E}[S_n^2]$, $g(t) = \sup\{n : s_n^2 \leq t\}$, and for $t \in [0,1]$, let $\eta_n(t)$ be the usual linear interpolation, i.e.*

(2.5)
$$\eta_n(t) = \frac{S_k + (s_n^2 t - s_k^2)(s_{k+1}^2 - s_k^2)^{-1} X_{k+1}}{\sqrt{2 s_n^2 \log \log s_n^2}}, \quad s_k^2 \leq s_n^2 t \leq s_{k+1}^2, k = 0, 1, \ldots, n-1.$$

Then, $g(e) < \infty$, $\{\eta_n, n > g(e)\}$ is relatively compact in $C[0,1]$, the set of continuous functions on $[0,1]$ equipped with uniform topology, and the set of limit points is the set of absolutely continuous functions $f(\cdot)$ on $[0,1]$ such that $f(0) = 0$ and $\int_0^1 f'(t)^2 dt \leq 1$.

2.2 Proofs

This section is devoted to the proof of Theorem 6. We use a standard central limit theorem, i.e. Theorem 8. In our proof, we need the fact that $\mathbb{E}[N[0,1]^2] < \infty$, which is proved in Lemma 3. Lemma 3 is proved by proving a stronger result first, i.e. Lemma 2. We will also prove Lemma 4 to guarantee that $\sigma > 0$ so that the central limit theorem is not degenerate.

Let us first quote the two necessary central limit theorems from Billingsley [8]. In both Theorem 8 and Theorem 9, the filtrations are the natural ones, i.e. given a stochastic process $(X_n)_{n \in \mathbb{Z}}$, $\mathcal{F}_b^a := \sigma(X_n, a \leq n \leq b)$, for $-\infty \leq a \leq b \leq \infty$.

Theorem 8 (Page 197 [8]). *Suppose X_n, $n \in \mathbb{Z}$, is an ergodic stationary sequence such that $\mathbb{E}[X_n] = 0$ and*

$$(2.6) \qquad \sum_{n \geq 1} \|\mathbb{E}[X_0|\mathcal{F}_{-n}^{-\infty}]\|_2 < \infty,$$

where $\|Y\|_2 = (\mathbb{E}[Y^2])^{1/2}$. Let $S_n = X_1 + \cdots + X_n$. Then $S_{[n\cdot]}/\sqrt{n} \to \sigma B(\cdot)$ weakly, where the weak convergence is on $D[0,1]$ equipped with the Skorohod topology and $\sigma^2 = \mathbb{E}[X_0^2] + 2\sum_{n=1}^{\infty} \mathbb{E}[X_0 X_n]$. The series converges absolutely.

Theorem 9 (Page 196 [8]). *Suppose X_n, $n \in \mathbb{Z}$, is an erogdic stationary sequence of square integrable martingale differences, i.e. $\sigma^2 = \mathbb{E}[X_n^2] < \infty$, and let $\mathbb{E}[X_n|\mathcal{F}_{n-1}^{-\infty}] = 0$. Let $S_n = X_1 + \cdots + X_n$. Then $S_{[n\cdot]}/\sqrt{n} \to \sigma B(\cdot)$ weakly, where the weak convergence is on $D[0,1]$ equipped with the Skorohod topology.*

Now, we are ready to prove our main result.

Proof of Theorem 6. Since in the stationary regime, $\mathbb{E}[N[n, n+1]] = \mathbb{E}[N[0,1]]$ for any $n \in \mathbb{Z}$ and let us denote $\mathbb{E}[N[0,1]] = \mu$. In order to apply Theorem 8, let us

first prove that

$$(2.7) \qquad \sum_{n=1}^{\infty} \left\{ \mathbb{E}\left[(\mathbb{E}[N(n,n+1] - \mu|\mathcal{F}_0^{-\infty}])^2 \right] \right\}^{1/2} < \infty.$$

Let $\mathbb{E}^{\omega_1^-}[N(n,n+1]]$ and $\mathbb{E}^{\omega_2^-}[N(n,n+1]]$ be two independent copies of $\mathbb{E}[N(n,n+1]|\mathcal{F}_0^{-\infty}]$. It is easy to check that

$$(2.8) \qquad \frac{1}{2}\mathbb{E}\left\{ \left[\mathbb{E}^{\omega_1^-}[N(n,n+1]] - \mathbb{E}^{\omega_2^-}[N(n,n+1]] \right]^2 \right\}$$

$$= \frac{1}{2}\mathbb{E}\left[\mathbb{E}^{\omega_1^-}[N(n,n+1]]^2 \right] + \frac{1}{2}\mathbb{E}\left[\mathbb{E}^{\omega_2^-}[N(n,n+1]]^2 \right]$$

$$\qquad - \mathbb{E}\left[\mathbb{E}^{\omega_1^-}[N(n,n+1]]\mathbb{E}^{\omega_2^-}[N(n,n+1]] \right]$$

$$= \mathbb{E}\left[\mathbb{E}[N(n,n+1]|\mathcal{F}_0^{-\infty}]^2 \right] - \mu^2$$

$$= \mathbb{E}\left[(\mathbb{E}[N(n,n+1] - \mu|\mathcal{F}_0^{-\infty}])^2 \right].$$

Therefore, we have

$$(2.9) \qquad \mathbb{E}\left[(\mathbb{E}[N(n,n+1] - \mu|\mathcal{F}_0^{-\infty}])^2 \right]$$

$$= \frac{1}{2}\mathbb{E}\left\{ \left[\mathbb{E}^{\omega_1^-}[N(n,n+1]] - \mathbb{E}^{\omega_2^-}[N(n,n+1]] \right]^2 \right\}$$

$$\leq \mathbb{E}\left\{ \left[\mathbb{E}^{\omega_1^-}[N(n,n+1]] - \mathbb{E}^{\varnothing}[N(n,n+1]] \right]^2 \right\}$$

$$\qquad + \mathbb{E}\left\{ \left[\mathbb{E}^{\omega_2^-}[N(n,n+1]] - \mathbb{E}^{\varnothing}[N(n,n+1]] \right]^2 \right\}$$

$$= 2\mathbb{E}\left\{ \left[\mathbb{E}^{\omega_1^-}[N(n,n+1]] - \mathbb{E}^{\varnothing}[N(n,n+1]] \right]^2 \right\},$$

where $\mathbb{E}^{\varnothing}[N(n,n+1]]$ denotes the expectation of the number of points in $(n,n+1]$ for the Hawkes process with the same dynamics (1.2) and empty history, i.e. $N(-\infty,0] = 0$.

Next, let us estimate $\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]]$. $\mathbb{E}^{\omega_1^-}[N(n, n+1]]$ is the expectation of the number of points in $(n, n+1]$ for the Hawkes process with intensity $\lambda_t = \lambda\left(\sum_{\tau:\tau\in\omega_1^-\cup\omega[0,t)} h(t-\tau)\right)$. It is well defined for a.e. ω_1^- under $\mathbb{P}$ because, under Assumption 1,

$$(2.10) \quad \mathbb{E}[\lambda_t] \leq \lambda(0) + \alpha\mathbb{E}\left[\int_{-\infty}^t h(t-s)N(ds)\right] = \lambda(0) + \alpha\|h\|_{L^1}\mathbb{E}[N[0,1]] < \infty,$$

which implies that $\lambda_t < \infty$ $\mathbb{P}$-a.s.

It is clear that $\mathbb{E}^{\omega_1^-}[N(n, n+1]] \geq \mathbb{E}^{\varnothing}[N(n, n+1]]$ almost surely, so we can use a coupling method to estimate the difference. We will follow the ideas in Brémaud and Massoulié [14] using the Poisson embedding method. Consider $(\Omega, \mathcal{F}, \mathcal{P})$, the canonical space of a point process on $\mathbb{R}^+ \times \mathbb{R}^+$ in which $\overline{N}$ is Poisson with intensity 1 under the probability measure $\mathcal{P}$. Then the Hawkes process N^0 with empty past history and intensity λ_t^0 satisfies the following.

$$(2.11) \quad \begin{cases} \lambda_t^0 = \lambda\left(\int_{(0,t)} h(t-s)N^0(ds)\right) & t \in \mathbb{R}^+, \\ N^0(C) = \int_C \overline{N}(dt \times [0, \lambda_t^0]) & C \in \mathcal{B}(\mathbb{R}^+). \end{cases}$$

For $n \geq 1$, let us define recursively λ_t^n, D_n and N^n as follows.

$$(2.12) \quad \begin{cases} \lambda_t^n = \lambda\left(\int_{(0,t)} h(t-s)N^{n-1}(ds) + \sum_{\tau\in\omega_1^-} h(t-\tau)\right) & t \in \mathbb{R}^+, \\ D_n(C) = \int_C \overline{N}(dt \times [\lambda_t^{n-1}, \lambda_t^n]) & C \in \mathcal{B}(\mathbb{R}^+), \\ N^n(C) = N^{n-1}(C) + D_n(C) & C \in \mathcal{B}(\mathbb{R}^+). \end{cases}$$

Following the arguments as in Brémaud and Massoulié [14], we know that each λ_t^n is an $\mathcal{F}_t^{\overline{N}}$-intensity of N^n, where $\mathcal{F}_t^{\overline{N}}$ is the σ-algebra generated by $\overline{N}$ up to time

46

t. By our Assumption 1, $\lambda(\cdot)$ is increasing, and it is clear that $\lambda^n(t)$ and $N^n(C)$ increase in n for all $t \in \mathbb{R}^+$ and $C \in \mathcal{B}(\mathbb{R}^+)$. Thus, D_n is well defined and also that as $n \to \infty$, the limiting processes λ_t and N exist. N counts the number of points of $\overline{N}$ below the curve $t \mapsto \lambda_t$ and admits λ_t as an $\mathcal{F}_t^{\overline{N}}$-intensity. By the monotonicity properties of λ_t^n and N^n, we have

$$(2.13) \qquad \lambda_t^n \le \lambda \left(\int_{(0,t)} h(t-s) N(ds) + \sum_{\tau \in \omega_1^-} h(t-\tau) \right),$$

$$(2.14) \qquad \lambda_t \ge \lambda \left(\int_{(0,t)} h(t-s) N^n(ds) + \sum_{\tau \in \omega_1^-} h(t-\tau) \right).$$

Letting $n \to \infty$ (it is valid since we assume that $\lambda(\cdot)$ is Lipschitz and thus continuous), we conclude that N, λ_t satisfies the dynamics (1.2). Therefore, with intensity λ_t, $N = N^0 + \sum_{i=1}^\infty D_i$ is the Hawkes process with past history ω_1^-.

We can then estimate the difference by noticing that

$$(2.15) \qquad \mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^\varnothing[N(n, n+1]] = \sum_{i=1}^\infty \mathbb{E}^{\mathcal{P}}[D_i(n, n+1]].$$

Here $\mathbb{E}^{\mathcal{P}}$ means the expectation with respect to $\mathcal{P}$, the probability measure on the canonical space that we defined earlier.

We have

$$(2.16) \quad \mathbb{E}^{\mathcal{P}}[D_1(n, n+1]]$$

$$= \mathbb{E}^{\mathcal{P}}\left[\int_n^{n+1} (\lambda^1(t) - \lambda^0(t))dt\right]$$

$$= \mathbb{E}^{\mathcal{P}}\left[\int_n^{n+1} \lambda\left(\sum_{\tau < t, \tau \in N^0 \cup \omega_1^-} h(t-\tau)\right) - \lambda\left(\sum_{\tau < t, \tau \in N^0 \cup \varnothing} h(t-\tau)\right)dt\right]$$

$$\leq \alpha \int_n^{n+1} \sum_{\tau \in \omega_1^-} h(t-\tau)dt,$$

where the first equality in (2.16) is due to the construction of D_1 in (2.12), the second equality in (2.16) is due to the definitions of λ^1 and λ^0 in (2.12) and finally the inequality in (2.16) is due to the fact that $\lambda(\cdot)$ is α-Lipschitz. Similarly,

$$(2.17) \qquad \mathbb{E}^{\mathcal{P}}[D_2(n, n+1]] \leq \mathbb{E}^{\omega_1^-}\left[\alpha \int_n^{n+1} \sum_{\tau \in D_1, \tau < t} h(t-\tau)dt\right]$$

$$\leq \sum_{\tau \in \omega_1^-} \alpha^2 \int_n^{n+1} \int_0^t h(t-s)h(s-\tau)ds dt.$$

Iteratively, we have, for any $k \in \mathbb{N}$,

$$\mathbb{E}^{\mathcal{P}}[D_k(n, n+1]] \leq \sum_{\tau \in \omega_1^-} \alpha^k \int_n^{n+1} \int_0^{t_k} \cdots \int_0^{t_2} h(t_k - t_{k-1})h(t_{k-1} - t_{k-2})$$

$$\cdots h(t_2 - t_1)h(t_1 - \tau)dt_1 \cdots dt_k =: \sum_{\tau \in \omega_1^-} K_k(n, \tau).$$

Now let $K(n, \tau) := \sum_{k=1}^{\infty} K_k(n, \tau)$. Then,

(2.18)
$$\mathbb{E}\left\{ \left[\mathbb{E}^{\omega_1^-}[N(n, n+1]] - \mathbb{E}^{\varnothing}[N(n, n+1]] \right]^2 \right\}$$

$$\leq \mathbb{E}\left[\left(\sum_{\tau \in \omega_1^-} K(n, \tau) \right)^2 \right]$$

$$\leq \mathbb{E}\left[\sum_{i,j \leq 0} K(n, i)K(n, j)N[i, i+1]N[j, j+1] \right]$$

$$= \sum_{i,j \leq 0} K(n, i)K(n, j)\mathbb{E}[N[i, i+1]N[j, j+1]]$$

$$\leq \sum_{i,j \leq 0} K(n, i)K(n, j)\frac{1}{2}\left\{ \mathbb{E}[N[i, i+1]^2] + \mathbb{E}[N[j, j+1]^2] \right\}$$

$$= \mathbb{E}[N[0, 1]^2]\left(\sum_{i \leq 0} K(n, i) \right)^2.$$

Here, $\mathbb{E}[N[0, 1]^2] < \infty$ by Lemma 3. Therefore, we have

(2.19)
$$\sum_{n=1}^{\infty}\left\{ \mathbb{E}\left[\left(\mathbb{E}[N(n, n+1] - \mu|\mathcal{F}_0^{-\infty}] \right)^2 \right] \right\}^{1/2}$$

$$\leq \sqrt{2\mathbb{E}[N[0, 1]^2]} \sum_{n=1}^{\infty} \sum_{i=-\infty}^{0} K(n, i)$$

$$\leq \sqrt{2\mathbb{E}[N[0, 1]^2]} \sum_{k=1}^{\infty} \alpha^k \int_0^{\infty} \int_0^{t_k} \cdots \int_0^{t_2} \int_{-\infty}^{0}$$

$$h(t_k - t_{k-1})h(t_{k-1} - t_{k-2}) \cdots h(t_2 - t_1)h(t_1 - s)dsdt_1 \cdots dt_k.$$

Let $H(t) := \int_t^{\infty} h(s)ds$. It is easy to check that $\int_0^{\infty} H(t)dt = \int_0^{\infty} th(t)dt < \infty$ by

Assumption 1. We have

(2.20)

$$\alpha^k \int_0^\infty \int_0^{t_k} \cdots \int_0^{t_2} \int_{-\infty}^0$$

$$h(t_k - t_{k-1})h(t_{k-1} - t_{k-2}) \cdots h(t_2 - t_1)h(t_1 - s)dsdt_1 \cdots dt_k$$

$$= \alpha^k \int_0^\infty \int_0^{t_k} \cdots \int_0^{t_2} h(t_k - t_{k-1})h(t_{k-1} - t_{k-2}) \cdots h(t_2 - t_1)H(t_1)dt_1 \cdots dt_k$$

$$= \alpha^k \int_0^\infty \cdots \int_{t_{k-2}}^\infty \int_{t_{k-1}}^\infty h(t_k - t_{k-1})dt_k h(t_{k-1} - t_{k-2})dt_{k-1} \cdots H(t_1)dt_1$$

$$= \alpha^k \|h\|_{L^1}^{k-1} \int_0^\infty H(t_1)dt_1 = \alpha^k \|h\|_{L^1}^{k-1} \int_0^\infty th(t)dt.$$

Since $\alpha\|h\|_{L^1} < 1$, we conclude that

(2.21)
$$\sum_{n=1}^\infty \left\{ \mathbb{E}\left[\left(\mathbb{E}[N(n, n+1] - \mu | \mathcal{F}_0^{-\infty}] \right)^2 \right] \right\}^{1/2}$$

$$\leq \sum_{k=1}^\infty \sqrt{2\mathbb{E}[N[0,1]^2]} \alpha^k \|h\|_{L^1}^{k-1} \int_0^\infty th(t)dt$$

$$= \sqrt{2\mathbb{E}[N[0,1]^2]} \cdot \frac{\alpha}{1 - \alpha\|h\|_{L^1}} \cdot \int_0^\infty th(t)dt < \infty.$$

Hence, by Theorem 8, we have

(2.22)
$$\frac{N_{[\cdot t]} - \mu[\cdot t]}{\sqrt{t}} \to \sigma B(\cdot) \quad \text{as } t \to \infty,$$

where

(2.23) $$\sigma^2 = \mathbb{E}[(N[0,1] - \mu)^2] + 2\sum_{j=1}^\infty \mathbb{E}[(N[0,1] - \mu)(N[j, j+1] - \mu)] < \infty.$$

By Lemma 4, $\sigma > 0$. Now, finally, for any $\epsilon > 0$, for t sufficiently large,

$$(2.24) \qquad \mathbb{P}\left(\sup_{0 \leq s \leq 1} \left| \frac{N_{[st]} - \mu[st]}{\sqrt{t}} - \frac{N_{st} - \mu st}{\sqrt{t}} \right| > \epsilon \right)$$

$$= \mathbb{P}\left(\sup_{0 \leq s \leq 1} \left| (N_{[st]} - N_{st}) + \mu(st - [st]) \right| > \epsilon\sqrt{t} \right)$$

$$\leq \mathbb{P}\left(\sup_{0 \leq s \leq 1} \left| N_{[st]} - N_{st} \right| + \mu > \epsilon\sqrt{t} \right)$$

$$\leq \mathbb{P}\left(\max_{0 \leq k \leq [t], k \in \mathbb{Z}} N[k, k+1] > \epsilon\sqrt{t} - \mu \right)$$

$$\leq ([t] + 1)\mathbb{P}(N[0, 1] > \epsilon\sqrt{t} - \mu)$$

$$\leq \frac{[t] + 1}{(\epsilon\sqrt{t} - \mu)^2} \int_{N[0,1] > \epsilon\sqrt{t} - \mu} N[0, 1]^2 d\mathbb{P} \to 0,$$

as $t \to \infty$ by Lemma 3. Hence, we conclude that $\frac{N_{\cdot t} - \cdot \mu t}{\sqrt{t}} \to \sigma B(\cdot)$ as $t \to \infty$. $\qquad \square$

The following Lemma 2 is used to prove Lemma 3.

Lemma 2. *There exists some $\theta > 0$ such that* $\sup_{t \geq 0} \mathbb{E}^\varnothing \left[e^{\int_0^t \theta h(t-s)N(ds)} \right] < \infty$.

Proof. Notice first that for any bounded deterministic function $f(\cdot)$,

$$(2.25) \qquad \exp\left\{ \int_0^t f(s)N(ds) - \int_0^t (e^{f(s)} - 1)\lambda(s)ds \right\}$$

is a martingale. Therefore, using the Lipschitz assumption of $\lambda(\cdot)$, i.e. $\lambda(z) \leq$

$\lambda(0) + \alpha z$ and applying Hölder's inequality, for $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$(2.26) \qquad \mathbb{E}^{\varnothing}\left[e^{\int_0^t \theta h(t-s)N(ds)}\right]$$

$$= \mathbb{E}^{\varnothing}\left[e^{\int_0^t \theta h(t-s)N(ds) - \frac{1}{p}\int_0^t (e^{p\theta h(t-s)}-1)\lambda(s)ds + \frac{1}{p}\int_0^t (e^{p\theta h(t-s)}-1)\lambda(s)ds}\right]$$

$$\leq \mathbb{E}^{\varnothing}\left[e^{\frac{q}{p}\int_0^t (e^{p\theta h(t-s)}-1)\lambda(s)ds}\right]^{\frac{1}{q}}$$

$$\leq \mathbb{E}^{\varnothing}\left[e^{\frac{q}{p}\int_0^t (e^{p\theta h(t-s)}-1)(\lambda(0)+\alpha\int_0^s h(s-u)N(du))ds}\right]^{\frac{1}{q}}$$

$$\leq \mathbb{E}^{\varnothing}\left[e^{\int_0^t \frac{q}{p}(e^{p\theta h(t-s)}-1)\alpha\int_0^s h(s-u)N(du)ds}\right]^{\frac{1}{q}} \cdot e^{\frac{1}{p}\int_0^\infty (e^{p\theta h(s)}-1)\lambda(0)ds}.$$

Let $C(t) = \int_0^t \frac{q}{p}(e^{p\theta h(t-s)} - 1)\alpha ds$. Then, for any $t \in [0,T]$,

$$(2.27) \qquad \mathbb{E}^{\varnothing}\left[e^{\int_0^t \frac{q}{p}(e^{p\theta h(t-s)}-1)\alpha\int_0^s h(s-u)N(du)ds}\right]$$

$$= \mathbb{E}^{\varnothing}\left[e^{\frac{1}{C(t)}\int_0^t \frac{q}{p}(e^{p\theta h(t-s)}-1)\alpha C(t)\int_0^s h(s-u)N(du)ds}\right]$$

$$\leq \mathbb{E}^{\varnothing}\left[\frac{1}{C(t)}\int_0^t \frac{q}{p}(e^{p\theta h(t-s)}-1)\alpha e^{C(t)\int_0^s h(s-u)N(du)}ds\right]$$

$$\leq \sup_{0\leq s\leq T} \mathbb{E}^{\varnothing}\left[e^{C(\infty)\int_0^s h(s-u)N(du)}\right],$$

where in the first inequality in (2.27), we used the Jensen's inequality since $x \mapsto e^x$ is convex and $\frac{1}{C(t)}\int_0^t \frac{q}{p}(e^{p\theta h(t-s)} - 1)\alpha ds = 1$, and in the second inequality in (2.27), we used the fact that $C(t) \leq C(\infty)$ and again $\frac{1}{C(t)}\int_0^t \frac{q}{p}(e^{p\theta h(t-s)} - 1)\alpha ds = 1$. Now choose $q > 1$ so small that $q\alpha\|h\|_{L^1} < 1$. Once p and q are fixed, choose $\theta > 0$ so small that

$$(2.28) \qquad C(\infty) = \int_0^\infty \frac{q}{p}(e^{p\theta h(s)} - 1)\alpha ds < \theta.$$

This implies that for any $t \in [0, T]$,

$$(2.29) \quad \mathbb{E}^{\varnothing}\left[e^{\int_0^t \theta h(t-s)N(ds)}\right] \leq \sup_{0 \leq s \leq T} \mathbb{E}^{\varnothing}\left[e^{\theta \int_0^s h(s-u)N(du)}\right]^{\frac{1}{q}} \cdot e^{\frac{1}{p}\int_0^{\infty}(e^{p\theta h(s)}-1)\lambda(0)ds}.$$

Hence, we conclude that for any $T > 0$,

$$(2.30) \quad \sup_{0 \leq t \leq T} \mathbb{E}^{\varnothing}\left[e^{\theta \int_0^t h(t-s)N(ds)}\right] \leq e^{\int_0^{\infty}(e^{p\theta h(s)}-1)\lambda(0)ds} < \infty.$$

$\square$

Lemma 3. *There exists some $\theta > 0$ such that $\mathbb{E}[e^{\theta N[0,1]}] < \infty$. Hence $\mathbb{E}[N[0,1]^2] < \infty$.*

Proof. By Assumption 1, $h(\cdot)$ is positive and decreasing. Thus, $\delta = \inf_{t \in [0,1]} h(t) > 0$. Hence,

$$(2.31) \quad \mathbb{E}^{\varnothing}[e^{\theta N[t-1,t]}] \leq \mathbb{E}^{\varnothing}[e^{\frac{\theta}{\delta}\int_0^t h(t-s)N(ds)}].$$

By Lemma 2, we can choose $\theta > 0$ so small that

$$(2.32) \quad \limsup_{t \to \infty} \mathbb{E}^{\varnothing}[e^{\theta N[t-1,t]}] < \infty.$$

Finally, $\mathbb{E}[e^{\theta N[0,1]}] \leq \liminf_{t \to \infty} \mathbb{E}^{\varnothing}[e^{\theta N[t-1,t]}] < \infty.$ $\square$

It is intuitively clear that $\sigma > 0$. But still we need a proof.

Lemma 4. $\sigma > 0$, *where σ is defined in* (2.23).

Proof. Let $\eta_n = \sum_{j=n}^{\infty} \mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+1}^{-\infty}]$, where $\mu = \mathbb{E}[N[0,1]]$. η_n is well

defined because we proved (2.7). To see this, notice that

$$(2.33) \qquad \|\eta_n\|_2 = \left\| \sum_{j=n}^{\infty} \mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+1}^{-\infty}] \right\|_2$$

$$\leq \sum_{j=n}^{\infty} \|\mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+1}^{-\infty}]\|_2 < \infty,$$

by (2.7). Also, it is easy to check that

$$(2.34) \qquad \mathbb{E}[\eta_{n+1} - \eta_n + N(n, n+1] - \mu | \mathcal{F}_{n+1}^{-\infty}]$$

$$= \mathbb{E}\left[\sum_{j=n+1}^{\infty} \mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+2}^{-\infty}] \Big| \mathcal{F}_{n+1}^{-\infty} \right]$$

$$- \mathbb{E}\left[\sum_{j=n}^{\infty} \mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+1}^{-\infty}] \Big| \mathcal{F}_{n+1}^{-\infty} \right] + N(n, n+1] - \mu$$

$$= \sum_{j=n+1}^{\infty} \mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+1}^{-\infty}] - \sum_{j=n+1}^{\infty} \mathbb{E}[N(j, j+1] - \mu | \mathcal{F}_{n+1}^{-\infty}]$$

$$- N(n, n+1] + \mu + N(n, n+1] - \mu = 0.$$

Let $Y_n = \eta_{n-1} - \eta_{n-2} + N(n-2, n-1] - \mu$. This is an ergodic, stationary sequence such that $\mathbb{E}[Y_n | \mathcal{F}_{n-1}^{-\infty}] = 0$. By (2.7), $\mathbb{E}[Y_n^2] < \infty$ and by Theorem 9, $S'_{[n \cdot]}/\sqrt{n} \to \sigma' B(\cdot)$, where $S'_n = \sum_{j=1}^{n} Y_j$. It is clear that $\sigma = \sigma' < \infty$ since for any

$\epsilon > 0$,

$$(2.35) \qquad \mathbb{P}\left(\max_{1\leq k\leq [n], k\in\mathbb{Z}} \frac{1}{\sqrt{n}} \sum_{j=1}^{k} (\eta_{j-1} - \eta_{j-2}) > \epsilon\right)$$

$$= \mathbb{P}\left(\max_{1\leq k\leq [n], k\in\mathbb{Z}} (\eta_{k-1} - \eta_{-1}) > \epsilon\sqrt{n}\right)$$

$$\leq \mathbb{P}\left(\left\{\max_{1\leq k\leq [n], k\in\mathbb{Z}} |\eta_{k-1}| > \frac{\epsilon\sqrt{n}}{2}\right\} \bigcup \left\{|\eta_{-1}| > \frac{\epsilon\sqrt{n}}{2}\right\}\right)$$

$$\leq \sum_{k=1}^{[n]} \mathbb{P}\left(|\eta_{k-1}| > \frac{\epsilon\sqrt{n}}{2}\right) + \mathbb{P}\left(|\eta_{-1}| > \frac{\epsilon\sqrt{n}}{2}\right)$$

$$= ([n] + 1)\mathbb{P}\left(|\eta_{-1}| > \frac{\epsilon\sqrt{n}}{2}\right)$$

$$\leq \frac{4([n] + 1)}{\epsilon^2 n} \int_{|\eta_{-1}| > \frac{\epsilon\sqrt{n}}{2}} |\eta_{-1}|^2 d\mathbb{P} \to 0,$$

as $n \to \infty$, where we used the stationarity of $\mathbb{P}$, Chebychev's inequality and (2.7).

Now, it becomes clear that

$$(2.36) \quad \sigma^2 = (\sigma')^2 = \mathbb{E}[Y_1^2]$$

$$= \mathbb{E}\left(\eta_0 - \eta_{-1} + N(-1, 0] - \mu\right)^2$$

$$= \mathbb{E}\left(\sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1] - \mu|\mathcal{F}_1^{-\infty}] - \sum_{j=0}^{\infty} \mathbb{E}[N(j, j+1] - \mu|\mathcal{F}_0^{-\infty}]\right)^2.$$

Consider $D = \{\omega : \omega^- \neq \varnothing, \omega(0, 1] = \varnothing\}$. Notice that $\mathbb{P}(\omega^- = \varnothing) = 0$. By

Jensen's inequality and Assumption 1, we have

$$(2.37) \qquad \mathbb{P}(D) = \int \mathbb{P}^{\omega^-}(N(0,1] = 0)\mathbb{P}(d\omega^-)$$

$$= \mathbb{E}\left[e^{-\int_0^1 \lambda(\sum_{\tau \in \omega^-} h(t-\tau))dt}\right]$$

$$\geq \exp\left\{-\mathbb{E}\int_0^1 \lambda\left(\sum_{\tau \in \omega^-} h(t-\tau)\right)dt\right\}$$

$$\geq \exp\left\{-\lambda(0) - \alpha\mathbb{E}\int_0^1 \sum_{\tau \in \omega^-} h(t-\tau)dt\right\}$$

$$\geq \exp\left\{-\lambda(0) - \alpha\mathbb{E}[N[0,1]] \cdot \|h\|_{L^1}\right\} > 0.$$

It is clear that given the event D,

$$(2.38) \qquad \sum_{j=0}^{\infty} \mathbb{E}[N(j,j+1] - \mu|\mathcal{F}_1^{-\infty}] < \sum_{j=0}^{\infty} \mathbb{E}[N(j,j+1] - \mu|\mathcal{F}_0^{-\infty}].$$

Therefore,

$$(2.39) \qquad \mathbb{P}\left(\sum_{j=0}^{\infty} \mathbb{E}[N(j,j+1] - \mu|\mathcal{F}_1^{-\infty}] \neq \sum_{j=0}^{\infty} \mathbb{E}[N(j,j+1] - \mu|\mathcal{F}_0^{-\infty}]\right) > 0,$$

which implies that $\sigma > 0$. $\qquad \square$

Proof of Theorem 7. By Heyde and Scott [57], the Strassen's invariance principle holds if we have (2.7) and $\sigma > 0$. $\qquad \square$

Chapter 3

Process-Level Large Deviations

for Nonlinear Hawkes Processes

3.1 Main Results

In this chaper, we prove a process-level, i.e. level-3 large deviation principle for nonlinear Hawkes processes. As a corollary, a level-1 large deviation principle is obtained by a contraction principle.

Let us recall that N is a nonlinear Hawkes process with intensity

$$(3.1) \qquad\qquad \lambda_t := \lambda\left(\int_{(-\infty,t)} h(t-s)N(ds)\right).$$

Throughout this chapter, we assume that

- The exciting function $h(t)$ is positive, continuous and decreasing for $t \geq 0$ and $h(t) = 0$ for any $t < 0$. We also assume that $\int_0^\infty h(t)dt < \infty$.

- The rate function $\lambda(\cdot) : [0,\infty) \to \mathbb{R}^+$ is increasing and $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$. We

also assume that $\lambda(\cdot)$ is Lipschitz with constant $\alpha > 0$, i.e. $|\lambda(x) - \lambda(y)| \leq \alpha|x - y|$ for any $x, y \geq 0$.

Let Ω be the set of countable, locally finite subsets of $\mathbb{R}$ and for any $\omega \in \Omega$ and $A \subseteq \mathbb{R}$, write $\omega(A) := \omega \cap A$. For any $t \in \mathbb{R}$, we write $\omega(t) = \omega(\{t\})$. Let $N(A) = \#|\omega \cap A|$ denote the number of points in the set A for any $A \subset \mathbb{R}$. We also use the notation N_t to denote $N[0, t]$, the number of points up to time t, starting from time 0. We define the shift operator θ_t by $\theta_t(\omega)(s) = \omega(t + s)$. We equip the sample space Ω with the topology in which the convergence $\omega_n \to \omega$ as $n \to \infty$ is defined by

$$(3.2) \qquad \sum_{\tau \in \omega_n} f(\tau) \to \sum_{\tau \in \omega} f(\tau),$$

for any continuous f with compact support.

This topology is equivalent to the vague topology for random measures, for which, see for example Grandell [45]. One can equip the space of locally finite random measures with the vague topology. The subspace of integer valued random measures is then the space of point processes. A simple point processes is a point process without multiple jumps. The space of point processes is closed. But the space of simple point processes is not closed.

Denote $\mathcal{F}_t^s = \sigma(\omega[s, t])$ for any $s < t$, i.e. the σ-algebra generated by all the possible configurations of points in the interval $[s, t]$. Denote $\mathcal{M}(\Omega)$ the space of probability measures on Ω. We also define $\mathcal{M}_S(\Omega)$ as the space of simple point processes that are invariant with respect to θ_t with bounded first moment, i.e. for any $Q \in \mathcal{M}_S(\Omega)$, $\mathbb{E}^Q[N[0, 1]] < \infty$. Define $\mathcal{M}_E(\Omega)$ as the set of ergodic simple point processes in $\mathcal{M}_S(\Omega)$. We define the topology of $\mathcal{M}_S(\Omega)$ as follows. For a

sequence Q_n in $\mathcal{M}_S(\Omega)$ and $Q \in \mathcal{M}_S(\Omega)$, we say $Q_n \to Q$ as $n \to \infty$ if and only if

$$(3.3) \qquad \int f dQ_n \to \int f dQ,$$

as $n \to \infty$ for any continuous and bounded f and

$$(3.4) \qquad \int N[0,1](\omega)Q_n(d\omega) \to \int N[0,1](\omega)Q(d\omega),$$

as $n \to \infty$. In other words, the topology is the weak topology strengthened by the convergence of the first moment of $N[0,1]$. For any Q_1, Q_2 in $\mathcal{M}_S(\Omega)$, one can define the metric $d(\cdot,\cdot)$ by

$$(3.5) \qquad d(Q_1,Q_2) = d_p(Q_1,Q_2) + \left| \mathbb{E}^{Q_1}[N[0,1]] - \mathbb{E}^{Q_2}[N[0,1]] \right|,$$

where $d_p(\cdot,\cdot)$ is the usual Prokhorov metric. Because this is an unusual topology, the compactness is different from that in the usual weak topology; later, when we prove the exponential tightness, we need to take some extra care. See Lemma 22 and (iii) of Lemma 21.

We denote by $C(\Omega)$ the set of real-valued continous functions on Ω. We similarly define $C(\Omega \times \mathbb{R})$. We also denote by $\mathcal{B}(\mathcal{F}_t^{-\infty})$ the set of all bounded $\mathcal{F}_t^{-\infty}$ progressively measurable and $\mathcal{F}_t^{-\infty}$ predictable functions.

Before we proceed, recall that a sequence $(P_n)_{n\in\mathbb{N}}$ of probability measures on a topological space X satisfies the large deviation principle (LDP) with rate function $I : X \to \mathbb{R}$ if I is non-negative, lower semicontinuous and for any measurable set

A,

$$- \inf_{x \in A^o} I(x) \le \liminf_{n \to \infty} \frac{1}{n} \log P_n(A) \le \limsup_{n \to \infty} \frac{1}{n} \log P_n(A) \le - \inf_{x \in \overline{A}} I(x).$$

Here, A^o is the interior of A and $\overline{A}$ is its closure. See Dembo and Zeitouni [30] or Varadhan [106] for general background regarding large deviations and their applications. Also Varadhan [107] has an excellent survey article on this subject.

In the pioneering work by Donsker and Varadhan [31], they obtained a level-3 large deviation result for certain stationary Markov processes.

We would like to prove the large deviation principle for nonlinear Hawkes processes by proving a process-level, also known as level-3 large deviation principle first. We can then use the contraction principle to obtain the level-1 large deviation principle for $(N_t/t \in \cdot)$.

Let us define the empirical measure for the process as

(3.7) $$R_{t,\omega}(A) = \frac{1}{t} \int_0^t \chi_A(\theta_s \omega_t) ds,$$

for any A, where $\omega_t(s) = \omega(s)$ for $0 \le s \le t$ and $\omega_t(s+t) = \omega_t(s)$ for any s. Donsker and Varadhan [31] proved that in the case when Ω is a space of càdlàg functions $\omega(\cdot)$ on $-\infty < t < \infty$ endowed with Skorohod topology and taking values in a Polish space X, under certain conditions, $P^{0,x}(R_{t,\omega} \in \cdot)$ satisfies a large deviation principle, where $P^{0,x}$ is a Markov process on Ω_∞^0 with initial value $x \in X$. The rate function $H(Q)$ is some entropy function.

Let $h(\alpha, \beta)_\Sigma$ be the relative entropy of α with respect to β restricted to the σ-algebra Σ. For any $Q \in \mathcal{M}_S(\Omega)$, let Q^{ω^-} be the regular conditional probability distribution of Q. Similarly we define P^{ω^-}.

Let us define the entropy function $H(Q)$ as

$$H(Q) = \mathbb{E}^Q[h(Q^{\omega^-}, P^{\omega^-})_{\mathcal{F}_1^0}]. \tag{3.8}$$

Notice that P^{ω^-} describes the Hawkes process conditional on the past history ω^-. It has rate $\lambda(\sum_{\tau \in \omega[0,s) \cup \omega^-} h(s - \tau))$ at time $0 \le s \le 1$, which is well defined for almost every ω^- under Q if $\mathbb{E}^Q[N[0,1]] < \infty$ since $\mathbb{E}^Q[\sum_{\tau \in \omega^-} h(-\tau)] = \|h\|_{L^1}\mathbb{E}^Q[N[0,1]] < \infty$ implies $\sum_{\tau \in \omega^-} h(s - \tau) \le \sum_{\tau \in \omega^-} h(-\tau) < \infty$ for all $0 \le s \le 1$.

When $H(Q) < \infty$, $h(Q^{\omega^-}, P^{\omega^-}) < \infty$ for a.e. ω^- under Q, which implies that $Q^{\omega^-} \ll P^{\omega^-}$ on $\mathcal{F}_1^0$. By the theory of absolute continuity of point processes, see for example Chapter 19 of Lipster and Shiryaev [72] or Chapter 13 of Daley and Vere-Jones [27], the compensator of Q^{ω^-} is absolutely continuous, i.e. it has some density $\hat{\lambda}$ say, such that by the Girsanov formula,

$$\begin{aligned}
H(Q) &= \int_{\Omega^-} \int \left[\int_0^1 \left(\lambda - \hat{\lambda}\right) ds + \int_0^1 \log(\hat{\lambda}/\lambda) dN_s\right] dQ^{\omega^-} Q(d\omega^-) \\
&= \int_\Omega \left[\int_0^1 \lambda(\omega, s) - \hat{\lambda}(\omega, s) + \log\left(\frac{\hat{\lambda}(\omega, s)}{\lambda(\omega, s)}\right) \hat{\lambda} ds\right] Q(d\omega),
\end{aligned} \tag{3.9}$$

where $\lambda = \lambda\left(\sum_{\tau \in \omega[0,s) \cup \omega^-} h(s - \tau)\right)$. Both λ and $\hat{\lambda}$ are $\mathcal{F}_s^{-\infty}$-predictable for $0 \le s \le 1$. For the equality in (3.9), we used the fact that $N_t - \int_0^t \hat{\lambda}(\omega, s)ds$ is a martingale under Q and for any $f(\omega, s)$ which is bounded, $\mathcal{F}_s^{-\infty}$ progressively measurable and predictable, we have

$$\int_\Omega \int_0^1 f(\omega, s)dN_s Q(d\omega) = \int_\Omega \int_0^1 f(\omega, s)\hat{\lambda}(\omega, s)ds Q(d\omega). \tag{3.10}$$

61

We will use the above fact repeatedly in this chapter.

The following theorem is the main result of this chapter.

Theorem 10. *For any open set $G \subset \mathcal{M}_S(\Omega)$,*

$$(3.11) \qquad \liminf_{t \to \infty} \frac{1}{t} \log P\left(R_{t,\omega} \in G\right) \geq - \inf_{Q \in G} H(Q),$$

and for any closed set $C \subset \mathcal{M}_S(\Omega)$,

$$(3.12) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P\left(R_{t,\omega} \in C\right) \leq - \inf_{Q \in C} H(Q).$$

We will prove the lower bound in Section 3.2, the upper bound in Section 3.3, and the superexponential estimates that are needed in the proof of the upper bound in Section 3.4.

Once we establish the level-3 large deviation result, we can obtain the large deviation principle for $(N_t/t \in \cdot)$ directly by using the contraction principle.

Theorem 11. *$(N_t/t \in \cdot)$ satisfies a large deviation principle with the rate function $I(\cdot)$ given by*

$$(3.13) \qquad I(x) = \inf_{Q \in \mathcal{M}_S(\Omega), \mathbb{E}^Q[N[0,1]] = x} H(Q).$$

Proof. Since $Q \mapsto \mathbb{E}^Q[N[0,1]]$ is continuous, $\int_\Omega N[0,1] dR_{t,\omega}$ satisfies a large deviation principle with the rate function $I(\cdot)$ by the contraction principle. (For a

62

discussion on contraction principle, see for example Varadhan [106].)

$$(3.14) \quad \int_\Omega N[0,1]dR_{t,\omega} = \frac{1}{t}\int_0^t N[0,1](\theta_s\omega_t)ds$$

$$= \frac{1}{t}\int_0^{t-1} N[s,s+1](\omega)ds + \frac{1}{t}\int_{t-1}^t N[s,s+1](\omega_t)ds.$$

Notice that

$$(3.15) \quad 0 \le \frac{1}{t}\int_{t-1}^t N[s,s+1](\omega_t)ds \le \frac{1}{t}(N[t-1,t](\omega)+N[0,1](\omega)),$$

and

$$(3.16) \quad \frac{1}{t}\int_0^{t-1} N[s,s+1](\omega)ds = \frac{1}{t}\left[\int_{t-1}^t N_s(\omega)ds - \int_0^1 N_s(\omega)ds\right] \le \frac{N_t}{t},$$

and

$$(3.17) \quad \frac{1}{t}\int_0^{t-1} N[s,s+1](\omega)ds \ge \frac{N_{t-1}-N_1}{t} = \frac{N_t}{t} - \frac{N[t-1,t]+N_1}{t}.$$

Hence,

$$(3.18) \quad \frac{N_t}{t} - \frac{N[t-1,t]+N_1}{t} \le \int_\Omega N[0,1]dR_{t,\omega} \le \frac{N_t}{t} + \frac{N[t-1,t]+N_1}{t}.$$

For the lower bound, for any open ball $B_\epsilon(x)$ centered at x with radius $\epsilon > 0$,

$$(3.19) \quad P\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \ge P\left(\int_\Omega N[0,1]dR_{t,\omega} \in B_{\epsilon/2}(x)\right)$$

$$- P\left(\frac{N[t-1,t]}{t} \ge \frac{\epsilon}{4}\right) - P\left(\frac{N_1}{t} \ge \frac{\epsilon}{4}\right).$$

63

For the upper bound, for any closed set C and $C^\epsilon = \bigcup_{x \in C} \overline{B_\epsilon(x)}$,

$$(3.20) \qquad P\left(\frac{N_t}{t} \in C\right) \leq P\left(\int_\Omega N[0,1]dR_{t,\omega} \in C^\epsilon\right)$$
$$+ P\left(\frac{N[t-1,t]}{t} \geq \frac{\epsilon}{4}\right) + P\left(\frac{N_1}{t} \geq \frac{\epsilon}{4}\right).$$

Finally, by Lemma 17, we have the following superexponential estimates

$$(3.21) \quad \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{N[t-1,t]}{t} \geq \frac{\epsilon}{4}\right) = \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{N_1}{t} \geq \frac{\epsilon}{4}\right) = -\infty.$$

Hence, for the lower bound, we have

$$(3.22) \qquad\qquad \liminf_{t\to\infty} \frac{1}{t} \log P\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq -I(x),$$

and for the upper bound, we have

$$(3.23) \qquad\qquad \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{N_t}{t} \in C\right) \leq -\inf_{x \in C^\epsilon} I(x),$$

which holds for any $\epsilon > 0$. Letting $\epsilon \downarrow 0$, we get the desired result. $\qquad\square$

3.2 Lower Bound

Lemma 5. *For any $\lambda, \hat{\lambda} \geq 0$, $\lambda - \hat{\lambda} + \hat{\lambda}\log(\hat{\lambda}/\lambda) \geq 0$.*

Proof. Write $\lambda - \hat{\lambda} + \hat{\lambda}\log(\hat{\lambda}/\lambda) = \hat{\lambda}\left[(\lambda/\hat{\lambda}) - 1 - \log(\lambda/\hat{\lambda})\right]$. Thus, it is sufficient to show that $F(x) = x - 1 - \log x \geq 0$ for any $x \geq 0$. Note that $F(0) = F(\infty) = 0$ and $F'(x) = 1 - \frac{1}{x} < 0$ when $0 < x < 1$ and $F'(x) > 0$ when $x > 1$ and finally $F(1) = 0$. Hence $F(x) \geq 0$ for any $x \geq 0$. $\qquad\square$

Lemma 6. *Assume $H(Q) < \infty$. Then,*

$$
(3.24) \qquad\qquad \mathbb{E}^Q[N[0,1]] \leq C_1 + C_2 H(Q),
$$

where $C_1, C_2 > 0$ are some constants independent of Q.

Proof. If $H(Q) < \infty$, then $h(Q^{\omega^-}, P^{\omega^-})_{\mathcal{F}_1^0} < \infty$ for a.e. ω^- under Q, which implies that $Q^{\omega^-} \ll P^{\omega^-}$ and thus $\hat{A}_t \ll A_t$, where $\hat{A}_t$ and A_t are the compensators of N_t under Q^{ω^-} and P^{ω^-} respectively. (For the theory of absolute continuity of point processes and Girsanov formula, see for example Lipster and Shiryaev [72] or Daley and Vere-Jones [27].) Since $A_t = \int_0^t \lambda(\omega, s) ds$, we have $\hat{A}_t = \int_0^t \hat{\lambda}(\omega, s) ds$ for some $\hat{\lambda}$. By the Girsanov formula,

$$
(3.25) \qquad\qquad H(Q) = \mathbb{E}^Q \left[\int_0^1 \lambda - \hat{\lambda} + \log\left(\hat{\lambda}/\lambda\right) \hat{\lambda} ds \right].
$$

Notice that $\mathbb{E}^Q[N[0,1]] = \int \int_0^1 \hat{\lambda} \, ds \, dQ$.

$$
(3.26) \qquad \int \int_0^1 \lambda \, ds \, dQ \leq \epsilon \int \int_0^1 \sum_{\tau < s} h(s - \tau) ds \, dQ + C_\epsilon
$$

$$
\leq \epsilon \int h(0) N[0,1] dQ + \epsilon \int \sum_{\tau < 0} h(-\tau) dQ + C_\epsilon
$$

$$
= \epsilon(h(0) + \|h\|_{L^1}) \mathbb{E}^Q[N[0,1]] + C_\epsilon
$$

$$
= \epsilon(h(0) + \|h\|_{L^1}) \int \int_0^1 \hat{\lambda} \, ds \, dQ + C_\epsilon.
$$

Therefore, we have

$$
(3.27) \qquad \int \int_0^1 \hat{\lambda} \cdot 1_{\hat{\lambda} < K\lambda} ds \, dQ \leq K\epsilon(h(0) + \|h\|_{L^1}) \int \int_0^1 \hat{\lambda} \, ds \, dQ + KC_\epsilon.
$$

65

On the other hand, by Lemma 5,

$$(3.28) \qquad H(Q) \geq \int \int_0^1 \left[\lambda - \hat{\lambda} + \hat{\lambda} \log(\hat{\lambda}/\lambda) \right] \cdot 1_{\hat{\lambda} \geq K\lambda} ds dQ$$

$$\geq (\log K - 1) \int \int_0^1 \hat{\lambda} \cdot 1_{\hat{\lambda} \geq K\lambda} ds dQ.$$

Thus,

$$(3.29) \qquad \int \int_0^1 \hat{\lambda} ds dQ \leq K\epsilon(h(0) + \|h\|_{L^1}) \int \int_0^1 \hat{\lambda} ds dQ + KC_\epsilon + \frac{H(Q)}{\log K - 1}.$$

Choosing $K > e$ and $\epsilon < \frac{1}{K(h(0) + \|h\|_{L^1})}$, we get

$$(3.30) \qquad \mathbb{E}^Q[N[0,1]] \leq \frac{KC_\epsilon}{1 - K\epsilon(h(0) + \|h\|_{L^1})} + \frac{H(Q)}{(\log K - 1)K\epsilon(h(0) + \|h\|_{L^1})}.$$

$\square$

Lemma 7. *We have the following alternative expression for $H(Q)$.*

$$(3.31) \qquad H(Q) = \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \leq s \leq 1} \mathbb{E}^Q \left[\int_0^1 \lambda(1 - e^f) ds + \int_0^1 f dN_s \right].$$

Proof. $\mathbb{E}^Q[N[0,1]] < \infty$ implies that $\mathbb{E}^{Q^{\omega^-}}[N[0,1]] < \infty$ for almost every ω^- under Q, also $\sum_{\tau \in \omega^-} h(-\tau) < \infty$ since $\mathbb{E}^Q[\sum_{\tau \in \omega^-} h(-\tau)] = \|h\|_{L^1} \mathbb{E}^Q[N[0,1]] < \infty$. Thus,

$$(3.32) \qquad \mathbb{E}^{P^{\omega^-}}[N[0,1]] = \mathbb{E}^{P^{\omega^-}} \left[\int_0^1 \lambda \left(\sum_{\tau \in \omega[0,s) \cup \omega^-} h(s - \tau) \right) ds \right]$$

$$\leq C_\epsilon + \epsilon h(0) \mathbb{E}^{P^{\omega^-}}[N[0,1]] + \epsilon \sum_{\tau \in \omega^-} h(-\tau) < \infty,$$

66

so $\mathbb{E}^{P^{\omega^-}}[N[0,1]] < \infty$ by choice of $\epsilon < \frac{1}{h(0)}$.

By the theory of absolute continuity of point processes, see for example Chapter 13 of Daley and Vere-Jones [27], if $\mathbb{E}^{Q^{\omega^-}}[N[0,1]], \mathbb{E}^{P^{\omega^-}}[N[0,1]] < \infty$, $Q^{\omega^-} \ll P^{\omega^-}$ if and only if $\hat{A}_t \ll A_t$, where $\hat{A}_t$ and $A_t = \int_0^t \lambda(\omega^-, \omega, s)ds$ are the compensators of N_t under Q^{ω^-} and P^{ω^-} respectively. If that's the case, we can write $\hat{A}_t = \int_0^t \hat{\lambda}(\omega^-, \omega, s)ds$ for some $\hat{\lambda}$ and there is Girsanov formula

$$(3.33) \qquad \log \frac{dQ^{\omega^-}}{dP^{\omega^-}}\bigg|_{\mathcal{F}_1^0} = \int_0^1 (\lambda - \hat{\lambda})ds + \int_0^1 \log\left(\hat{\lambda}/\lambda\right) dN_s,$$

which implies that

$$(3.34) \qquad H(Q) = \mathbb{E}^Q\left[\int_0^1 \lambda - \hat{\lambda} + \log\left(\hat{\lambda}/\lambda\right)\hat{\lambda}ds\right].$$

For any f, $\hat{\lambda}f + (1 - e^f)\lambda \le \hat{\lambda}\log(\hat{\lambda}/\lambda) + \lambda - \hat{\lambda}$ and the equality is achieved when $f = \log(\hat{\lambda}/\lambda)$. Thus, clearly, we have

$$(3.35) \qquad \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \le s \le 1} \mathbb{E}^Q\left[\int_0^1 \lambda(1 - e^f)ds + \int_0^1 f dN_s\right] \le H(Q).$$

On the other hand, we can always find a sequence f_n convergent to $\log(\hat{\lambda}/\lambda)$ and by Fatou's lemma, we get the opposite inequality.

Now, assume that we do not have $Q^{\omega^-} \ll P^{\omega^-}$ for a.e. ω^- under Q. That implies that $H(Q) = \infty$. We want to show that

$$(3.36) \qquad \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \le s \le 1} \mathbb{E}^Q\left[\int_0^1 \lambda(1 - e^f)ds + \int_0^1 f dN_s\right] = \infty.$$

Let us assume that

$$(3.37) \qquad \sup_{f(\omega,s)\in\mathcal{B}(\mathcal{F}_s^{-\infty})\cap C(\Omega\times\mathbb{R}),0\leq s\leq 1} \mathbb{E}^Q\left[\int_0^1 \lambda(1-e^f)ds + \int_0^1 fdN_s\right] < \infty.$$

We want to prove that $H(Q) < \infty$.

Let $P_\epsilon^{\omega^-}$ be the point process on $[0,1]$ with compensator $A_t + \epsilon\hat{A}_t$. Clearly $\hat{A}_t \ll A_t + \epsilon\hat{A}_t$ and $Q^{\omega^-} \ll P_\epsilon^{\omega^-}$.

For any f,

$$
\begin{aligned}
(3.38) \quad & \mathbb{E}^Q\left[\int_0^1 (1-e^f)d(A_s + \epsilon\hat{A}_s) + fd\hat{A}_s\right] \\
&= \mathbb{E}^Q\left[\int_0^1 (1-e^f)\chi_{f<0}d(A_s + \epsilon\hat{A}_s) + f\chi_{f<0}d\hat{A}_s\right] \\
&\quad + \mathbb{E}^Q\left[\int_0^1 (1-e^f)\chi_{f\geq 0}d(A_s + \epsilon\hat{A}_s) + f\chi_{f\geq 0}d\hat{A}_s\right] \\
&\leq \mathbb{E}^Q\left[\int_0^1 d(A_s + \epsilon\hat{A}_s)\right] + \mathbb{E}^Q\left[\int_0^1 (1-e^f)\chi_{f\geq 0}dA_s + f\chi_{f\geq 0}d\hat{A}_s\right] \\
&= \mathbb{E}^Q\left[\int_0^1 d(A_s + \epsilon\hat{A}_s)\right] + \mathbb{E}^Q\left[\int_0^1 (1-e^{f\chi_{f\geq 0}})dA_s + f\chi_{f\geq 0}d\hat{A}_s\right] \\
&\leq C_\delta + \delta(h(0) + \|h\|_{L^1})\mathbb{E}^Q[N[0,1]] \\
&\quad + \sup_{f(\omega,s)\in\mathcal{B}(\mathcal{F}_s^{-\infty})\cap C(\Omega\times\mathbb{R}),0\leq s\leq 1} \mathbb{E}^Q\left[\int_0^1 \lambda(1-e^f)ds + \int_0^1 fdN_s\right] < \infty.
\end{aligned}
$$

Therefore,

(3.39)

$$\infty > \liminf_{\epsilon \downarrow 0} \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \leq s \leq 1} \mathbb{E}^Q \left[\int_0^1 (1 - e^f) d(A_s + \epsilon \hat{A}_s) + f d\hat{A}_s \right]$$

$$= \liminf_{\epsilon \downarrow 0} \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \leq s \leq 1}$$

$$\mathbb{E}^Q \left[\int_0^1 \left(1 - e^f + f \cdot \frac{d\hat{A}_s}{d(A_s + \epsilon \hat{A}_s)} \right) d(A_s + \epsilon \hat{A}_s) \right]$$

$$= \liminf_{\epsilon \downarrow 0} \mathbb{E}^Q [h(Q^{\omega^-}, P_\epsilon^{\omega^-})_{\mathcal{F}_1^0}]$$

$$= \mathbb{E}^Q [h(Q^{\omega^-}, P^{\omega^-})_{\mathcal{F}_1^0}] = H(Q),$$

by lower semicontinuity of the relative entropy $h(\cdot, \cdot)$, Fatou's lemma, and the fact that $P_\epsilon^{\omega^-} \to P^{\omega^-}$ weakly as $\epsilon \downarrow 0$. Hence $H(Q) < \infty$. $\qquad \square$

Lemma 8. $H(Q)$ *is lower semicontinuous and convex in* Q.

Proof. By Lemma 7, we can rewrite $H(Q)$ as

$$(3.40) \qquad H(Q) = \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \leq s \leq 1} \mathbb{E}^Q \left[\int_0^1 \lambda(1 - e^f) + \hat{\lambda} f ds \right]$$

$$= \sup_{f(\omega,s) \in \mathcal{B}(\mathcal{F}_s^{-\infty}) \cap C(\Omega \times \mathbb{R}), 0 \leq s \leq 1} \mathbb{E}^Q \left[\int_0^1 \lambda(1 - e^f) ds + \int_0^1 f dN_s \right].$$

If $Q_n \to Q$, then $\mathbb{E}^{Q_n}[N[0,1]] \to \mathbb{E}^Q[N[0,1]]$ and $Q_n \to Q$ weakly. Since $f(\omega, s) \in C(\Omega \times \mathbb{R}) \cap \mathcal{B}(\mathcal{F}_s^{-\infty})$, $\int_0^1 f(\omega, s) dN_s$ is continuous on Ω, and since f is uniformly bounded, $\int_0^1 f(\omega, s) dN_s \leq \|f\|_{L^\infty} N[0, 1]$. Hence,

$$(3.41) \qquad \mathbb{E}^{Q_n} \left[\int_0^1 f(\omega, s) dN_s \right] \to \mathbb{E}^Q \left[\int_0^1 f(\omega, s) dN_s \right].$$

69

Let $\lambda^M = \lambda\left(\sum_{\tau<s} h^M(s-\tau)\right)$, where $h^M(s) = h(s)\chi_{s\leq M}$. Then, $\lambda^M(\omega,s) \in C(\Omega \times \mathbb{R})$ and thus $\int_0^1 \lambda^M(1 - e^{f(\omega,s)})ds \in C(\Omega)$. Also, $\int_0^1 \lambda^M(1 - e^{f(\omega,s)})ds \leq K(1 + e^{\|f\|_{L^\infty}})N[-M, 1]$, where $K > 0$ is some constant. Therefore,

$$(3.42) \qquad \mathbb{E}^{Q_n}\left[\int_0^1 \lambda^M(1 - e^{f(\omega,s)})ds\right] \to \mathbb{E}^Q\left[\int_0^1 \lambda^M(1 - e^{f(\omega,s)})ds\right]$$

as $n \to \infty$. Next, notice that

$$(3.43) \qquad \left|\mathbb{E}^Q\left[\int_0^1 \lambda^M(1 - e^{f(\omega,s)})ds\right] - \mathbb{E}^Q\left[\int_0^1 \lambda(1 - e^{f(\omega,s)})ds\right]\right|$$
$$\leq \mathbb{E}^Q(1 + e^{\|f\|_{L^\infty}})\alpha\mathbb{E}^Q[N[0,1]]\int_M^\infty h(s)ds \to 0$$

as $M \to \infty$. Similarly, we have

(3.44)
$$\limsup_{M\to\infty}\limsup_{n\to\infty}\left|\mathbb{E}^{Q_n}\left[\int_0^1 \lambda^M(1 - e^{f(\omega,s)})ds\right] - \mathbb{E}^{Q_n}\left[\int_0^1 \lambda(1 - e^{f(\omega,s)})ds\right]\right| = 0.$$

Hence,

$$(3.45) \qquad \mathbb{E}^{Q_n}\left[\int_0^1 \lambda(\omega, s)(1 - e^{f(\omega,s)})ds\right] \to \mathbb{E}^Q\left[\int_0^1 \lambda(\omega, s)(1 - e^{f(\omega,s)})ds\right].$$

The supremum is taken over a linear functional of Q, which is continuous in Q, therefore the supremum over these linear functionals will be lower semicontinuous. Similarly, since in the variational formula expression of $H(Q)$ in Lemma 7, the supremum is taken over a linear functional of Q, $H(Q)$ is convex in Q. $\qquad\square$

Lemma 9. $H(Q)$ *is linear in* Q.

Proof. It is in general true that the process-level entropy function $H(Q)$ is linear in Q. Following the arguments in Donsker and Varadhan [31], there exists a subset

$\Omega_0 \subset \Omega$ which is $\mathcal{F}_0^{-\infty}$ measurable and a $\mathcal{F}_0^{-\infty}$ measurable map $\hat{Q} : \Omega_0 \to \mathcal{M}_E(\Omega)$ such that $Q(\Omega_0) = 1$ for all $Q \in \mathcal{M}_S(\Omega)$ and $Q(\omega : \hat{Q} = Q) = 1$ for all $Q \in \mathcal{M}_E(\Omega)$. Therefore, there exists a universal version, say $\hat{Q}^{\omega^-}$ independent of Q such that $\int \hat{Q}^{\omega^-} Q(d\omega^-) = Q$. Since that is true for all $Q \in \mathcal{M}_E(\Omega)$, it also holds for $Q \in \mathcal{M}_S(\Omega)$. Hence,

$$(3.46) \qquad H(Q) = \mathbb{E}^Q \left[h(Q^{\omega^-}, P^{\omega^-})_{\mathcal{F}_1^0} \right] = \mathbb{E}^Q \left[h(\hat{Q}^{\omega^-}, P^{\omega^-})_{\mathcal{F}_1^0} \right],$$

i.e. $H(Q)$ is linear in Q. $\qquad\qquad\square$

In this chapter, we are proving the large deviation principle for Hawkes processes started with empty history, i.e. with probability measure $P^\varnothing$. But when time elapses, the Hawkes process generates points and that create a new history. We need to understand how the history created affects the future. What we want to prove is some uniform estimates to the effect that if the past history is well controlled, then the new history will also be well controlled. This is essentially what the following Lemma 10 says. Consider the configuration of points starting from time 0 up to time t. We shift it by t and denote that by w_t such that $w_t \in \Omega^-$, where Ω^- is Ω restricted to $\mathbb{R}^-$. These notations will be used in Lemma 10.

Remark 2. *At the very beginning of the chapter, we defined ω_t. It should not be confused with w_t in this section.*

Lemma 10. *For any $Q \in \mathcal{M}_E(\Omega)$ such that $H(Q) < \infty$ and any open neighborhood N of Q, there exists some K_ℓ^- such that $\varnothing \in K_\ell^-$ and $Q(K_\ell^-) \to 1$ as $\ell \to \infty$ and*

$$(3.47) \qquad \liminf_{t \to \infty} \frac{1}{t} \inf_{w_0 \in K_\ell^-} \log P^{w_0}(R_{t,\omega} \in N, w_t \in K_\ell^-) \geq -H(Q).$$

Proof. Let us abuse the notations a bit by defining

$$
(3.48) \qquad \lambda(\omega^-) = \lambda \left(\sum_{\tau \in \omega^-, \tau \in \omega[0,s)} h(s-\tau) \right).
$$

For any $t > 0$, since $\lambda(\cdot) \geq c > 0$ and $\lambda(\cdot)$ is Lipschitz with constant α, we have

(3.49)

$$
\begin{aligned}
\log \frac{dP^{\omega^-}}{dP^{w_0}} \bigg|_{\mathcal{F}_t^0} &= \int_0^t \lambda(w_0) - \lambda(\omega^-) ds + \int_0^t \log \left(\frac{\lambda(\omega^-)}{\lambda(w_0)} \right) dN_s \\
&\leq \int_0^t |\lambda(w_0) - \lambda(\omega^-)| ds + \int_0^t \log \left(1 + \frac{|\lambda(w_0) - \lambda(\omega^-)|}{\lambda(w_0)} \right) dN_s \\
&\leq \int_0^t \alpha \sum_{\tau \in \omega^- \cup w_0} h(s-\tau) ds + \int_0^t \frac{\alpha}{c} \sum_{\tau \in \omega^- \cup w_0} h(s-\tau) dN_s.
\end{aligned}
$$

Define

$$
(3.50) \qquad K_\ell^- = \left\{ \omega : N[-t,0](\omega) \leq \ell(1+t), \forall t > 0 \right\}.
$$

By the maximal ergodic theorem,

(3.51)

$$
\begin{aligned}
Q((K_\ell^-)^c) &= Q \left(\sup_{t>0} \frac{N[-t,0]}{t+1} > \ell \right) \\
&\leq Q \left(\sup_{t>0} \frac{N[-([t]+1),0]}{[t]+1} > \ell \right) \\
&= Q \left(\sup_{n \geq 1, n \in \mathbb{N}} \frac{N[-n,0]}{n} > \ell \right)
\end{aligned}
$$

(3.52)

$$
\leq \frac{\mathbb{E}^Q[N[0,1]]}{\ell} \to 0
$$

as $\ell \to \infty$. Thus $Q(K_\ell^-) \to 1$ as $\ell \to \infty$.

Fix any $s > 0$ and $\omega^- \in K_\ell^-$. Since h is decreasing, $h' \leq 0$, integration by parts

shows that

$$(3.53) \qquad \sum_{\tau \in \omega^-} h(s-\tau) = \int_0^\infty h(s+\sigma) dN[-\sigma, 0]$$

$$= -\int_0^\infty N[-\sigma, 0] h'(s+\sigma) d\sigma$$

$$\leq -\int_0^\infty \ell(1+\sigma) h'(s+\sigma) d\sigma$$

$$= \ell h(s) + \ell \int_0^\infty h(s+\sigma) d\sigma$$

$$= \ell h(s) + \ell H(s),$$

where $H(t) = \int_t^\infty h(s) ds$.

Therefore, uniformly for $\omega_-, w_0 \in K_\ell^-$,

$$(3.54) \qquad \int_0^t \alpha \sum_{\tau \in \omega^- \cup w_0} h(s-\tau) ds \leq 2\ell\alpha \|h\|_{L^1} + 2\ell\alpha u(t),$$

where $u(t) = \int_0^t H(s) ds$ and

$$(3.55) \qquad \int_0^t \frac{\alpha}{c} \sum_{\tau \in \omega^- \cup w_0} h(s-\tau) dN_s \leq \frac{2\ell\alpha}{c} \int_0^t (h(s) + H(s)) dN_s.$$

Define

$$(3.56) \qquad K_{\ell,t}^+ = \left\{ \omega : \frac{2\ell\alpha}{c} \int_0^t (h(s) + H(s)) dN_s \leq \ell^2 (\|h\|_{L^1} + u(t)) \right\}.$$

Then, uniformly in $t > 0$,

$$(3.57) \qquad Q((K_{\ell,t}^+)^c) \leq \frac{2\alpha \mathbb{E}^Q[N[0,1]]}{c \cdot \ell} \to 0,$$

as $\ell \to \infty$. Thus $\inf_{t>0} Q(K^+_{\ell,t}) \to 1$ as $\ell \to \infty$.

Hence, uniformly for $\omega_-, w_0 \in K^-_\ell$ and $\omega \in K^+_{\ell,t}$,

$$(3.58) \qquad \log \frac{dP^{\omega_-}}{dP^{w_0}}\bigg|_{\mathcal{F}^0_t} \leq 2\ell\alpha\|h\|_{L^1} + 2\ell\alpha u(t) + \ell^2(\|h\|_{L^1} + u(t))$$

$$(3.59) \qquad\qquad\qquad = C_1(\ell) + C_2(\ell)u(t),$$

where $C_1(\ell) = 2\ell\alpha\|h\|_{L^1} + \ell^2\|h\|_{L^1}$ and $C_2(\ell) = 2\ell\alpha + \ell^2$.

Observe that

$$(3.60) \qquad\qquad \limsup_{t\to\infty} \frac{u(t)}{t} = \limsup_{t\to\infty} \frac{1}{t}\int_0^t H(s)ds = 0.$$

Let $D_t = \{R_{t,\omega} \in N, w_t \in K^-_\ell\}$.

Uniformly for $w_0 \in K^-_{\ell,t}$,

(3.61)

$P^{w_0}(D_t)$

$\geq e^{-t(H(Q)+\epsilon)-C_1(\ell)-C_2(\ell)u(t)}$

$\cdot Q\left[D_t \cap \left\{\frac{1}{t}\log\frac{dP^{\omega_-}}{dQ^{\omega_-}}\bigg|_{\mathcal{F}^0_t} \leq H(Q)+\epsilon\right\} \cap \left\{\log\frac{dP^{\omega_-}}{dP^{w_0}}\bigg|_{\mathcal{F}^0_t} \leq C_1(\ell)+C_2(\ell)u(t)\right\}\right]$

$\geq e^{-t(H(Q)+\epsilon)-C_1(\ell)-C_2(\ell)u(t)}$

$\qquad\qquad \cdot Q\left[D_t \cap \left\{\frac{1}{t}\log\frac{dP^{\omega_-}}{dQ^{\omega_-}}\bigg|_{\mathcal{F}^0_t} \leq H(Q)+\epsilon\right\} \cap \{K^+_{\ell,t} \cap K^-_\ell\}\right].$

Since $Q \in \mathcal{M}_E(\Omega)$, by ergodic theorem,

$$(3.62) \qquad\qquad\qquad \lim_{t\to\infty} Q(R_{t,\omega} \in N) = 1,$$

and since $\psi(\omega, t) = \log \frac{dQ^\omega}{dP^\omega}\big|_{\mathcal{F}_t^0}$ satisfies,

(3.63) $\qquad \psi(\omega, t+s) = \psi(\omega, t) + \psi(\theta_t \omega, s), \quad \mathbb{E}^Q[\psi(\omega, t)] = tH(Q),$

for almost every ω^- under Q,

(3.64) $$\lim_{t\to\infty} \frac{1}{t} \log \frac{dP^{\omega^-}}{dQ^{\omega^-}}\bigg|_{\mathcal{F}_t^0} = H(Q).$$

Q is stationary, so $Q(w_t \in K_\ell^-) \geq Q(K_\ell^-) \to 1$ as $\ell \to \infty$. Also, $Q(K_{\ell,t}^+) \geq \inf_{t>0} Q(K_{\ell,t}^+) \to 1$ as $\ell \to \infty$. Remember that $\limsup_{t\to\infty} \frac{u(t)}{t} = 0$. By choosing ℓ big enough, we conclude that

(3.65) $$\liminf_{t\to\infty} \frac{1}{t} \inf_{w_0 \in K_\ell^-} \log P^{w_0}(R_{t,\omega} \in N, w_t \in K_\ell^-) \geq -H(Q) - \epsilon.$$

Since it holds for any $\epsilon > 0$, we get the desired result. $\qquad \square$

Theorem 12 (Lower Bound). *For any open set G,*

(3.66) $$\liminf_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in G) \geq - \inf_{Q \in G} H(Q).$$

Proof. It is sufficent to prove that for any $Q \in \mathcal{M}_S(\Omega)$, $H(Q) < \infty$, for any neighborhood N of Q, $\liminf_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in N) \geq -H(Q)$. Since for every invariant measure $P \in \mathcal{M}_S$, there exists a probability measure μ_P on the space $\mathcal{M}_E$ of ergodic measures such that $P = \int_{\mathcal{M}_E} Q \mu_P(dQ)$, for any $Q \in \mathcal{M}_S(\Omega)$ such that $H(Q) < \infty$, without loss of generality, we can assume that $Q = \sum_{j=1}^\ell \alpha_j Q_j$, where $\alpha_j \geq 0$, $1 \leq j \leq \ell$ and $\sum_{j=1}^\ell \alpha_j = 1$. By linearity of $H(\cdot)$, $H(Q) = \sum_{j=1}^\ell \alpha_j H(Q_j)$. Divide the interval $[0, t]$ into subintervals of length $\alpha_j t$, let t_j, $1 \leq j \leq \ell$ be the

75

right hand endpoints of these subintervals, and let $t_0 = 0$. For each Q_j, take K_M^- as in Lemma 10. We have $\min_{1 \leq j \leq \ell} Q_j(K_M^-) \to 1$, as $M \to \infty$. Choose neighborhoods N_j of Q_j, $1 \leq j \leq \ell$ such that $\bigcup_{j=1}^{\ell} \alpha_j N_j \subseteq N$. We have

$$(3.67) \qquad P^{\varnothing}(R_{t,\omega} \in N) \geq P^{\varnothing}(R_{t_1,\omega} \in N_1, w_{t_1} \in K_M^-)$$

$$\cdot \prod_{j=2}^{\ell} \inf_{w_0 \in K_{t_{j-1}-t_{j-2}}^-} P^{w_0}(R_{t_j - t_{j-1}, \omega} \in N_j, w_{t_j - t_{j-1}} \in K_M^-).$$

Now, applying Lemma 10 and the linearity of $H(\cdot)$,

$$(3.68) \qquad \liminf_{t \to \infty} \frac{1}{t} \log P^{\varnothing}(R_{t,\omega} \in N) \geq -\sum_{j=1}^{\ell} \alpha_j H(Q_j) = -H(Q).$$

$\square$

3.3 Upper Bound

Remark 3. *By following the argument in Donsker and Varadhan [31], if $\omega^- \mapsto P^{\omega^-}$ is weakly continuous, then*

$$(3.69) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in A) \leq -\inf_{Q \in A} H(Q),$$

for any compact A. If the Hawkes process has finite range of memory, i.e. $h(\cdot)$ has compact support, and if it is continuous, then, for any $a < b$, if $\omega_n^- \to \omega^-$, we

have

$$(3.70) \qquad \left| \int_a^b \lambda(\omega_n^-, \omega, s)ds - \int_a^b \lambda(\omega^-, \omega, s)ds \right|$$

$$\leq \alpha \int_a^b \left| \sum_{\tau \in \omega_n^-} h(s - \tau) - \sum_{\tau \in \omega^-} h(s - \tau) \right| ds \to \infty,$$

as $n \to \infty$, which implies that $P^{\omega_n^-} \to P^{\omega^-}$.

If the Hawkes process does not have finite range of memory, then we should use the specific features of the Hawkes process to obtain the upper bound.

Before we proceed, let us prove an easy but very useful lemma that we will use repeatedly in the proofs of the estimates in this chapter.

Lemma 11. *Let $f(\omega, s)$ be $\mathcal{F}_s^{-\infty}$ progressively measurable and predictable. Then,*

$$(3.71) \qquad \mathbb{E}\left[e^{\int_0^t f(\omega,s)dN_s} \right] \leq \mathbb{E}\left[e^{\int_0^t (e^{2f(\omega,s)}-1)\lambda(\omega,s)ds} \right]^{1/2}.$$

Proof. Since $\exp\left\{ \int_0^t 2f(\omega, s)dN_s - \int_0^t (e^{2f(\omega,s)} - 1)\lambda(\omega, s)ds \right\}$ is a martingale, by Cauchy-Schwarz inequality,

(3.72)
$$\mathbb{E}\left[e^{\int_0^t f(\omega,s)dN_s} \right] = \mathbb{E}\left[e^{\frac{1}{2}\int_0^t 2f(\omega,s)dN_s - \frac{1}{2}\int_0^t (e^{2f(\omega,s)}-1)\lambda(\omega,s)ds + \frac{1}{2}\int_0^t (e^{2f(\omega,s)}-1)\lambda(\omega,s)ds} \right]$$

$$\leq \mathbb{E}\left[e^{\int_0^t (e^{2f(\omega,s)}-1)\lambda(\omega,s)ds} \right]^{1/2}.$$

$\square$

Define $\mathcal{C}_T$

$$(3.73) \qquad \mathcal{C}_T = \left\{ F(\omega) := \int_0^T f(\omega, s) dN_s - \int_0^T (e^{f(\omega, s)} - 1) \lambda(\omega, s) ds, \right.$$

$$\left. f(\omega, s) \in \mathcal{B}(\mathcal{F}_s^0) \cap C(\Omega \times \mathbb{R}) \right\}.$$

Here $\lambda(\omega, s)$ is $\mathcal{F}_s^{-\infty}$ progressively measurable and predictable, and $f(\omega, s) \in \mathcal{B}(\mathcal{F}_s^0) \cap C(\Omega \times \mathbb{R})$ means that f is $\mathcal{F}_s^0$ progressively measurable, predictable and also bounded and continuous.

Lemma 12. *For any $T > 0$ and $F \in \mathcal{C}_T$, we have, for any $t > 0$,*

$$(3.74) \qquad \mathbb{E}^{P^\varnothing} \left[e^{\frac{1}{T} \int_0^t F(\theta_s \omega) ds} \right] \leq 1.$$

Proof. For any $t > 0$, writing $\psi(s) = \sum_{k : s+kT \leq t} F(\theta_{s+kT} \omega)$,

$$(3.75) \qquad \mathbb{E}^{P^\varnothing} \left[e^{\frac{1}{T} \int_0^t F(\theta_s \omega) ds} \right] = \mathbb{E}^{P^\varnothing} \left[e^{\frac{1}{T} \int_0^T \psi(s) ds} \right]$$

$$\leq \frac{1}{T} \int_0^T \mathbb{E}^{P^\varnothing} \left[e^{\psi(s)} \right] ds = 1,$$

by Jensen's inequality and the fact that $\mathbb{E}^{P^\varnothing} \left[e^{\psi(s)} \right] = 1$ by iteratively conditioning since $\mathbb{E}^{P^{\omega^-}} \left[e^{F(\omega)} \right] = 1$ for any ω^-. $\qquad \square$

Remark 4. *Under $P^\varnothing$, the $\mathcal{F}_s^{-\infty}$ progressively measurable rate function λ is well defined since it only creates a history between time 0 and time t. Similarly, in the proof in Lemma 12, $\mathbb{E}^{P^{\omega^-}} \left[e^{F(\omega)} \right] = 1$ for any ω^- should be interpreted as the expectation is 1 given any history created between time 0 and t, which is well defined.*

Next, we need to compare $\frac{1}{T} \int_0^t F(\theta_s \omega_t) ds$ and $\frac{1}{T} \int_0^t F(\theta_s \omega) ds$.

Lemma 13. *For any $q > 0$, $T > 0$ and $F \in \mathcal{C}_T$,*

$$(3.76) \qquad \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[\exp \left\{ q \left| \frac{1}{T} \int_0^t F(\theta_s \omega_t) ds - \frac{1}{T} \int_0^t F(\theta_s \omega) ds \right| \right\} \right] = 0.$$

Proof.

$$(3.77) \qquad \left| \frac{1}{T} \int_0^t F(\theta_s \omega_t) ds - \frac{1}{T} \int_0^t F(\theta_s \omega) ds \right|$$

$$\leq \left| \frac{1}{T} \int_0^t \int_0^T f(u, \theta_s \omega) dN_u ds - \frac{1}{T} \int_0^t \int_0^T f(u, \theta_s \omega_t) dN_u ds \right|$$

$$+ \left| \frac{1}{T} \int_0^t \int_0^T (e^{f(u, \theta_s \omega)} - 1) \lambda(\theta_s \omega, u) du ds \right.$$

$$\left. - \frac{1}{T} \int_0^t \int_0^T (e^{f(u, \theta_s \omega_t)} - 1) \lambda(\theta_s \omega_t, u) du ds \right|.$$

It is easy to see that $\int_0^T f(u, \theta_s \omega) dN_u ds$ is $\mathcal{F}_{s+T}^s$-measurable and

$$(3.78) \qquad \int_0^T f(u, \theta_s \omega) dN_u ds = \int_0^T f(u, \theta_s \omega_t) dN_u ds$$

for any $0 \leq s \leq t - T$. Hence,

$$(3.79) \qquad \left| \frac{1}{T} \int_0^t \int_0^T f(u, \theta_s \omega) dN_u ds - \frac{1}{T} \int_0^t \int_0^T f(u, \theta_s \omega_t) dN_u ds \right|$$

$$\leq \frac{1}{T} \int_{t-T}^t \int_0^T |f(u, \theta_s \omega)| dN_u ds + \frac{1}{T} \int_{t-T}^t \int_0^T |f(u, \theta_s \omega_t)| dN_u ds$$

$$\leq \frac{\|f\|_{L^\infty}}{T} \int_{t-T}^t N[s, s+T](\omega) ds + \frac{\|f\|_{L^\infty}}{T} \int_{t-T}^t N[s, s+T](\omega_t) ds$$

$$\leq \frac{\|f\|_{L^\infty}}{T} [N[t-T, t+T](\omega) + N[t-T, t+T](\omega_t)]$$

$$= \frac{\|f\|_{L^\infty}}{T} [N[t-T, t+T](\omega) + N[t-T, T](\omega) + N[0, T](\omega)].$$

By Hölder's inequality and Lemma 17, we have

$$(3.80) \qquad \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{\left| \frac{1}{T} \int_0^t \int_0^T f(u, \theta_s \omega) dN_u ds - \frac{1}{T} \int_0^t \int_0^T f(u, \theta_s \omega_t) dN_u ds \right|} \right]$$

$$\leq \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{\frac{\|f\|_{L^\infty}}{T} [N[t-T, t+T](\omega) + N[t-T, T](\omega) + N[0, T](\omega)]} \right] = 0.$$

Furthermore,

$$(3.81)$$
$$\left| \frac{1}{T} \int_0^t \int_0^T (e^{f(u, \theta_s \omega)} - 1) \lambda(\theta_s \omega, u) du ds - \frac{1}{T} \int_0^t \int_0^T (e^{f(u, \theta_s \omega_t)} - 1) \lambda(\theta_s \omega_t, u) du ds \right|$$

$$\leq \frac{1}{T} \int_0^t \int_0^T \left| e^{f(u, \theta_s \omega)} - e^{f(u, \theta_s \omega_t)} \right| \lambda(\theta_s \omega, u) du ds$$

$$+ \frac{1}{T} \int_0^t \int_0^T (e^{f(u, \theta_s \omega_t)} - 1) \left| \lambda(\theta_s \omega_t, u) - \lambda(\theta_s \omega, u) \right| du ds.$$

For the first term

$$(3.82) \qquad \frac{1}{T} \int_0^t \int_0^T \left| e^{f(u,\theta_s\omega)} - e^{f(u,\theta_s\omega_t)} \right| \lambda(\theta_s\omega, u) du\, ds$$

$$= \frac{1}{T} \int_{t-T}^t \int_0^T \left| e^{f(u,\theta_s\omega)} - e^{f(u,\theta_s\omega_t)} \right| \lambda(\theta_s\omega, u) du\, ds$$

$$\leq \frac{2e^{\|f\|_{L^\infty}}}{T} \int_{t-T}^t \int_0^T \lambda(\theta_s\omega, u) du\, ds$$

$$= \frac{2e^{\|f\|_{L^\infty}}}{T} \int_{t-T}^t \int_0^T \lambda\left(\sum_{\tau \in \omega[0,u+s)} h(u+s-\tau) \right) du\, ds$$

$$\leq 2e^{\|f\|_{L^\infty}} T C_\epsilon + \frac{2e^{\|f\|_{L^\infty}}}{T} \epsilon \int_{t-T}^t \int_0^T \sum_{\tau \in \omega[0,u+s)} h(u+s-\tau) du\, ds$$

$$\leq 2e^{\|f\|_{L^\infty}} T C_\epsilon + \frac{2e^{\|f\|_{L^\infty}}}{T} \epsilon \int_{t-T}^t \int_0^T N[0, u+s] h(0) du\, ds$$

$$\leq 2e^{\|f\|_{L^\infty}} T C_\epsilon + 2e^{\|f\|_{L^\infty}} \epsilon \int_{t-T}^t N[0, s+T] h(0) ds$$

$$\leq 2e^{\|f\|_{L^\infty}} T C_\epsilon + 2e^{\|f\|_{L^\infty}} \epsilon T (N[0, t] + N[t, t+T]) h(0).$$

Therefore,

$$(3.83) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{\frac{1}{T} \int_0^t \int_0^T \left| e^{f(u,\theta_s\omega)} - e^{f(u,\theta_s\omega_t)} \right| \lambda(\theta_s\omega,u) du\, ds} \right] \leq c(\epsilon),$$

where $c(\epsilon) \to 0$ as $\epsilon \to 0$; in other words, it vanishes.

81

For the second term,

(3.84)

$$\frac{1}{T} \int_0^t \int_0^T (e^{f(u,\theta_s \omega_t)} - 1) \, |\lambda(\theta_s \omega_t, u) - \lambda(\theta_s \omega, u)| \, duds$$

$$\leq \frac{e^{\|f\|_{L^\infty}} + 1}{T}$$

$$\cdot \int_0^t \int_0^T \alpha \left| \sum_{\tau \in \omega_t [0,u+s) \cup (\omega_t)^-} h(u+s-\tau) - \sum_{\tau \in \omega [0,u+s)} h(u+s-\tau) \right| duds$$

$$\leq \frac{e^{\|f\|_{L^\infty}} + 1}{T} \int_{t-T}^t \int_0^T \alpha \sum_{\tau \in \omega_t [0,u+s)} h(u+s-\tau) duds$$

$$+ \frac{e^{\|f\|_{L^\infty}} + 1}{T} \int_{t-T}^t \int_0^T \alpha \sum_{\tau \in \omega [0,u+s)} h(u+s-\tau) duds$$

$$+ \frac{e^{\|f\|_{L^\infty}} + 1}{T} \int_0^t \int_0^T \alpha \sum_{\tau \in (\omega_t)^-} h(u+s-\tau) duds$$

Assume that $h(\cdot)$ is decreasing and $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$. By applying Jensen's inequality twice, we can estimate the second term above,

(3.85)
$$\mathbb{E}^{P^\varnothing} \left[e^{\frac{e^{\|f\|_{L^\infty}}+1}{T} \alpha \int_{t-T}^t \int_0^T \int_0^{u+s} h(u+s-v) dN_v duds} \right]$$

$$\leq \frac{1}{T} \int_{t-T}^t \mathbb{E}^{P^\varnothing} \left[e^{(e^{\|f\|_{L^\infty}}+1)\alpha \int_0^T \int_0^{u+s} h(u+s-v) dN_v du} \right] ds$$

$$\leq \frac{1}{T^2} \int_{t-T}^t \int_0^T \mathbb{E}^{P^\varnothing} \left[e^{(e^{\|f\|_{L^\infty}}+1)\alpha T \int_0^{u+s} h(u+s-v) dN_v} \right] duds$$

$$\leq \frac{1}{T^2} \int_{t-T}^t \int_0^T \mathbb{E}^{P^\varnothing} \left[e^{\int_0^{u+s} C(\alpha,T,h)\lambda(v) dv} \right]^{1/2} duds$$

$$\leq \frac{e^{C(\alpha,T,h)C_\epsilon}}{T^2} \int_{t-T}^t \int_0^T \mathbb{E}^{P^\varnothing} \left[e^{\epsilon C(\alpha,T,h)N[0,u+s]h(0)} \right]^{1/2} duds$$

$$\leq e^{C(\alpha,T,h)C_\epsilon} \mathbb{E}^{P^\varnothing} \left[e^{\epsilon C(\alpha,T,h)N[0,t+T]h(0)} \right]^{1/2}.$$

where $C(\alpha, T, h) = \exp(2(e^{\|f\|_{L^\infty}} + 1)\alpha T h(0)) - 1$. Thus,

$$(3.86) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{\frac{e^{\|f\|_{L^\infty}} + 1}{T} \alpha \int_{t-T}^t \int_0^T \int_0^{u+s} h(u+s-v)dN_v du ds} \right] = 0.$$

Similarly, we can estimate the first term.

For the third term, by Jensen's inequality, we have

$$(3.87) \qquad \mathbb{E}^{P^\varnothing} \left[e^{\frac{e^{\|f\|_{L^\infty}} + 1}{T} \int_0^t \int_0^T \alpha \sum_{\tau \in (\omega_t)^-} h(u+s-\tau)du ds} \right]$$

$$\leq \frac{1}{T} \int_0^T \mathbb{E}^{P^\varnothing} \left[e^{\alpha(\exp(\|f\|_{L^\infty})+1) \int_0^t \sum_{\tau \in (\omega_t)^-} h(u+s-\tau)ds} \right] du$$

$$\leq \mathbb{E}^{P^\varnothing} \left[e^{\alpha(\exp(\|f\|_{L^\infty})+1) \int_0^t \sum_{\tau \in (\omega_t)^-} h(s-\tau)ds} \right]$$

$$= \mathbb{E}^{P^\varnothing} \left[e^{\alpha(\exp(\|f\|_{L^\infty})+1) \int_0^t \int_0^t \sum_{k=0}^\infty h(s+kt+t-u)dN_u ds} \right]$$

Since $h(\cdot)$ is decreasing, $\int_{kt}^{(k+1)t} h(s)ds \geq th((k+1)t)$. Thus

$$(3.88) \qquad \sum_{k=0}^\infty h(s+kt+t-u) \leq h(s+t-u) + \frac{1}{t} \int_{s+t-u}^\infty h(v)dv.$$

Let $C(\alpha, f) = \alpha(\exp(\|f\|_{L^\infty}) + 1)$ and $H(t) = \int_t^\infty h(s)ds$. Then,

$$(3.89) \qquad \mathbb{E}^{P^\varnothing} \left[e^{\alpha(\exp(\|f\|_{L^\infty})+1) \int_0^t \int_0^t \sum_{k=0}^\infty h(s+kt+t-u)dN_u ds} \right]$$

$$\leq \mathbb{E}^{P^\varnothing} \left[e^{C(\alpha,f) \int_0^t \int_0^t \frac{1}{t} H(s+t-u)dN_u ds + C(\alpha,f) \int_0^t [\int_0^t h(s+t-u)ds]dN_u} \right]$$

$$= \mathbb{E}^{P^\varnothing} \left[e^{\int_0^t [\frac{C(\alpha,f)}{t} \int_0^t H(s+t-u)ds]dN_u + \int_0^t [\int_0^t C(\alpha,f)h(s+t-u)ds]dN_u} \right].$$

Notice that

$$(3.90) \qquad \mathbb{E}^{P^\varnothing} \left[e^{\int_0^t [\frac{C(\alpha,f)}{t} \int_0^t H(s+t-u)ds]dN_u} \right] \leq \mathbb{E}^{P^\varnothing} \left[e^{[\frac{C(\alpha,f)}{t} \int_0^t H(s)ds]N_t} \right],$$

83

where $\frac{C(\alpha,f)}{t}\int_0^t H(s)ds \to 0$ as $t \to \infty$, which implies that

$$(3.91) \qquad \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}^{P^\varnothing}\left[e^{\int_0^t\left[\frac{C(\alpha,f)}{t}\int_0^t H(s+u)ds\right]dN_u}\right] = 0.$$

Moreover,

(3.92)

$$\mathbb{E}^{P^\varnothing}\left[e^{\int_0^t\left[\int_0^t C(\alpha,f)h(s+t-u)ds\right]dN_u}\right]$$

$$\leq \mathbb{E}^{P^\varnothing}\left[e^{\int_0^t(e^{2\int_0^t C(\alpha,f)h(s+t-u)ds}-1)\lambda(u)du}\right]^{1/2}$$

$$\leq e^{\frac{1}{2}C_\epsilon\int_0^t(e^{2\int_0^t C(\alpha,f)h(s+t-u)ds}-1)du}\mathbb{E}^{P^\varnothing}\left[e^{\int_0^t(e^{\int_0^t 2C(\alpha,f)h(s+t-u)ds}-1)\epsilon\sum_{\tau<u}h(u-\tau)du}\right]^{1/2}$$

$$\leq e^{\frac{1}{2}C_\epsilon\int_0^t(e^{2C(\alpha,f)H(t-u)}-1)du}\mathbb{E}^{P^\varnothing}\left[e^{\int_0^t(e^{2C(\alpha,f)\|h\|_{L^1}}-1)\epsilon\sum_{\tau<u}h(u-\tau)du}\right]^{1/2}$$

$$\leq e^{\frac{1}{2}C_\epsilon\int_0^t(e^{2C(\alpha,f)H(u)}-1)du}\mathbb{E}^{P^\varnothing}\left[e^{\epsilon(e^{2C(\alpha,f)\|h\|_{L^1}}-1)\|h\|_{L^1}N_t}\right]^{1/2}$$

Notice that it holds for any $\epsilon > 0$ and that $\frac{1}{t}\int_0^t(e^{2C(\alpha,f)H(u)}-1)du \to 0$ as $t \to \infty$, which implies

$$(3.93) \qquad \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}^{P^\varnothing}\left[e^{\int_0^t\left[\int_0^t C(\alpha,f)h(s+t-u)ds\right]dN_u}\right] = 0.$$

Putting all these things together and applying Hölder's inequality several times, we find that for any $q > 0$, $T > 0$ and $F \in \mathcal{C}_T$,

$$(3.94) \qquad \limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}^{P^\varnothing}\left[\exp\left\{q\left|\frac{1}{T}\int_0^t F(\theta_s\omega_t)ds - \frac{1}{T}\int_0^t F(\theta_s\omega)ds\right|\right\}\right] = 0.$$

$\square$

Lemma 14.

$$(3.95) \qquad \lim_{T \to \infty} \frac{1}{T} \sup_{F \in \mathcal{C}_T} \int_{\Omega} F(\omega) Q(d\omega) \geq H(Q).$$

Proof. Assume $H(Q) < \infty$. For any $\epsilon > 0$, there exists some f_ϵ such that

$$(3.96) \qquad \mathbb{E}^Q \left[\int_0^1 f_\epsilon dN_s - \int_0^1 (e^{f_\epsilon} - 1)\lambda ds \right] \geq H(Q) - \epsilon.$$

We can find a sequence $f_T \in \mathcal{B}\left(\mathcal{F}_s^{-(T-1)}\right) \cap C(\Omega \times \mathbb{R}) \to f_\epsilon$ as $T \to \infty$. By Fatou's lemma,

$$(3.97) \qquad \liminf_{T \to \infty} \frac{1}{T} \sup_{F \in \mathcal{C}_T} \int_{\Omega} F(\omega) Q(d\omega)$$
$$\geq \liminf_{T \to \infty} \mathbb{E}^Q \left[\int_0^1 f_T dN_s - \int_0^1 (e^{f_T} - 1)\lambda ds \right] \geq H(Q) - \epsilon.$$

If $H(Q) = \infty$, then, for any $M > 0$, there exists some f_M such that

$$(3.98) \qquad \mathbb{E}^Q \left[\int_0^1 f_M dN_s - \int_0^1 (e^{f_M} - 1)\lambda ds \right] \geq M.$$

Repeat the same argument as in the case that $H(Q) < \infty$. $\qquad \square$

Lemma 15. *For any compact set A,*

$$(3.99) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in A) \leq - \inf_{Q \in A} H(Q).$$

Proof. Notice that

(3.100)
$$\mathbb{E}^{P^\varnothing} \left[e^{N[0,t]} \right] \leq \mathbb{E}^{P^\varnothing} \left[e^{(e^2-1) \int_0^t \lambda(s) ds} \right]^{1/2} \leq \mathbb{E}^{P^\varnothing} \left[e^{(e^2-1)\epsilon h(0) N[0,t] + C_\epsilon (e^2-1)} \right]^{1/2}.$$

By choosing $\epsilon > 0$ small enough, we have $\mathbb{E}^{P^{\varnothing}}[e^{N[0,t]}] \leq e^{Ct}$ for some constant $C > 0$. Therefore

$$(3.101) \qquad \limsup_{\ell \to \infty} \limsup_{t \to \infty} \frac{1}{t} \log P^{\varnothing}\left(N[0,t] > \ell t\right) = -\infty,$$

which implies (by comparing $\int_{\Omega} N[0,1]dR_{t,\omega}$ and $N[0,t]/t$ and the superexponential estimates in Lemma 17)

$$(3.102) \qquad \limsup_{\ell \to \infty} \limsup_{t \to \infty} \frac{1}{t} \log P^{\varnothing}\left(\int_{\Omega} N[0,1]dR_{t,\omega} > \ell\right) = -\infty.$$

Therefore, we need only to consider compact sets A such that for any $Q \in A$, $\mathbb{E}^Q[N[0,1]] < \infty$.

Now for any A compact consisting of Q with $\mathbb{E}^Q[N[0,1]] < \infty$ and for any $F \in \mathcal{C}_T$ and for any $p,q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, by Hölder's inequality, Chebychev's inequality, and Lemma 12,

$$(3.103) \quad P^{\varnothing}(R_{t,\omega} \in A)$$

$$\leq \mathbb{E}^{P^{\varnothing}}\left[e^{\frac{1}{pT}\int_0^t F(\theta_s \omega_t)ds}\right] \cdot \exp\left\{-\frac{t}{pT} \inf_{Q \in A} \int_{\Omega} F(\omega)Q(d\omega)\right\}$$

$$\leq \mathbb{E}^{P^{\varnothing}}\left[e^{\frac{1}{T}\int_0^t F(\theta_s \omega)ds}\right]^{1/p} \mathbb{E}^{P^{\varnothing}}\left[e^{\frac{q}{pT}\left|\int_0^t F(\theta_s \omega_t)ds - \int_0^t F(\theta_s \omega)ds\right|}\right]^{1/q}$$

$$\cdot \exp\left\{-\frac{t}{pT} \inf_{Q \in A} \int_{\Omega} F(\omega)Q(d\omega)\right\}$$

$$\leq \mathbb{E}^{P^{\varnothing}}\left[e^{\frac{q}{pT}\left|\int_0^t F(\theta_s \omega_t)ds - \int_0^t F(\theta_s \omega)ds\right|}\right]^{1/q} \cdot \exp\left\{-\frac{t}{pT} \inf_{Q \in A} \int_{\Omega} F(\omega)Q(d\omega)\right\}$$

By Lemma 13,

$$(3.104) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P^{\varnothing}(R_{t,\omega} \in A) \leq -\frac{1}{p} \inf_{Q \in A} \frac{1}{T} \int_{\Omega} F(\omega)Q(d\omega).$$

Since it holds for any $p > 1$, we get

$$(3.105) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P^{\varnothing}(R_{t,\omega} \in A) \leq - \inf_{Q \in A} \frac{1}{T} \int_{\Omega} F(\omega) Q(d\omega).$$

For any compact A, given $Q \in A$ and $\epsilon > 0$, by Lemma 14, there exists $T_Q > 0$ and $F_Q \in \mathcal{C}_{T_Q}$ such that $\frac{1}{T_Q} \int_{\Omega} F_Q(\omega) Q(d\omega) \geq \inf_{A \in Q} H(Q) - \frac{1}{2}\epsilon$. Since the linear integral is a continuous functional of Q (see the proof of Lemma 8), there exists a neighborhood G_Q of Q such that $\frac{1}{T_Q} \int_{\Omega} F_Q(\omega) Q(d\omega) \geq \inf_{A \in Q} H(Q) - \epsilon$ for all $Q \in G_Q$. Since A is compact, there exists $G_{Q_1}, \ldots, G_{Q_\ell}$ such that $A \subset \bigcup_{j=1}^{\ell} G_{Q_j}$. Hence

$$(3.106) \qquad \inf_{1 \leq j \leq \ell} \sup_{T > 0} \sup_{F \in \mathcal{C}_T} \inf_{Q \in G_j} \frac{1}{T} \int_{\Omega} F(\omega) Q(d\omega) \geq \inf_{Q \in A} H(Q) - \epsilon.$$

Note that for any A and B,

$$(3.107) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in A \cup B)$$
$$\leq \max \left\{ \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in A), \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in B) \right\}.$$

Thus, for $A \subset \bigcup_{j=1}^{\ell} G_j$,

$$(3.108) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in A) \leq - \inf_{1 \leq j \leq \ell} \sup_{T > 0} \sup_{F \in \mathcal{C}_T} \inf_{Q \in G_j} \frac{1}{T} \int F(\omega) Q(d\omega),$$

whence $\limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in A) \leq - \inf_{Q \in A} H(Q)$ for any compact A. $\qquad \square$

Theorem 13 (Upper Bound). *For any closed set C,*

$$(3.109) \qquad \limsup_{t \to \infty} \frac{1}{t} \log P(R_{t,\omega} \in C) \leq - \inf_{Q \in C} H(Q).$$

87

Proof. For any closed set C and compact $\mathcal{A}^n$ which is defined in Lemma 22, we have

(3.110)

$$\limsup_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in C)$$

$$\leq \max \left\{ \limsup_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in C \cap \mathcal{A}^n), \limsup_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in (\mathcal{A}^n)^c) \right\}.$$

Since $C \cap \mathcal{A}_n$ is compact, Lemma 15 implies

(3.111) $$\limsup_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in C \cap \mathcal{A}^n) \leq - \inf_{Q \in C \cap \mathcal{A}^n} H(Q) \leq - \inf_{Q \in C} H(Q).$$

Furthermore, by Lemma 21,

(3.112) $$\limsup_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in (\mathcal{A}^n)^c)$$

$$= \limsup_{t\to\infty} \frac{1}{t} \log P\left(R_{t,\omega} \in \bigcup_{j=n}^{\infty} \mathcal{A}_{\frac{1}{j},j,j}^c \right)$$

$$\leq \max_{j\geq n} \max \left\{ \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t} \int_0^t \chi_{N[0,1]\geq j}(\theta_s \omega_t) ds \geq \varepsilon(j) \right), \right.$$

$$\limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t} \int_0^t \chi_{N[0,1/j]\geq 2}(\theta_s \omega_t) ds \geq (1/j)g(1/j) \right),$$

$$\left. \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t} \int_0^t N[0,1]\chi_{N[0,1]\geq \ell}(\theta_s \omega_t) ds \geq m(\ell) \right) \right\} \to -\infty$$

as $n \to \infty$. Hence,

(3.113) $$\limsup_{t\to\infty} \frac{1}{t} \log P(R_{t,\omega} \in C) \leq - \inf_{Q \in C} H(Q).$$

$\square$

88

3.4 Superexponential Estimates

In order to get the full large deviation principle, we need the upper bound inequality valid for any closed set instead of for any compact set, which requires some superexponential estimates.

Lemma 16. *For any $q > 0$,*

$$(3.114) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{q \int_0^t h(t-s)dN_s} \right] = 0.$$

Proof.

$$(3.115) \qquad \mathbb{E}^{P^\varnothing} \left[e^{q \int_0^t h(t-s)dN_s} \right] \leq \mathbb{E}^{P^\varnothing} \left[e^{\int_0^t (e^{2qh(t-s)}-1)\lambda(\sum_{0<\tau<s} h(s-\tau))ds} \right]^{1/2}$$

$$\leq \mathbb{E}^{P^\varnothing} \left[e^{(C_\epsilon + h(0)\epsilon N_t) \int_0^t (e^{2qh(t-s)}-1)ds} \right]^{1/2}.$$

Note that $\int_0^t (e^{2qh(t-s)} - 1)ds = \int_0^t (e^{2qh(s)} - 1)ds \in L_1$ since $h \in L^1$. Therefore,

$$(3.116) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{q \int_0^t h(t-s)dN_s} \right] \leq c(\epsilon),$$

where $c(\epsilon) \to 0$ as $\epsilon \to 0$. Since it holds for any ϵ, we get the desired result. $\qquad \square$

Lemma 17. *For any $q > 0$ and $T > 0$,*

$$(3.117) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}^{P^\varnothing} \left[e^{qN[t,t+T]} \right] = 0.$$

Therefore, for any $\epsilon > 0$,

$$(3.118) \qquad \limsup_{t\to\infty} \frac{1}{t} \log P^\varnothing \left(N[t, t+T] \geq \epsilon t \right) = -\infty.$$

Proof. By Hölder's inequality,

(3.119)

$$\mathbb{E}^{P^{\varnothing}}\left[e^{qN[t,t+T]}\right] \leq \mathbb{E}^{P^{\varnothing}}\left[e^{(e^{2q}-1)\int_t^{t+T}\lambda(\sum_{0<\tau<s}h(s-\tau))ds}\right]^{1/2}$$

$$\leq e^{\frac{1}{2}(e^{2q}-1)C_{\epsilon}T} \cdot \mathbb{E}^{P^{\varnothing}}\left[e^{\epsilon(e^{2q}-1)h(0)N[t,t+T]+\epsilon(e^{2q}-1)\int_0^t h(t-s)dN_s}\right]^{1/2}$$

$$\leq e^{\frac{1}{2}(e^{2q}-1)C_{\epsilon}T} \cdot \mathbb{E}^{P^{\varnothing}}\left[e^{2\epsilon(e^{2q}-1)h(0)N[t,t+T]}\right]^{1/4}\mathbb{E}^{P^{\varnothing}}$$

$$\cdot \left[e^{2\epsilon(e^{2q}-1)\int_0^t h(t-s)dN_s}\right]^{1/4}.$$

Choose $\epsilon < q[2(e^{2q}-1)h(0)]^{-1}$. Then

(3.120) $$\mathbb{E}^{P^{\varnothing}}\left[e^{qN[t,t+T]}\right]^{3/4} \leq e^{\frac{1}{2}(e^{2q}-1)C_{\epsilon}T} \cdot \mathbb{E}^{P^{\varnothing}}\left[e^{2\epsilon(e^{2q}-1)\int_0^t h(t-s)dN_s}\right]^{1/4}.$$

Lemma 16 completes the proof. $\square$

Lemma 18. *We have the following superexponential estimates.*

(i) For any $\epsilon > 0$,

(3.121) $$\limsup_{\delta \to 0}\limsup_{t \to \infty}\frac{1}{t}\log P\left(\frac{1}{\delta t}\int_0^t \chi_{N[0,\delta]\geq 2}(\theta_s\omega)ds \geq \epsilon\right) = -\infty.$$

(ii) For any $\epsilon > 0$,

(3.122) $$\limsup_{M\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log P\left(\frac{1}{t}\int_0^t \chi_{N[0,1]\geq M}(\theta_s\omega)ds \geq \epsilon\right) = -\infty.$$

(iii) For any $\epsilon > 0$,

(3.123) $$\limsup_{\ell\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log P\left(\frac{1}{t}\int_0^t N[0,1]\chi_{N[0,1]\geq\ell}(\theta_s\omega)ds \geq \epsilon\right) = -\infty.$$

90

Proof. (i) Define

(3.124) $$N_{\ell'}[0,t] = \int_0^t \chi_{\lambda(s)<\ell'} dN_s, \quad \hat{N}_{\ell'}[0,t] = \int_0^t \chi_{\lambda(s)\geq\ell'} dN_s.$$

Then $N[0,t] = N_{\ell'}[0,t] + \hat{N}_{\ell'}[0,t]$ and $N_{\ell'}[0,t]$ has compensator $\int_0^t \lambda(s)\chi_{\lambda(s)<\ell'} ds$ and $\hat{N}_{\ell'}[0,t]$ has compensator $\int_0^t \lambda(s)\chi_{\lambda(s)\geq\ell'} ds$. Notice that

(3.125) $$\chi_{N[0,\delta]\geq 2} \leq \chi_{N_{\ell'}[0,\delta]\geq 2} + \chi_{\hat{N}_{\ell'}[0,\delta]\geq 1}.$$

It is clear that $N_{\ell'}$ is dominated by the usual Poisson process with rate ℓ'. By Lemma 19,

(3.126) $$\limsup_{\delta\to 0} \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{\delta t} \int_0^t \chi_{N_{\ell'}[0,\delta]\geq 2}(\theta_s\omega)ds \geq \frac{\epsilon}{2}\right) = -\infty.$$

On the other hand,

(3.127) $$\begin{aligned}
\frac{1}{\delta} \int_0^t \chi_{\hat{N}_{\ell'}[0,\delta]\geq 1}(\theta_s\omega)ds &= \frac{1}{\delta} \int_0^t \chi_{\hat{N}_{\ell'}[s,s+\delta]\geq 1}(\omega)ds \\
&\leq \frac{1}{\delta} \int_0^t \hat{N}_{\ell'}[s,s+\delta]ds \\
&= \frac{1}{\delta} \int_\delta^{t+\delta} \hat{N}_{\ell'}[0,s]ds - \frac{1}{\delta} \int_0^t \hat{N}_{\ell'}[0,s]ds \\
&\leq \hat{N}_{\ell'}[0,t] + N[t,t+\delta].
\end{aligned}$$

By Lemma 17, we have

(3.128) $$\limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t}N[t,t+\delta] \geq \frac{\epsilon}{4}\right) = -\infty,$$

91

for any $\delta > 0$. Hence

$$(3.129) \qquad \limsup_{\delta \to 0} \limsup_{t \to \infty} \frac{1}{t} \log P\left(\frac{1}{t} N[t, t+\delta] \geq \frac{\epsilon}{4}\right) = -\infty.$$

Finally, for some positive $h(\ell')$ to be chosen later,

$$(3.130) \qquad P\left(\frac{1}{t}\hat{N}_{\ell'}[0, t] \geq \frac{\epsilon}{4}\right) \leq \mathbb{E}\left[e^{h(\ell')\hat{N}_{\ell'}[0,t]}\right] e^{-th(\ell')\epsilon/4}$$

$$\leq \mathbb{E}\left[e^{(e^{2h(\ell')}-1)\int_0^t \lambda(s)\chi_{\lambda(s)\geq \ell'}ds}\right]^{1/2} e^{-th(\ell')\epsilon/4}.$$

Let $f(z) = \frac{z}{\lambda(z)}$. Then $f(z) \to \infty$ as $z \to \infty$. Let $Z_s = \sum_{\tau \in \omega[0,s]} h(s-\tau)$. Then, by the definition of $\lambda(s)$ and abusing the notation a little bit, we see that $\lambda(s) = \lambda(Z_s)$. Since $\lambda(\cdot)$ is increasing, its inverse function λ^{-1} exists and $\lambda^{-1}(\ell') \to \infty$ as $\ell' \to \infty$. We have

$$(3.131) \quad \mathbb{E}\left[e^{(e^{2h(\ell')}-1)\int_0^t \lambda(s)\chi_{\lambda(s)\geq \ell'}ds}\right]^{1/2} \leq \mathbb{E}\left[e^{(e^{2h(\ell')}-1)\int_0^t \lambda(Z_s)\chi_{Z_s\geq \lambda^{-1}(\ell')}ds}\right]^{1/2}$$

$$\leq \mathbb{E}\left[e^{(e^{2h(\ell')}-1)\int_0^t \lambda(Z_s)\frac{f(Z_s)}{\inf_{z\geq \ell'} f(\lambda^{-1}(z))}ds}\right]^{1/2}.$$

It is clear that $\lim_{\ell' \to \infty} \inf_{z \geq \ell'} f(\lambda^{-1}(z)) = \infty$. Choose

$$(3.132) \qquad h(\ell') = \frac{1}{2}\log\left[\inf_{z\geq \ell'} f(\lambda^{-1}(z)) + 1\right].$$

Then $h(\ell') \to \infty$ as $\ell' \to \infty$ and

$$(3.133) \qquad \mathbb{E}\left[e^{(e^{2h(\ell')}-1)\int_0^t \lambda(Z_s)\frac{f(Z_s)}{\inf_{z\geq\ell'} f(\lambda^{-1}(z))}ds}\right]^{1/2} = \mathbb{E}\left[e^{\int_0^t Z_s ds}\right]^{1/2}$$

$$= \mathbb{E}\left[e^{\int_0^t \sum_{\tau\in\omega[0,s]} h(s-\tau)ds}\right]^{1/2}$$

$$\leq \mathbb{E}\left[e^{\|h\|_{L^1} N_t}\right]^{1/2}.$$

Hence,

$$(3.134) \qquad \limsup_{\ell'\to\infty}\limsup_{\delta\to 0}\limsup_{t\to\infty} \frac{1}{t}\log P\left(\frac{1}{t}\hat{N}_{\ell'}[0,t]\geq\frac{\epsilon}{4}\right) = -\infty.$$

(ii) It is easy to see that (iii) implies (ii).

(iii) Observe first that

$$(3.135)$$

$$N[s,s+1]\chi_{N[s,s+1]\geq\ell} \leq N_{\ell'}[s,s+1]\chi_{N_{\ell'}[s,s+1]\geq\frac{\ell}{2}} + \hat{N}_{\ell'}[s,s+1]\chi_{\hat{N}_{\ell'}[s,s+1]\geq\frac{\ell}{2}}$$

$$+ \frac{\ell}{2}\chi_{N_{\ell'}[s,s+1]\geq\frac{\ell}{2}} + \frac{\ell}{2}\chi_{\hat{N}_{\ell'}[s,s+1]\geq\frac{\ell}{2}}.$$

For the first term, notice that $N_{\ell'}$ is dominated by a usual Poisson process with rate ℓ'. Thus, by Lemma 20,

$$(3.136)$$

$$\limsup_{\ell\to\infty}\limsup_{t\to\infty} \frac{1}{t}\log P\left(\frac{1}{t}\int_0^t N_{\ell'}[s,s+1]\chi_{N_{\ell'}[s,s+1]\geq\frac{\ell}{2}}(\omega)ds \geq \frac{\epsilon}{4}\right) = -\infty.$$

For the second term, $\hat{N}_{\ell'}[s,s+1]\chi_{\hat{N}_{\ell'}[s,s+1]\geq\frac{\ell}{2}} \leq \hat{N}_{\ell'}[s,s+1]$ and

$$(3.137) \qquad \int_0^t \hat{N}_{\ell'}[s,s+1]ds \leq \hat{N}_{\ell'}[0,t] + N[t,t+1].$$

93

By Lemma 17,

$$(3.138) \qquad \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t} N[t, t+1] \geq \frac{\epsilon}{8}\right) = -\infty,$$

and by the same argument as in (i),

$$(3.139) \qquad \limsup_{\ell'\to\infty} \limsup_{\ell\to\infty} \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t} \hat{N}_{\ell'}[0, t] \geq \frac{\epsilon}{8}\right) = -\infty.$$

For the third term, notice that

$$(3.140) \qquad \int_0^t \frac{\ell}{2} \chi_{N_{\ell'}[s,s+1]\geq\frac{\ell}{2}} ds \leq \int_0^t N_{\ell'}[s, s+1] \chi_{N_{\ell'}[s,s+1]\geq\frac{\ell}{2}}(\omega) ds.$$

So we can get the same superexponential estimate as before. Finally, for the fourth term,

$$(3.141) \qquad \int_0^t \frac{\ell}{2} \chi_{\hat{N}_{\ell'}[s,s+1]\geq\frac{\ell}{2}} ds \leq \int_0^t \hat{N}_{\ell'}[s, s+1](\omega) ds.$$

We can get the same superexponential estimate as before. $\qquad\square$

Lemma 19. *Assume N_t is a Poisson process with constant rate λ. Then for any $\epsilon > 0$,*

$$(3.142) \qquad \limsup_{\delta\to 0} \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{\delta t} \int_0^t \chi_{N[s,s+\delta]\geq 2}(\omega) ds \geq \epsilon\right) = -\infty.$$

Proof. Let $f(\delta, \omega) = \frac{1}{h(\delta)} \chi_{N[0,\delta]\geq 2}(\omega)$, where $h(\delta)$ is to be chosen later. By Jensen's

inequality and stationarity and independence of increments of the Poisson process,

$$(3.143) \quad \mathbb{E}\left[e^{\int_0^t \frac{1}{\delta} f(\delta, \theta_s \omega) ds}\right] \leq \mathbb{E}\left[e^{\frac{1}{\delta} \int_0^\delta \sum_{j=0}^{[t/\delta]} f(\delta, \theta_{s+j\delta}\omega) ds}\right]$$

$$\leq \mathbb{E}\left[\frac{1}{\delta}\int_0^\delta e^{\sum_{j=0}^{[t/\delta]} f(\delta, \theta_{s+j\delta}\omega)} ds\right]$$

$$= \mathbb{E}\left[e^{\sum_{j=0}^{[t/\delta]} f(\delta, \theta_{j\delta}\omega)}\right]$$

$$= \mathbb{E}\left[e^{f(\delta, \omega)}\right]^{[t/\delta]+1}$$

$$= \left\{e^{1/h(\delta)}(1 - e^{-\lambda\delta} - \lambda\delta e^{-\lambda\delta}) + e^{-\lambda\delta} + \lambda\delta e^{-\lambda\delta}\right\}^{[t/\delta]+1}$$

$$\leq (M' e^{1/h(\delta)} \lambda^2 \delta^2 + 1)^{[t/\delta]+1},$$

for some $M' > 0$. Choose $h(\delta) = \frac{1}{\log(1/\delta)}$. Then,

$$(3.144) \qquad \mathbb{E}\left[e^{\int_0^t \frac{1}{\delta} f(\delta, \theta_s \omega) ds}\right] \leq (M'\delta + 1)^{[t/\delta]+1} \leq e^{Mt},$$

for some $M > 0$. Therefore, by Chebychev's inequality,

$$(3.145) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{1}{\delta h(\delta) t} \int_0^t \chi_{N[s,s+\delta]\geq 2}(\omega) ds \geq \frac{\epsilon}{h(\delta)}\right) \leq M - \frac{\epsilon}{h(\delta)},$$

which holds for any $\delta > 0$. Letting $\delta \to 0$, we get the desired result. □

Lemma 20. *Assume N_t is a Poisson process with constant rate λ. Then for any* $\epsilon > 0$,

$$(3.146) \qquad \limsup_{\ell\to\infty} \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t}\int_0^t N[0,1]\chi_{N[0,1]\geq\ell}(\theta_s\omega) ds \geq \epsilon\right) = -\infty.$$

Proof. Let $h(\ell)$ be some function of ℓ to be chosen later. Following the same

95

argument as in the proof of Lemma 19, we have

$$(3.147) \qquad \mathbb{P}\left(h(\ell) \int_0^t N[0,1] \chi_{N[0,1] \geq \ell}(\theta_s \omega) ds \geq \epsilon h(\ell) t \right)$$

$$\leq \mathbb{E}\left[e^{h(\ell) \int_0^t N[0,1] \chi_{N[0,1] \geq \ell}(\theta_s \omega) ds} \right] e^{-\epsilon h(\ell) t}$$

$$\leq \mathbb{E}\left[e^{h(\ell) N[0,1] \chi_{N[0,1] \geq \ell}} \right]^{[t]+1} e^{-\epsilon h(\ell) t}$$

$$= \left\{ \mathbb{P}(N[0,1] < \ell) + \sum_{k=\ell}^{\infty} e^{h(\ell)k} e^{-\lambda} \frac{\lambda^k}{k!} \right\}^{[t]+1} e^{-\epsilon h(\ell) t}$$

$$\leq \left\{ 1 + C_1 \sum_{k=\ell}^{\infty} e^{h(\ell)k + \log(\lambda)k - \log(k)k} \right\}^{[t]+1} e^{-\epsilon h(\ell) t}$$

$$\leq \left\{ 1 + C_2 e^{h(\ell)\ell + \log(\lambda)\ell - \log(\ell)\ell} \right\}^{[t]+1} e^{-\epsilon h(\ell) t}.$$

Choosing $h(\ell) = (\log(\ell))^{1/2}$ will do the work. $\qquad\qquad\square$

The following Lemma 21 provides us the superexponential estimates that we need. These superexponential estimates have basically been done in Lemma 18. The difference is that in the statement in Lemma 18, we used ω and in Lemma 21 it is changed to ω_t which is what we needed. Lemma 21 has three statements. Part (i) says if you start with a sequence of simple point processes, the limiting point process may not be simple, but this has probability that is superexponentially small. Part (ii) is the usual superexponential we would expect if $\mathcal{M}_S(\Omega)$ were equipped with weak topology. But since we are using a strengthened weak topology with the convergence of first moment as well, we will also need Part (iii).

Lemma 21. *We have the following superexponential estimates.*

(i) For some $g(\delta) \to 0$ as $\delta \to 0$,

$$(3.148) \qquad \limsup_{\delta \to 0} \limsup_{t \to \infty} \frac{1}{t} \log P\left(\frac{1}{\delta t} \int_0^t \chi_{N[0,\delta] \geq 2}(\theta_s \omega_t) ds \geq g(\delta) \right) = -\infty.$$

(ii) For some $\varepsilon(M) \to 0$ as $M \to \infty$,

$$(3.149) \qquad \limsup_{M\to\infty} \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t}\int_0^t \chi_{N[0,1]\geq M}(\theta_s\omega_t)ds \geq \varepsilon(M)\right) = -\infty.$$

(iii) For some $m(\ell) \to 0$ as $\ell \to \infty$,

$$(3.150) \qquad \limsup_{\ell\to\infty} \limsup_{t\to\infty} \frac{1}{t} \log P\left(\frac{1}{t}\int_0^t N[0,1]\chi_{N[0,1]\geq \ell}(\theta_s\omega_t)ds \geq m(\ell)\right) = -\infty.$$

Proof. We can replace the ϵ in the statement of Lemma 18 by $g(\delta)$, $\varepsilon(M)$ and $m(\ell)$ by a standard analysis argument. Here, we can also replace the ω in Lemma 18 by ω_t since

$$(3.151) \qquad \left|\int_0^t \chi_{N[0,\delta]\geq 2}(\theta_s\omega_t)ds - \int_0^t \chi_{N[0,\delta]\geq 2}(\theta_s\omega)ds\right| \leq 2\delta,$$

$$(3.152) \qquad \left|\int_0^t \chi_{N[0,1]\geq M}(\theta_s\omega_t)ds - \int_0^t \chi_{N[0,1]\geq M}(\theta_s\omega)ds\right| \leq 2,$$

and

$$
\begin{aligned}
(3.153) \qquad & \left|\int_0^t N[0,1]\chi_{N[0,1]\geq \ell}(\theta_s\omega_t)ds - \int_0^t N[0,1]\chi_{N[0,1]\geq \ell}(\theta_s\omega)ds\right| \\
& \leq \int_{t-1}^t N[s,s+1](\omega)ds + \int_{t-1}^t N[s,s+1](\omega_t)ds \\
& \leq N[t-1,t+1](\omega) + N[t-1,t+1](\omega_t) \\
& = N[t-1,t+1](\omega) + N[t-1,t](\omega) + N[0,1](\omega).
\end{aligned}
$$

By Lemma 17, we have the superexponential estimate, for any $\epsilon > 0$,

(3.154)
$$\limsup_{t \to \infty} \frac{1}{t} \log P \left(\frac{1}{t} \left\{ N[t-1, t+1](\omega) + N[t-1, t](\omega) + N[0, 1](\omega) \right\} \geq \epsilon \right) = -\infty.$$

$\square$

Lemma 22. *For any* $\delta, M > 0, \ell > 0$, *define*

(3.155)
$$\mathcal{A}_\delta = \left\{ Q \in \mathcal{M}_S(\Omega) : Q(N[0, \delta] \geq 2) \leq \delta g(\delta) \right\},$$

$$\mathcal{A}_M = \left\{ Q \in \mathcal{M}_S(\Omega) : Q(N[0, 1] \geq M) \leq \varepsilon(M) \right\},$$

$$\mathcal{A}_\ell = \left\{ Q \in \mathcal{M}_S(\Omega) : \int_{N[0,1] \geq \ell} N[0, 1] dQ \leq m(\ell) \right\},$$

where $\varepsilon(M) \to 0$ *as* $M \to \infty$, $m(\ell) \to 0$ *as* $\ell \to \infty$ *and* $g(\delta) \to 0$ *as* $\delta \to 0$. *Let*
$\mathcal{A}_{\delta, M, \ell} = \mathcal{A}_\delta \cap \mathcal{A}_M \cap \mathcal{A}_\ell$ *and*

(3.156)
$$\mathcal{A}^n = \bigcap_{j=n}^{\infty} \mathcal{A}_{\frac{1}{j}, j, j}.$$

Then $\mathcal{A}^n$ *is compact.*

Proof. Observe that for $\beta > 0$, the sets

(3.157) $$K_\beta = \bigcap_{k=1}^{\infty} \{ \omega : \{ N[-k, -(k-1)](\omega) \leq \beta \ell_k \} \cap \{ N[k-1, k](\omega) \leq \beta \ell_k \} \}$$

are relatively compact in Ω. Let $\overline{K_\beta}$ be the closure of K_β, which is then compact.

For any $Q \in \mathcal{A}^n$, $Q(N[0, 1] \geq M) \leq \epsilon(M)$ for any $M \geq n$. We can choose β big

enough and an increasing sequence ℓ_k such that $\beta \ell_1 \geq n$ and $\infty > \sum_{k=1}^{\infty} \epsilon(\beta \ell_k) \to 0$

98

as $\beta \to \infty$, uniformly for $Q \in \mathcal{A}^n$,

(3.158)

$$
\begin{aligned}
Q\left(\overline{K_\beta}^c\right) &\leq Q(K_\beta^c) \\
&= Q\left(\bigcup_{k=1}^{\infty} \{N[-k, -(k-1)](\omega) > \beta\ell_k\} \cap \{N[k-1, k](\omega) > \beta\ell_k\}\right) \\
&\leq \sum_{k=1}^{\infty} \{Q(N[-(k-1), -k] > \beta\ell) + Q(N[k-1, k] > \beta\ell_k)\} \\
&= 2\sum_{k=1}^{\infty} Q(N[0,1] > \beta\ell_k) \\
&\leq 2\sum_{k=1}^{\infty} \epsilon(\beta\ell_k) \to 0
\end{aligned}
$$

as $\beta \to \infty$. Therefore, $\mathcal{A}^n$ is tight in the weak topology and by Prokhorov theorem $\mathcal{A}^n$ is precompact in the weak topology. In other words, for any sequence in $\mathcal{A}^n$, there exists a subsequence, say Q_n such that $Q_n \to Q$ weakly as $n \to \infty$ for some Q. By the definition of $\mathcal{A}^n$, Q_n are uniformly integrable, which implies that $\int N[0,1]dQ_n \to \int N[0,1]dQ$ as $n \to \infty$. It is also easy to see that $\mathcal{A}^n$ is closed by checking that each $\mathcal{A}_{\frac{1}{j},j,j}$ is closed. That implies that $Q \in \mathcal{A}^n$. Finally, we need to check that Q is a simple point process. Let $I_{j,\delta} = [(j-1)\delta, j\delta]$. We have for any

$Q \in \mathcal{A}^n$,

$$(3.159) \quad Q\left(\exists t : N[t-, t] \geq 2\right) = Q\left(\bigcup_{k=1}^{\infty}\{\exists t \in [-k, k] : N[t-, t] \geq 2\}\right)$$

$$= Q\left(\bigcup_{k=1}^{\infty}\bigcap_{\delta>0}\bigcup_{j=-[k/\delta]+1}^{[k/\delta]}\{\omega : \#\{\omega \cup I_{j,\delta}\} \geq 2\}\right)$$

$$\leq \sum_{k=1}^{\infty}\inf_{\delta=\frac{1}{m}, m \geq n}\sum_{j=-[k/\delta]+1}^{[k/\delta]}Q(\#\{\omega \cup I_{j,\delta}\} \geq 2)$$

$$\leq \sum_{k=1}^{\infty}\inf_{\delta=\frac{1}{m}, m \geq n}\{2[k/\delta]\delta g(\delta)\}$$

$$= 0.$$

Hence, $\mathcal{A}^n$ is precompact in our topology. Since $\mathcal{A}^n$ is closed, it is compact. □

3.5 Concluding Remarks

In this chapter, we obtained a process-level large deviation principle for a wide class of simple point processes, i.e. nonlinear Hawkes processes. Indeed, the methods and ideas should apply to other simple point processes as well and we should expect to get the same expression for the rate function $H(Q)$. For $H(Q) < \infty$, it should be of the form

$$(3.160) \quad H(Q) = \int_{\Omega}\int_0^1 \lambda(\omega, s) - \hat{\lambda}(\omega, s) + \log\left(\frac{\hat{\lambda}(\omega, s)}{\lambda(\omega, s)}\right)\hat{\lambda}(\omega, s)ds Q(d\omega),$$

where $\lambda(\omega, s)$ is the intensity of the underlying simple point process. Now, it would be interesting to ask for what conditions for a simple point process would guarantee the process-level large deviation principle that we obtained in this chapter? First,

we have to assume that $\lambda(\omega, t)$ is predictable and progressively measurable. Second, in our proof of the upper bound in this chapter, the key assumption we used about nonlinear Hawkes process was that $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$. That is crucial to guarantee the superexponential estimates we needed for the upper bound. If for a simple point process, we have $\lambda(\omega, t) \leq F(N(t, \omega))$ for some sublinear function $F(\cdot)$, we would expect the superexponential estimates still to work for the upper bound. Third, it is not enough to have $\lambda(\omega, t) \leq F(N(t, \omega))$ for sublinear $F(\cdot)$ to get the full large deviation principle. The reason is that in the proof of lower bound, in particular, in Lemma 10, we need to use the fact that any memory in $\lambda(\omega, t)$ has memory will decay to zero over time. For nonlinear Hawkes processes, this is guaranteed by the assumption that $\int_0^\infty h(t)dt < \infty$, which is crucial in the proof of Lemma 10. Indeed for any simple point process P, if you want to define P^{ω^-}, the probability measure conditional on the past history ω^-, to make sense of it, you have to have some regularities to ensure that the memory of the history will decay to zero eventually over time. From this perspective, nonlinear Hawkes processes form a rich and ideal class for which the process-level large deviation principle holds.

Chapter 4

Large Deviations for Markovian Nonlinear Hawkes Processes

In Chapter 3, we studied the large deviations for $(N_t/t \in \cdot)$ by proving first a process-level, i.e. level-3 large deviation principle and then applying the contraction principle. In this chapter, we will obtain an alternative expression for the rate function of the large deviation principle of $(N_t/t \in \cdot)$ when $h(\cdot)$ is exponential or sums of exponentials. The main idea is that when $h(\cdot)$ is exponential or sums of exponentials, the system is Markovian and we can use Feynman-Kac formula to obtain an upper bound and some tilting method to get a lower bound. The assumption $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$ will provide us the compactness in order to apply a minmax theorem to match the lower bound and the upper bound.

4.1 An Ergodic Lemma

In this section, we prove an ergodic theorem for a class of Markovian processes with jumps more general than the Markovian nonlinear Hawkes processes.

Let $Z_i(t) := \sum_{\tau_j < t} a_i e^{-b_i(t - \tau_j)}$, $1 \le i \le d$, where $b_i > 0$, $a_i \ne 0$ (might be negative), and τ_j's are the arrivals of the simple point process with intensity $\lambda(Z_1(t), \cdots, Z_d(t))$ at time t, where $\lambda : \mathcal{Z} \to \mathbb{R}^+$ and $\mathcal{Z} := \mathbb{R}^{\epsilon_1} \times \cdots \mathbb{R}^{\epsilon_d}$ is the domain for $(Z_1(t), \ldots, Z_d(t))$, where $\mathbb{R}^{\epsilon_i} := \mathbb{R}^+$ or $\mathbb{R}^-$ depending on whether $\epsilon_i = +1$ or -1, where $\epsilon_i = +1$ if $a_i > 0$ and $\epsilon_i = -1$ otherwise. If we assume the exciting function to be $h(t) = \sum_{i=1}^{d} a_i e^{-b_i t}$, then a Markovian nonlinear Hawkes process is a simple point process with intensity of the form $\lambda(\sum_{i=1}^{d} Z_i(t))$.

The generator $\mathcal{A}$ for $(Z_1(t), \ldots, Z_d(t))$ is given by

$$(4.1) \quad \mathcal{A}f = -\sum_{i=1}^{d} b_i z_i \frac{\partial f}{\partial z_i} + \lambda(z_1, \ldots, z_d)[f(z_1 + a_1, \ldots, z_d + a_d) - f(z_1, \ldots, z_d)].$$

We want to prove the existence and uniqueness of the invariant probability measure for $(Z_1(t), \ldots, Z_d(t))$. Here the invariance is in time.

The lecture notes [47] by Martin Hairer gives the criterion for the existence and uniqueness of the invariant probability measure for Markov processes.

Suppose we have a jump diffusion process with generator $\mathcal{L}$. If we can find u such that $u \ge 0$, $\mathcal{L}u \le C_1 - C_2 u$ for some constants $C_1, C_2 > 0$, then, there exists an invariant probability measure. We thereby have the following lemma.

Lemma 23. *Consider* $h(t) = \sum_{i=1}^{d} a_i e^{-b_i t} > 0$. *Let* $\epsilon_i = +1$ *if* $a_i > 0$ *and* $\epsilon_i = -1$ *if* $a_i < 0$. *Assume* $\lambda(z_1, \ldots, z_n) \le \sum_{i=1}^{d} \alpha_i |z_i| + \beta$, *where* $\beta > 0$ *and* $\alpha_i > 0$, $1 \le i \le d$, *satisfies* $\sum_{i=1}^{d} \frac{|a_i|}{b_i} \alpha_i < 1$. *Then, there exists a unique invariant probability measure for* $(Z_1(t), \ldots, Z_d(t))$.

103

Proof. The lecture notes [47] by Martin Hairer gives the criterion for the existence of an invariant probability measure for Markov processes. Suppose we have a jump diffusion process with generator $\mathcal{L}$. If we can find u such that $u \geq 0$, $\mathcal{L}u \leq C_1 - C_2 u$ for some constants $C_1, C_2 > 0$, then, there exists an invariant probability measure.

Try $u(z_1, \ldots, z_d) = \sum_{i=1}^{d} \epsilon_i c_i z_i \geq 0$, where $c_i > 0$, $1 \leq i \leq d$. Then,

$$(4.2) \qquad \mathcal{A}u = -\sum_{i=1}^{d} b_i \epsilon_i c_i z_i + \lambda(z_1, \ldots, z_d) \sum_{i=1}^{d} a_i \epsilon_i c_i$$

$$\leq -\sum_{i=1}^{d} b_i c_i |z_i| + \sum_{i=1}^{d} \alpha_i |z_i| \sum_{i=1}^{d} |a_i| c_i + \beta \sum_{i=1}^{d} |a_i| c_i.$$

Taking $c_i = \frac{\alpha_i}{b_i} > 0$, we get

$$(4.3) \qquad \mathcal{A}u \leq -\left(1 - \sum_{i=1}^{d} \frac{|a_i|\alpha_i}{b_i}\right) \sum_{i=1}^{d} \alpha_i |z_i| + \beta \sum_{i=1}^{d} \frac{|a_i|\alpha_i}{b_i}$$

$$\leq -\min_{1 \leq i \leq d} b_i \cdot \left(1 - \sum_{i=1}^{d} \frac{|a_i|\alpha_i}{b_i}\right) u + \beta \sum_{i=1}^{d} \frac{|a_i|\alpha_i}{b_i}.$$

Next, we will prove the uniqueness of the invariant probability measure. It is sufficient to prove that for any $x, y \in \mathcal{Z}_d$, there exist times $T_1, T_2 > 0$ such that $\mathcal{P}^x(T_1, \cdot)$ and $\mathcal{P}^y(T_2, \cdot)$ are not mutually singular. Here $\mathcal{P}^x(T, \cdot) := \mathbb{P}(Z_T^x \in \cdot)$, where Z_T^x is Z_T starting at $Z_0 = x$, i.e. $Z_T^x = xe^{-bT} + \sum_{\tau_j < T} ae^{-b(T-\tau_j)}$. To see this, let us prove by contradiction. If there were two distinct invariant probability measures μ_1 and μ_2, then there exist two disjoints sets E_1 and E_2 such that $\mu_1 : E_1 \to E_1$ and $\mu_2 : E_2 \to E_2$, see for example Varadhan [108]. Now, we can choose $x_1 \in E_1$ and $x_2 \in E_2$. so that $\mathcal{P}^{x_1}(T_1, \cdot)$ and $\mathcal{P}^{x_2}(T_2, \cdot)$ are supported on E_1 and E_2 respectively for any $T_1, T_2 > 0$, which implies that $\mathcal{P}^{x_1}(T_1, \cdot)$ and $\mathcal{P}^{x_2}(T_2, \cdot)$ are mutually singular. This leads to a contradiction.

Consider the simplest case $h(t) = ae^{-bt}$. Let us assume that $x > y > 0$. Conditioning on the event that Z_t^x and Z_t^y have exactly one jump during the time interval $(0, T)$ respectively, the laws of $\mathcal{P}^x(T, \cdot)$ and $\mathcal{P}^y(T, \cdot)$ have positive densities on the sets

$$(4.4) \qquad \big((a+x)e^{-bT}, xe^{-bT} + a\big) \quad \text{and} \quad \big((a+y)e^{-bT}, ye^{-bT} + a\big)$$

respectively. Choosing $T > \frac{1}{b}\log(\frac{x-y+a}{a})$, we have

$$(4.5) \qquad \big((a+x)e^{-bT}, xe^{-bT} + a\big) \bigcap \big((a+y)e^{-bT}, ye^{-bT} + a\big) \neq \varnothing,$$

which implies that $\mathcal{P}^x(T, \cdot)$ and $\mathcal{P}^y(T, \cdot)$ are not mutually singular.

Similarly, one can show the uniqueness of the invariant probability measure for the multidimensional case. Indeed, it is easy to see that for any $x, y \in \mathcal{Z}_d$, $Z_{T_1}^x$ and $Z_{T_2}^y$ hit a common point for some T_1 and T_2 after possibly different number of jumps. Here $Z_t^x := (Z_t^{x_1}, \ldots, Z_t^{x_d}) \in \mathcal{Z}_d$ and $Z_t^y := (Z_t^{y_1}, \ldots, Z_t^{y_d}) \in \mathcal{Z}_d$, where $Z_t^{x_i} = x_i e^{-b_i t} + \sum_{\tau_j < t} a_i e^{-b_i(t-\tau_j)}$, $1 \leq i \leq d$. Since $\mathcal{P}^x(T_1, \cdot)$ and $\mathcal{P}^y(T_2, \cdot)$ have probability densities, $\mathcal{P}^x(T_1, \cdot)$ and $\mathcal{P}^y(T_2, \cdot)$ are not mutually singular for some T_1 and T_2. $\qquad\qquad \square$

4.2 Large Deviations for Markovian Nonlinear Hawkes Processes with Exponential Exciting Function

We assume first that $h(t) = ae^{-bt}$, where $a, b > 0$, i.e. the process Z_t jumps upwards an amount a at each point and decays exponentially between points with rate b. In this case, Z_t is Markovian.

Notice first that $Z_0 = 0$ and

$$(4.6) \qquad\qquad dZ_t = -bZ_t dt + a dN_t,$$

which implies that $N_t = \frac{1}{a} Z_t + \frac{b}{a} \int_0^t Z_s ds$.

We prove first the existence of the limit of the logarithmic moment generating function of N_t.

Theorem 14. *Assume that $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$ and that $\lambda(\cdot)$ is continuous and bounded below by some positive constant. Then,*

$$(4.7) \qquad\qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \Gamma(\theta),$$

where

$$(4.8) \quad \Gamma(\theta) = \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e} \left\{ \int \frac{\theta b}{a} z \hat{\pi}(dz) + \int (\hat{\lambda} - \lambda) \hat{\pi}(dz) - \int \left(\log(\hat{\lambda}/\lambda) \right) \hat{\lambda} \hat{\pi}(dz) \right\},$$

where $\mathcal{Q}_e$ is defined as

(4.9) $\mathcal{Q}_e = \left\{ (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} : \hat{\mathcal{A}} \text{ has unique invariant probability measure } \hat{\pi} \right\},$

where

(4.10) $\mathcal{Q} = \left\{ (\hat{\lambda}, \hat{\pi}) : \hat{\pi} \in \mathcal{M}(\mathbb{R}^+), \int z\hat{\pi}(dz) < \infty, \hat{\lambda} \in L^1(\hat{\pi}), \hat{\lambda} > 0 \right\},$

where $\mathcal{M}(\mathbb{R}^+)$ denotes the space of probability measures on $\mathbb{R}^+$ and for any $\hat{\lambda}$ such that $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}$, we define the generator $\hat{\mathcal{A}}$ as

(4.11) $$\hat{\mathcal{A}}f(z) = -bz\frac{\partial f}{\partial z} + \hat{\lambda}(z)[f(z+a) - f(z)].$$

for any $f : \mathbb{R}^+ \to \mathbb{R}$ that is C^1, i.e. continuously differentiable.

Proof. By Lemma 24, $\mathbb{E}[e^{\theta N_t}] < \infty$ for any $\theta \in \mathbb{R}$, also

(4.12) $$\mathbb{E}[e^{\theta N_t}] = \mathbb{E}\left[e^{\frac{\theta}{a}\left(Z_t + b\int_0^t Z_s ds\right)} \right].$$

Define the set

(4.13) $\mathcal{U}_\theta = \left\{ u \in C^1(\mathbb{R}^+, \mathbb{R}^+) : u(z) = e^{f(z)}, \text{ where } f \in \mathcal{F} \right\},$

where

(4.14) $\mathcal{F} = \bigg\{ f : f(z) = Kz + g(z) + L, K > \dfrac{\theta}{a}, K, L \in \mathbb{R},$

$\qquad\qquad\qquad\qquad\qquad\qquad g \text{ is } C_1 \text{ with compact support} \bigg\}.$

107

Now for any $u \in \mathcal{U}_\theta$, define

$$(4.15) \qquad M := \sup_{z \geq 0} \frac{\mathcal{A}u(z) + \frac{\theta b}{a} z u(z)}{u(z)}.$$

By Dynkin's formula if $M < \infty$, for $V(z) := \frac{\theta b}{a} z$, we have

$$(4.16)$$

$$\mathbb{E}\left[u(Z_t)e^{\int_0^t V(Z_s)ds}\right] = u(Z_0) + \int_0^t \mathbb{E}\left[(\mathcal{A}u(Z_s) + V(Z_s)u(Z_s))e^{\int_0^s V(Z_v)dv}\right] ds$$

$$\leq u(Z_0) + M \int_0^t \mathbb{E}\left[u(Z_s)e^{\int_0^s V(Z_v)dv}\right] ds,$$

which implies by Gronwall's lemma that

$$(4.17) \qquad \mathbb{E}\left[u(Z_t)e^{\int_0^t V(Z_s)ds}\right] \leq u(Z_0)e^{Mt} = u(0)e^{Mt}.$$

Observe that by the definition of $\mathcal{U}_\theta$, for any $u \in \mathcal{U}_\theta$, we have $u(z) \geq c_1 e^{\frac{\theta}{a} z}$ for some constant $c_1 > 0$ and therefore by (4.12) and (4.17),

$$(4.18) \qquad \mathbb{E}\left[e^{\theta N_t}\right] \leq \frac{1}{c_1} \mathbb{E}\left[u(Z_t)e^{\int_0^t \frac{\theta b}{a} Z_s ds}\right] \leq \frac{1}{c_1} u(0)e^{Mt}.$$

Hence,

$$(4.19) \qquad \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta N_t}\right] \leq M = \sup_{z \geq 0} \frac{\mathcal{A}u(z) + \frac{\theta b}{a} z u(z)}{u(z)},$$

which is still true even if $M = \infty$. Since this holds for any $u \in \mathcal{U}_\theta$,

$$(4.20) \qquad \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta N_t}\right] \leq \inf_{u \in \mathcal{U}_\theta} \sup_{z \geq 0} \frac{\mathcal{A}u(z) + \frac{\theta b}{a} z u(z)}{u(z)}.$$

108

Define the tilted probability measure $\hat{\mathbb{P}}$ by

$$(4.21) \qquad \frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = \exp\left\{\int_0^t (\lambda(Z_s) - \hat{\lambda}(Z_s))ds + \int_0^t \log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right)dN_s\right\}.$$

Notice that $\hat{\mathbb{P}}$ defined in (4.21) is indeed a probability measure by Girsanov formula. (For the theory of absolute continuity for point processes and their Girsanov formulas, we refer to Lipster and Shiryaev [72].)

Now by Jensen's inequality,

$$(4.22) \quad \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$$

$$= \liminf_{t\to\infty} \frac{1}{t} \log \hat{\mathbb{E}}\left[\exp\left\{\theta N_t - \log\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t}\right\}\right]$$

$$\geq \liminf_{t\to\infty} \hat{\mathbb{E}}\left[\frac{1}{t}\theta N_t - \frac{1}{t}\log\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t}\right]$$

$$= \liminf_{t\to\infty} \hat{\mathbb{E}}\left[\frac{1}{t}\theta N_t - \frac{1}{t}\int_0^t (\lambda(Z_s) - \hat{\lambda}(Z_s))ds - \int_0^t \log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right)dN_s\right].$$

Since $N_t - \int_0^t \hat{\lambda}(Z_s)ds$ is a martingale under $\hat{\mathbb{P}}$, we have

$$(4.23) \qquad \hat{\mathbb{E}}\left[\int_0^t \log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right)(dN_s - \hat{\lambda}(Z_s)ds)\right] = 0.$$

Therefore, by the ergodic theorem, (for a reference, see Chapter 16.4 of Koralov

and Sinai [64]), for any $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$,

$$
(4.24) \quad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]
$$

$$
\geq \liminf_{t \to \infty} \hat{\mathbb{E}} \left[\frac{1}{t} \theta N_t - \frac{1}{t} \int_0^t (\lambda(Z_s) - \hat{\lambda}(Z_s)) ds - \int_0^t \log \left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)} \right) \hat{\lambda}(Z_s) ds \right]
$$

$$
= \int \frac{\theta b}{a} z \hat{\pi}(dz) + \int (\hat{\lambda} - \lambda) \hat{\pi}(dz) - \int \left(\log(\hat{\lambda}) - \log(\lambda) \right) \hat{\lambda} \hat{\pi}(dz).
$$

Hence,

$$
(4.25) \quad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]
$$

$$
\geq \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e} \left\{ \int \frac{\theta b}{a} z \hat{\pi} + \int (\hat{\lambda} - \lambda) \hat{\pi} - \int \left(\log(\hat{\lambda}) - \log(\lambda) \right) \hat{\lambda} \hat{\pi} \right\}.
$$

Recall that

$$
(4.26) \quad \mathcal{F} = \left\{ f : f(z) = Kz + g(z) + L, K > \frac{\theta}{a}, K, L \in \mathbb{R}, \right.
$$

$$
\left. g \text{ is } C_1 \text{ with compact support} \right\}.
$$

We claim that

$$
(4.27) \quad \inf_{f \in \mathcal{F}} \left\{ \int \hat{\mathcal{A}} f(z) \hat{\pi}(dz) \right\} = \begin{cases} 0 & \text{if } (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e, \\ -\infty & \text{if } (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} \backslash \mathcal{Q}_e. \end{cases}
$$

It is easy to see that for $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$, and g being C_1 with compact support, $\int \mathcal{A} g \hat{\pi} = 0$. Next, we can find a sequence $f_n(z) \to z$ pointwise under the bound $|f_n(z)| \leq \alpha z + \beta$, for some $\alpha, \beta > 0$, where $f_n(z)$ is C_1 with compact support. But by our definition of $\mathcal{Q}$, $\int z \hat{\pi} < \infty$. So by the dominated convergence theorem,

$\int \hat{\mathcal{A}} z \hat{\pi} = 0$. The nontrivial part is to prove that if for any $g \in \mathcal{G} = \{g(z) + L, g$ is C_1 with compact support$\}$ such that $\int \hat{\mathcal{A}} g \hat{\pi} = 0$, then $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$. We can easily check the conditions in Echevrría [32]. (For instance, $\mathcal{G}$ is dense in $C(\mathbb{R}^+)$, the set of continuous and bounded functions on $\mathbb{R}^+$ with limit that exists at infinity and $\hat{\mathcal{A}}$ satisfies the minimum principle, i.e. $\hat{\mathcal{A}} f(z_0) \geq 0$ for any $f(z_0) = \inf_{z \in \mathbb{R}^+} f(z)$. This is because at minimum, the first derivative of f vanishes and $\hat{\lambda}(z_0)(f(z_0 + a) - f(z_0)) \geq 0$. The other conditions in Echeverría [32] can also be easily verified.) Thus, Echevrría [32] implies that $\hat{\pi}$ is an invariant measure. Now, our proof in Lemma 23 shows that $\hat{\pi}$ has to be unique as well. Therefore, $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$. This implies that if $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} \backslash \mathcal{Q}_e$, there exists some $g \in \mathcal{G}$, such that $\int \hat{\mathcal{A}} g \hat{\pi} \neq 0$. Now, any constant multiplier of g still belongs to $\mathcal{G}$ and thus $\inf_{g \in \mathcal{G}} \int \hat{\mathcal{A}} g \hat{\pi} = -\infty$ and hence $\inf_{f \in \mathcal{F}} \int \hat{\mathcal{A}} f \hat{\pi} = -\infty$ if $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} \backslash \mathcal{Q}_e$.

Therefore,

$$(4.28) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \geq \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}} \inf_{f \in \mathcal{F}} \left\{ \int \frac{\theta b}{a} z \hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) + \int \hat{\mathcal{A}} f \hat{\pi} \right\}$$

$$(4.29) \qquad \geq \sup_{(\hat{\lambda}\hat{\pi}, \hat{\pi}) \in \mathcal{R}} \inf_{f \in \mathcal{F}} \left\{ \int \frac{\theta b}{a} z \hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) + \int \hat{\mathcal{A}} f \hat{\pi} \right\},$$

where $\mathcal{R} = \{(\hat{\lambda}\hat{\pi}, \hat{\pi}) : (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}\}$ and

$$(4.30) \qquad \hat{H}(\hat{\lambda}, \hat{\pi}) = \int \left[(\lambda - \hat{\lambda}) + \log \left(\hat{\lambda}/\lambda \right) \hat{\lambda} \right] \hat{\pi}.$$

Define

$$(4.31) \quad F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) = \int \frac{\theta b}{a} z\hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) + \int \hat{\mathcal{A}} f\hat{\pi}$$

$$= \int \frac{\theta b}{a} z\hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) - \int bz\frac{\partial f}{\partial z}\hat{\pi} + \int (f(z+a) - f(z))\hat{\lambda}\hat{\pi}.$$

Notice that F is linear in f and hence convex in f and also

$$(4.32) \qquad \hat{H}(\hat{\lambda}, \hat{\pi}) = \sup_{f \in C_b(\mathbb{R}^+)} \left\{ \int \left[\hat{\lambda}f + \lambda(1 - e^f) \right] \hat{\pi} \right\},$$

where $C_b(\mathbb{R}^+)$ denotes the set of bounded functions on $\mathbb{R}^+$. Inside the bracket above, it is linear in both $\hat{\pi}$ and $\hat{\lambda}\hat{\pi}$. Hence $\hat{H}$ is weakly lower semicontinuous and convex in $(\hat{\lambda}\hat{\pi}, \hat{\pi})$. Therefore, F is concave in $(\hat{\lambda}\hat{\pi}, \hat{\pi})$. Furthermore, for any $f = Kz + g + L \in \mathcal{F}$,

$$(4.33) \qquad F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) = \int \left(\frac{\theta}{a} - K \right) bz\hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) - \int bz\frac{\partial g}{\partial z}\hat{\pi}$$

$$+ \int (g(z+a) - g(z))\hat{\lambda}\hat{\pi} + Ka \int \hat{\lambda}\hat{\pi}.$$

If $\lambda_n \pi_n \to \gamma_\infty$ and $\pi_n \to \pi_\infty$ weakly, then, since g is C_1 with compact support, we have

$$(4.34) \qquad -\int bz\frac{\partial g}{\partial z}\pi_n + \int (g(z+a) - g(z))\lambda_n\pi_n + Ka \int \lambda_n\pi_n$$

$$\to -\int bz\frac{\partial g}{\partial z}\pi_\infty + \int (g(z+a) - g(z))\gamma_\infty + Ka \int \gamma_\infty,$$

as $n \to \infty$. Moreover, in general, if $P_n \to P$ weakly, then, for any f which is upper semicontinuous and bounded from above, we have $\limsup_n \int f dP_n \leq \int f dP$. Since

$\left(\frac{\theta}{a} - K\right) bz$ is continuous and nonpositive on $\mathbb{R}^+$, we have

(4.35) $$\limsup_{n \to \infty} \int \left(\frac{\theta}{a} - K\right) bz\pi_n \leq \int \left(\frac{\theta}{a} - K\right) bz\pi_\infty.$$

Hence, we conclude that F is upper semicontinuous in the weak topology.

In order to switch the supremum and infimum in (4.29), since we have already proved that F is concave, upper semicontinuous in $(\hat{\lambda}\hat{\pi}, \hat{\pi})$ and convex in f, it is sufficient to prove the compactness of $\mathcal{R}$ to apply Ky Fan's minmax theorem (see Fan [37]). Indeed, Joó developed some level set method and proved that it is sufficient to show the compactness of the level set (see Joó [60] and Frenk and Kassay [40]). In other words, it suffices to prove that, for any $C \in \mathbb{R}$ and $f \in \mathcal{F}$, the level set

(4.36) $$\left\{ (\hat{\lambda}\hat{\pi}, \hat{\pi}) \in \mathcal{R} : \hat{H} + \int bz\frac{\partial f}{\partial z}\hat{\pi} - \frac{\theta b}{a}z\hat{\pi} - \hat{\lambda}[f(z+a) - f(z)]\hat{\pi} \leq C \right\}$$

is compact.

Fix any $f = Kz + g + L \in \mathcal{F}$, where $K > \frac{\theta}{a}$ and g is C_1 with compact support and L is some constant, uniformly for any pair $(\hat{\lambda}\hat{\pi}, \hat{\pi})$ that is in the level set of

(4.36), there exists some $C_1, C_2 > 0$ such that

(4.37)

$$C_1 \geq \hat{H} + \left(K - \frac{\theta}{a} \right) b \int z\hat{\pi} - C_2 \int \hat{\lambda}\hat{\pi}$$

$$\geq \int_{\hat{\lambda} \geq cz+\ell} \left[\lambda - \hat{\lambda} + \hat{\lambda} \log(\hat{\lambda}/\lambda) \right] \hat{\pi} + \left(K - \frac{\theta}{a} \right) b \int z\hat{\pi}$$

$$- C_2 \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} - C_2 \int_{\hat{\lambda} < cz+\ell} \hat{\lambda}\hat{\pi}$$

$$\geq \left[\min_{z \geq 0} \log \frac{cz+\ell}{\lambda(z)} - 1 - C_2 \right] \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} + \left[-c \cdot C_2 + \left(K - \frac{\theta}{a} \right) b \right] \int z\hat{\pi} - \ell C_2.$$

We choose $0 < c < \left(K - \frac{\theta}{a} \right) \frac{b}{C_2}$ and ℓ large enough so that $\min_{z \geq 0} \log \frac{cz+\ell}{\lambda(z)} - 1 - C_2 > 0$, where we used the fact that $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$ and $\min_z \lambda(z) > 0$. Hence,

$$(4.38) \qquad \int z\hat{\pi} \leq C_3, \quad \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} \leq C_4,$$

where

$$(4.39) \qquad C_3 = \frac{C_1 + \ell C_2}{-c \cdot C_2 + \left(K - \frac{\theta}{a} \right) b}, \quad C_4 = \frac{C_1 + \ell C_2}{\min_{z \geq 0} \log \frac{cz+\ell}{\lambda(z)} - 1 - C_2}.$$

Therefore, we have

$$(4.40) \qquad \int \hat{\lambda}\hat{\pi} = \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} + \int_{\hat{\lambda} < cz+\ell} \hat{\lambda}\hat{\pi} \leq C_4 + c \cdot C_3 + \ell,$$

and hence

$$(4.41) \qquad \hat{H}(\hat{\lambda}, \hat{\pi}) \leq C_1 + C_2 \left[C_4 + c \cdot C_3 + \ell \right] < \infty.$$

Therefore, for any $(\lambda_n \pi_n, \pi_n) \in \mathcal{R}$, we get

$$(4.42) \qquad \limsup_{\ell \to \infty} \sup_n \int_{z \ge \ell} \pi_n \le \limsup_{\ell \to \infty} \sup_n \frac{1}{\ell} \int z \pi_n \le \lim_{\ell \to \infty} \frac{C_3}{\ell} = 0,$$

which implies the tightness of π_n. By Prokhorov's Theorem, there exists a subsequence of π_n which converges weakly to π_∞. We also want to show that there exists some γ_∞ such that $\lambda_n \pi_n \to \gamma_\infty$ weakly (passing to a subsequence if necessary). It is enough to show that

(i) $\sup_n \int \lambda_n \pi_n < \infty$.

(ii) $\lim_{\ell \to \infty} \sup_n \int_{z \ge \ell} \lambda_n \pi_n = 0$.

(i) and (ii) will give us tightness of $\lambda_n \pi_n$ and hence implies the weak convergence for a subsequence.

Now, let us prove statements (i) and (ii).

To prove (i), notice that

$$(4.43) \qquad \sup_n \int \lambda_n \pi_n = \sup_n \int \frac{b}{a} z \pi_n \le \frac{b}{a} [C_4 + c \cdot C_3 + \ell] < \infty.$$

To prove (ii), notice that $(\lambda - \lambda_n) + \lambda_n \log(\lambda_n/\lambda) \ge 0$. That is because $x - 1 - \log x \ge 0$ for any $x > 0$ and hence

$$(4.44) \qquad \lambda - \hat{\lambda} + \hat{\lambda} \log(\hat{\lambda}/\lambda) = \hat{\lambda} \left[(\lambda/\hat{\lambda}) - 1 - \log(\lambda/\hat{\lambda}) \right] \ge 0.$$

115

Notice that

$$(4.45) \qquad \limsup_{\ell \to \infty} \sup_n \int_{z \geq \ell} \lambda_n \pi_n \leq \limsup_{\ell \to \infty} \sup_n \int_{\lambda_n < \sqrt{\lambda z}, z \geq \ell} \lambda_n \pi_n$$

$$+ \limsup_{\ell \to \infty} \sup_n \int_{\lambda_n \geq \sqrt{\lambda z}, z \geq \ell} \lambda_n \pi_n.$$

For the first term, since $\sup_n \int z \pi_n < \infty$ and $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$,

$$(4.46) \qquad \limsup_{\ell \to \infty} \sup_n \int_{\lambda_n < \sqrt{\lambda z}, z \geq \ell} \lambda_n \pi_n \leq \limsup_{\ell \to \infty} \sup_n \int_{z \geq \ell} \sqrt{\lambda z} \pi_n = 0.$$

For the second term, since $\limsup_{z \to \infty} \frac{\lambda(z)}{z} = 0$,

$$(4.47) \qquad \limsup_{\ell \to \infty} \sup_n \int_{\lambda_n \geq \sqrt{\lambda z}, z \geq \ell} \lambda_n \pi_n$$

$$\leq \limsup_{\ell \to \infty} \sup_n \hat{H}(\lambda_n, \pi_n) \sup_{\lambda_n \geq \sqrt{\lambda z}, z \geq \ell} \frac{\lambda_n}{\lambda - \lambda_n + \lambda_n \log(\lambda_n/\lambda)} = 0.$$

Therefore, passing to some subsequence if necessary, we have $\lambda_n \pi_n \to \gamma_\infty$ and $\pi_n \to \pi_\infty$ weakly. Since we proved that F is upper semicontinuous in the weak topology, the level set is compact in the weak topology. Therefore, we can switch

116

the supremum and infimum in (4.29) and get

$$(4.48) \qquad \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta N_t}\right]$$

$$(4.49) \qquad \geq \inf_{f\in\mathcal{F}} \sup_{\hat{\pi}:\int z\hat{\pi}<\infty} \sup_{\hat{\lambda}\in L^1(\hat{\pi})} \left\{ \int \frac{\theta b}{a} z\hat{\pi} + (\hat{\lambda}-\lambda)\hat{\pi} - \log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi} + \hat{\mathcal{A}}f\hat{\pi} \right\}$$

$$(4.50) \qquad = \inf_{f\in\mathcal{F}} \sup_{\hat{\pi}:\int z\hat{\pi}<\infty} \int \left[\frac{\theta bz}{a} + \lambda(z)(e^{f(z+a)-f(z)}-1) - bz\frac{\partial f}{\partial z} \right] \hat{\pi}(dz)$$

$$(4.51) \qquad = \inf_{f\in\mathcal{F}} \sup_{z\geq 0} \left[\frac{\theta bz}{a} + \lambda(z)(e^{f(z+a)-f(z)}-1) - bz\frac{\partial f}{\partial z} \right]$$

$$(4.52) \qquad = \inf_{f\in\mathcal{F}} \sup_{z\geq 0} \left[\frac{\theta bze^{f(z)}}{ae^{f(z)}} + \frac{\lambda(z)}{e^{f(z)}}(e^{f(z+a)} - e^{f(z)}) - \frac{bz}{e^{f(z)}}\frac{\partial e^{f(z)}}{\partial z} \right]$$

$$(4.53) \qquad \geq \inf_{u\in\mathcal{U}_\theta} \sup_{z\geq 0} \left\{ \frac{\mathcal{A}u}{u} + \frac{\theta b}{a}z \right\}.$$

We need some justifications. Define $G(\hat{\lambda}) = \hat{\lambda} - \log(\hat{\lambda}/\lambda)\hat{\lambda} + \hat{\mathcal{A}}f$. The supremum of $G(\hat{\lambda})$ is achieved when $\frac{\partial G}{\partial \hat{\lambda}} = 0$ which implies $\hat{\lambda} = \lambda e^{f(z+a)-f(z)}$. Notice that for $f \in \mathcal{F}$, the optimal $\hat{\lambda} = \lambda e^{f(z+a)-f(z)}$ satisfies $\int \hat{\lambda}\hat{\pi} < \infty$ since $\int \lambda\hat{\pi} < \infty$ and $\int z\hat{\pi} < \infty$. This gives us (4.50). Next, let us explain (4.51). For any probability measure $\hat{\pi}$,

$$(4.54) \qquad \int \left[\frac{\theta bz}{a} + \lambda(z)(e^{f(z+a)-f(z)}-1) - bz\frac{\partial f}{\partial z} \right] \hat{\pi}(dz)$$

$$\leq \sup_{z\geq 0} \left[\frac{\theta bz}{a} + \lambda(z)(e^{f(z+a)-f(z)}-1) - bz\frac{\partial f}{\partial z} \right],$$

which implies the right hand side of (4.50) is less or equal to the right hand side of (4.51). To prove the other direction. For any $f = Kz + g + L \in \mathcal{F}$, we have

$$(4.55) \qquad \frac{\theta bz}{a} + \lambda(z)(e^{f(z+a)-f(z)}-1) - bz\frac{\partial f}{\partial z}$$

$$= \left(\frac{\theta b}{a} - Kb \right) z + \lambda(z)(e^{Ka+g(z+a)-g(z)}-1) - bz\frac{\partial g}{\partial z},$$

117

which is continuous in z and also bounded on $z \in [0, \infty)$ since g is C^1 with compact support and $K > \frac{\theta}{a}$ and $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$. Hence there exists some $z^* \geq 0$ such that

$$
(4.56) \quad \begin{aligned}
&\frac{\theta b z}{a} + \lambda(z)(e^{f(z+a)-f(z)} - 1) - bz\frac{\partial f}{\partial z} \\
&= \frac{\theta b z^*}{a} + \lambda(z^*)(e^{f(z^*+a)-f(z^*)} - 1) - bz^*\frac{\partial f}{\partial z}\bigg|_{z=z^*}.
\end{aligned}
$$

Take a sequence of probability measures $\hat{\pi}_n$ such that it has probability density function n if $z \in [z^* - \frac{1}{2n}, z^* + \frac{1}{2n}]$ and 0 otherwise. Then, for every n, $\int z\hat{\pi}_n(dz) < \infty$. Therefore, we have

$$
(4.57) \quad \begin{aligned}
&\lim_{n\to\infty} \int \left[\frac{\theta b z}{a} + \lambda(z)(e^{f(z+a)-f(z)} - 1) - bz\frac{\partial f}{\partial z} \right] \hat{\pi}_n(dz) \\
&= \lim_{n\to\infty} n \int_{z^*-\frac{1}{2n}}^{z^*+\frac{1}{2n}} \left[\frac{\theta b z}{a} + \lambda(z)(e^{f(z+a)-f(z)} - 1) - bz\frac{\partial f}{\partial z} \right] dz \\
&= \frac{\theta b z^*}{a} + \lambda(z^*)(e^{f(z^*+a)-f(z^*)} - 1) - bz^*\frac{\partial f}{\partial z}\bigg|_{z=z^*} \\
&= \sup_{z\geq 0} \left[\frac{\theta b z}{a} + \lambda(z)(e^{f(z+a)-f(z)} - 1) - bz\frac{\partial f}{\partial z} \right].
\end{aligned}
$$

We conclude that the right hand side of (4.50) is greater or equal to the right hand side of (4.51).

Notice that for any $f = Kz + g + L \in \mathcal{F}$,

$$
(4.58) \quad \begin{aligned}
&\frac{\theta b z}{a} + \lambda(z)(e^{f(z+a)-f(z)} - 1) - bz\frac{\partial f}{\partial z} \\
&= \frac{b(\theta - Ka)}{a}z + \lambda(z)(e^{Ka+g(z+a)-g(z)} - 1) - bz\frac{\partial g}{\partial z},
\end{aligned}
$$

whose supremum is achieved at some finite $z^* > 0$ since $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$, $K > \frac{\theta}{a}$

and $g \in C^1$ with compact support. Hence $\int z\hat{\pi} < \infty$ is satisified for the optimal $\hat{\pi}$. This gives us (4.51). Finally, for any $f \in \mathcal{F}$, $u = e^f \in \mathcal{U}_\theta$, which implies (4.53). $\quad\square$

Lemma 24. *Assume* $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$, *we have* $\mathbb{E}[e^{\theta N_t}] < \infty$ *for any* $\theta \in \mathbb{R}$.

Proof. Observe that for any $\gamma \in \mathbb{R}$,

$$(4.59) \qquad \exp\left\{\gamma N_t - \int_0^t (e^\gamma - 1)\lambda(Z_s)ds\right\}$$

is a martinagle. Since $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$, for any $\epsilon > 0$, there exists a constant $C_\epsilon > 0$ such that $\lambda(z) \le C_\epsilon + \epsilon z$ for any $z \ge 0$. Also,

$$
\begin{aligned}
(4.60) \qquad \int_0^t Z_s ds &= \int_0^t \int_0^s h(s-u)N(du)ds \\
&= \int_0^t \left[\int_u^t h(s-u)ds\right] N(du) \\
&\le \int_0^t \left[\int_u^\infty h(s-u)ds\right] N(du) = \|h\|_{L^1} N_t.
\end{aligned}
$$

Therefore, for any $\gamma > 0$,

$$
\begin{aligned}
(4.61) \qquad 1 &= \mathbb{E}\left[e^{\gamma N_t - \int_0^t (e^\gamma - 1)\lambda(Z_s)ds}\right] \\
&\ge \mathbb{E}\left[e^{\gamma N_t - (e^\gamma - 1)\int_0^t (C_\epsilon + \epsilon Z_s)ds}\right] \\
&\ge \mathbb{E}\left[e^{\gamma N_t - (e^\gamma - 1)C_\epsilon t - (e^\gamma - 1)\epsilon\|h\|_{L^1} N_t}\right].
\end{aligned}
$$

For any $\theta > 0$, choose $\gamma > \theta$ and ϵ small enough so that $\gamma - (e^\gamma - 1)\epsilon\|h\|_{L^1} \ge \theta$. Then,

$$(4.62) \qquad \mathbb{E}\left[e^{\theta N_t}\right] \le e^{(e^\gamma - 1)C_\epsilon t} < \infty.$$

□

Now, we are ready to prove the large deviations result.

Theorem 15. *Assume* $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$ *and that* $\lambda(\cdot)$ *is continuous and bounded below by some positive constant. Then,* $(\frac{N_t}{t} \in \cdot)$ *satisfies the large deviation principle with the rate function* $I(\cdot)$ *as the Fenchel-Legendre transform of* $\Gamma(\cdot)$,

$$(4.63) \qquad I(x) = \sup_{\theta \in \mathbb{R}} \{\theta x - \Gamma(\theta)\}.$$

Proof. If $\limsup_{z\to\infty} \frac{\lambda(z)}{z} = 0$, then the forthcoming Lemma 26 implies that $\Gamma(\theta) < \infty$ for any θ. Thus, by Gärtner-Ellis Theorem, we have the upper bound. For Gärtner-Ellis Theorem and a general theory of large deviations, see for example [30]. To prove the lower bound, it suffices to show that for any $x > 0$, $\epsilon > 0$, we have

$$(4.64) \qquad \liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq -\sup_{\theta}\{\theta x - \Gamma(\theta)\},$$

where $B_\epsilon(x)$ denotes the open ball centered at x with radius ϵ. Let $\hat{\mathbb{P}}$ denote the tilted probability measure with rate $\hat{\lambda}$ defined in Theorem 14. By Jensen's

inequality,

$$(4.65) \quad \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right)$$

$$= \frac{1}{t} \log \int_{\frac{N_t}{t} \in B_\epsilon(x)} \frac{d\mathbb{P}}{d\hat{\mathbb{P}}} d\hat{\mathbb{P}}$$

$$= \frac{1}{t} \log \hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) + \frac{1}{t} \log \left[\frac{1}{\hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right)} \int_{\frac{N_t}{t} \in B_\epsilon(x)} \frac{d\mathbb{P}}{d\hat{\mathbb{P}}} d\hat{\mathbb{P}}\right]$$

$$\geq \frac{1}{t} \log \hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) - \frac{1}{\hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right)} \cdot \frac{1}{t}\hat{\mathbb{E}}\left[1_{\frac{N_t}{t} \in B_\epsilon(x)} \log \frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\right].$$

By the ergodic theorem,

$$(4.66) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq -\Lambda(x),$$

where

$$(4.67) \qquad \Lambda(x) = \inf_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^x} \left\{\int (\lambda - \hat{\lambda})\hat{\pi} + \int \log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi}\right\},$$

and

$$(4.68) \qquad \mathcal{Q}_e^x = \left\{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e : \int \hat{\lambda}(z)\hat{\pi}(dz) = x\right\}.$$

Notice that

$$(4.69) \qquad \Gamma(\theta) = \sup_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e} \left\{ \int \theta \hat{\lambda} \hat{\pi} + \int (\hat{\lambda} - \lambda) \hat{\pi} - \int \log(\hat{\lambda}/\lambda) \hat{\lambda} \hat{\pi} \right\}$$

$$= \sup_x \sup_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^x} \left\{ \int \theta \hat{\lambda} \hat{\pi} + \int (\hat{\lambda} - \lambda) \hat{\pi} - \int \log(\hat{\lambda}/\lambda) \hat{\lambda} \hat{\pi} \right\}$$

$$= \sup_x \sup_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^x} \left\{ \int \frac{\theta b}{a} z \hat{\pi}(dz) + \int (\hat{\lambda} - \lambda) \hat{\pi} - \int \log(\hat{\lambda}/\lambda) \hat{\lambda} \hat{\pi} \right\}$$

$$= \sup_x \{ \theta x - \Lambda(x) \}.$$

We prove in Lemma 25 that $\Lambda(x)$ is convex in x, identify it as the convex conjugate of $\Gamma(\theta)$ and thus conclude the proof. $\qquad \square$

Lemma 25. $\Lambda(x)$ *in* (4.67) *is convex in* x.

Proof. Define

$$(4.70) \qquad \hat{H}(\hat{\lambda}, \hat{\pi}) = \int (\lambda - \hat{\lambda}) \hat{\pi} + \int \log(\hat{\lambda}/\lambda) \hat{\lambda} \hat{\pi}.$$

Then,

$$(4.71) \qquad \Lambda(x) = \inf_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^x} \hat{H}(\hat{\lambda}, \hat{\pi}).$$

We want to prove that $\Lambda(\alpha x_1 + \beta x_2) \leq \alpha \Lambda(x_1) + \beta \Lambda(x_2)$ for any $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$. For any $\epsilon > 0$, we can choose $(\hat{\lambda}_k, \hat{\pi}_k) \in \mathcal{Q}_e^{x_k}$ such that $\hat{H}(\hat{\lambda}_k, \hat{\pi}_k) \leq \Lambda(x_k) + \epsilon/2$, for $k = 1, 2$. Set

$$(4.72) \qquad \hat{\pi}_3 = \alpha \hat{\pi}_1 + \beta \hat{\pi}_2, \quad \hat{\lambda}_3 = \frac{d(\alpha \hat{\pi}_1)}{d(\alpha \hat{\pi}_1 + \beta \hat{\pi}_2)} \hat{\lambda}_1 + \frac{d(\beta \hat{\pi}_2)}{d(\alpha \hat{\pi}_1 + \beta \hat{\pi}_2)} \hat{\lambda}_2.$$

Then for any test function f,

$$(4.73) \qquad \int \hat{\mathcal{A}}_3 f \hat{\pi}_3 = \alpha \int \hat{\mathcal{A}}_1 f \hat{\pi}_1 + \beta \int \hat{\mathcal{A}}_2 f \hat{\pi}_2 = 0,$$

which implies $(\hat{\lambda}_3, \hat{\pi}_3) \in \mathcal{Q}_e$. Furthermore,

$$(4.74) \qquad \int \hat{\lambda}_3 \hat{\pi}_3 = \alpha \int \hat{\lambda}_1 \hat{\pi}_1 + \beta \int \hat{\lambda}_2 \hat{\pi}_2 = \alpha x_1 + \beta x_2.$$

Therefore, $(\hat{\lambda}_3, \hat{\pi}_3) \in \mathcal{Q}_e^{\alpha x_1 + \beta x_2}$. Finally, since $x \log x$ is a convex function and if we apply Jensen's inequality, we get

$$(4.75) \quad \hat{H}(\hat{\lambda}_3, \hat{\pi}_3) = \int \left[(\lambda - \hat{\lambda}_3 - \hat{\lambda}_3 \log \lambda) + \hat{\lambda}_3 \log \hat{\lambda}_3 \right] \hat{\pi}_3$$

$$\leq \int \left[(\lambda - \hat{\lambda}_3 - \hat{\lambda}_3 \log \lambda) + \alpha \frac{d\hat{\pi}_1}{d\hat{\pi}_3} \hat{\lambda}_1 \log \hat{\lambda}_1 + \beta \frac{d\hat{\pi}_2}{d\hat{\pi}_3} \hat{\lambda}_2 \log \hat{\lambda}_2 \right] \hat{\pi}_3$$

$$= \alpha \hat{H}(\hat{\lambda}_1, \hat{\pi}_1) + \beta \hat{H}(\hat{\lambda}_2, \hat{\pi}_2).$$

Therefore,

(4.76)

$$\Lambda(\alpha x_1 + \beta x_2) \leq \hat{H}(\hat{\lambda}_3, \hat{\pi}_3) \leq \alpha \hat{H}(\hat{\lambda}_1, \hat{\pi}_1) + \beta \hat{H}(\hat{\lambda}_2, \hat{\pi}_2) \leq \alpha \Lambda(x_1) + \beta \Lambda(x_2) + \epsilon.$$

$\square$

Lemma 26. *If* $\limsup_{z \to \infty} \frac{\lambda(z)}{bz} < \frac{1}{a}$, *then for any*

$$(4.77) \qquad \theta < \log \left(\frac{b}{a \limsup_{z \to \infty} \frac{\lambda(z)}{z}} \right) - 1 + \frac{a}{b} \cdot \limsup_{z \to \infty} \frac{\lambda(z)}{z},$$

we have $\Gamma(\theta) < \infty$. *If* $\limsup_{z \to \infty} \frac{\lambda(z)}{z} = 0$, *then* $\Gamma(\theta) < \infty$ *for any* $\theta \in \mathbb{R}$.

Proof. For $K \geq \frac{\theta}{a}$, we have $e^{Kz} \in \mathcal{U}_\theta$ and

$$(4.78) \qquad \Gamma(\theta) \leq \inf_{g \in \mathcal{U}_\theta} \sup_{z \geq 0} \frac{\mathcal{A}g(z) + \frac{\theta b}{a} z g(z)}{g(z)} \leq \sup_{z \geq 0} \left\{ \frac{\mathcal{A}e^{Kz}}{e^{Kz}} + \frac{\theta b}{a} z \right\}$$

$$= \sup_{z \geq 0} \left\{ - \left(bK - \frac{\theta b}{a} \right) z + \lambda(z)(e^{Ka} - 1) \right\}.$$

Define the function

$$(4.79) \qquad F(K) = -K + \limsup_{z \to \infty} \frac{\lambda(z)}{bz} \cdot (e^{Ka} - 1).$$

Then $F(0) = 0$, F is convex and $F(K) \to \infty$ as $K \to \infty$ and its minimum is attained at

$$(4.80) \qquad K^* = \frac{1}{a} \log \left(\frac{b}{a \limsup_{z \to \infty} \frac{\lambda(z)}{z}} \right) > 0,$$

and $F(K^*) < 0$. Therefore, $\Gamma(\theta) < \infty$ for any

$$(4.81) \qquad \theta < -a \min_{K > 0} \left\{ -K + \limsup_{z \to \infty} \frac{\lambda(z)}{bz} \cdot (e^{Ka} - 1) \right\}$$

$$= \log \left(\frac{b}{a \limsup_{z \to \infty} \frac{\lambda(z)}{z}} \right) - 1 + \frac{a}{b} \cdot \limsup_{z \to \infty} \frac{\lambda(z)}{z} < K^* a.$$

If $\limsup_{z \to \infty} \frac{\lambda(z)}{z} = 0$, trying $e^{Kz} \in \mathcal{U}_\theta$ for any $K > \frac{\theta}{a}$, we have $\Gamma(\theta) < \infty$ for any θ. $\qquad \square$

4.3 Large Deviations for Markovian Nonlinear Hawkes Processes with Sum of Exponentials Exciting Function

In this section, we consider the Markovian nonlinear Hawkes processes with sum of exponentials exciting functions, i.e. $h(t) = \sum_{i=1}^{d} a_i e^{-b_i t}$. Let

$$(4.82) \qquad Z_i(t) = \sum_{\tau_j < t} a_i e^{-b_i(t-\tau_j)}, \quad 1 \leq i \leq d,$$

and $Z_t = \sum_{i=1}^{d} Z_i(t) = \sum_{\tau_j < t} h(t - \tau_j)$, where τ_j's are the arrivals of the Hawkes process with intensity $\lambda(Z_t) = \lambda(Z_1(t) + \cdots + Z_d(t))$ at time t. Observe that this is a special case of the Markovian processes with jumps studied in Section 4.1 with $\lambda(Z_1(t), Z_2(t), \cdots, Z_d(t))$ taking the form $\lambda(\sum_{i=1}^{d} Z_i(t))$. It is easy to see that $(Z_1, \ldots, Z_d)$ is Markovian with generator

$$(4.83) \quad \mathcal{A}f = -\sum_{i=1}^{d} b_i z_i \frac{\partial f}{\partial z_i} + \lambda \left(\sum_{i=1}^{d} z_i \right) \cdot [f(z_1 + a_1, \ldots, z_d + a_d) - f(z_1, \ldots, z_d)].$$

Here $b_i > 0$ for any $1 \leq i \leq d$ and a_i can be negative. But we restrict ourselves to the set of b_i's and a_i's so that $h(t) = \sum_{i=1}^{d} a_i e^{-b_i t} > 0$ for any $t \geq 0$ for the rest of this paper. In particular, $h(0) = \sum_{i=1}^{d} a_i > 0$. If $a_i > 0$, then $Z_i(t) \geq 0$ almost surely; if $a_i < 0$, then $Z_i(t) \leq 0$ almost surely.

Theorem 16. *Assume* $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$, $\lambda(\cdot)$ *is continuous and bounded below by*

a positive constant. Then,

$$(4.84) \qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \inf_{u \in \mathcal{U}_\theta} \sup_{(z_1,\dots,z_d) \in \mathcal{Z}} \left\{ \frac{\mathcal{A}u}{u} + \frac{\theta}{\sum_{i=1}^d a_i} \sum_{i=1}^d b_i z_i \right\},$$

where $\mathcal{Z} = \{(z_1,\dots,z_d) : a_i z_i \geq 0, 1 \leq i \leq d\}$ *and*

$$(4.85) \qquad \mathcal{U}_\theta = \left\{ u \in C_1(\mathbb{R}^d, \mathbb{R}^+), u = e^f, f \in \mathcal{F} \right\},$$

where

$$(4.86) \qquad \mathcal{F} = \left\{ f = g + \frac{\theta \sum_{i=1}^d z_i}{\sum_{i=1}^d a_i} + L, L \in \mathbb{R}, g \in \mathcal{G} \right\},$$

where

$$(4.87) \qquad \mathcal{G} = \left\{ \sum_{i=1}^d K \epsilon_i z_i + g, K > 0, g \text{ is } C_1 \text{ with compact support} \right\}.$$

Proof. Notice that

$$(4.88) \qquad dZ_i(t) = -b_i Z_i(t)dt + a_i dN_t, \quad 1 \leq i \leq d.$$

Hence, $a_i N_t = Z_i(t) - Z_i(0) + \int_0^t b_i Z_i(s)ds$ and

$$(4.89) \quad \mathbb{E}[e^{\theta N_t}] = \mathbb{E}\left[\exp \left\{ \frac{\theta \sum_{i=1}^d Z_i(t) - Z_i(0)}{\sum_{i=1}^d a_i} + \frac{\theta}{\sum_{i=1}^d a_i} \int_0^t \sum_{i=1}^d b_i Z_i(s)ds \right\} \right].$$

Following the same arguments in the proof of Theorem 14, we obtain the upper

126

bound

$$(4.90) \qquad \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \leq \inf_{u \in \mathcal{U}_\theta} \sup_{(z_1,\dots,z_d) \in \mathcal{Z}} \left\{ \frac{\mathcal{A}u}{u} + \frac{\theta}{\sum_{i=1}^d a_i} \sum_{i=1}^d b_i z_i \right\}.$$

As before, we can obtain the lower bound

$$(4.91) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$$

$$\geq \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e} \int \left[\theta \hat{\lambda} - \lambda + \hat{\lambda} - \hat{\lambda} \log \left(\hat{\lambda}/\lambda \right) \right] \hat{\pi}(dz_1, \dots, dz_d)$$

$$\geq \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}} \inf_{g \in \mathcal{G}} \int \left[\theta \hat{\lambda} - \lambda + \hat{\lambda} - \hat{\lambda} \log \left(\hat{\lambda}/\lambda \right) + \hat{\mathcal{A}} g \right] \hat{\pi}$$

$$= \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}} \inf_{f \in \mathcal{F}} \int \left[\frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} - \lambda + \hat{\lambda} - \hat{\lambda} \log \left(\hat{\lambda}/\lambda \right) + \hat{\mathcal{A}} f \right] \hat{\pi}.$$

The equality in the last line above holds by taking $f = g + L + \frac{\theta \sum_{i=1}^d z_i}{\sum_{i=1}^d a_i} \in \mathcal{F}$ for $g \in \mathcal{G}$, where

$$(4.92) \qquad \mathcal{G} = \left\{ \sum_{i=1}^d K \epsilon_i z_i + g, K > 0, g \text{ is } C_1 \text{ with compact support} \right\}.$$

Here, $\epsilon_i = a_i/|a_i|$, $1 \leq i \leq d$. Define

$$(4.93) \qquad F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) = \int \left[\frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} + \hat{\mathcal{A}} f \right] \hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}).$$

F is linear in f and hence convex in f. Also $\hat{H}$ is weakly lower semicontinuous and convex in $(\hat{\lambda}\hat{\pi}, \hat{\pi})$. Therefore, F is concave in $(\hat{\lambda}\hat{\pi}, \hat{\pi})$. Furthermore, for any

$$f = \frac{\theta \sum_{i=1}^{d} z_i}{\sum_{i=1}^{d} a_i} + \sum_{i=1}^{d} K\epsilon_i z_i + g + L \in \mathcal{F},$$

$$(4.94) \quad F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) = \int \left[\theta + \sum_{i=1}^{d} K\epsilon_i a_i \right] \hat{\lambda}\hat{\pi} - \int \sum_{i=1}^{d} K\epsilon_i b_i z_i \hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) + \int \hat{\mathcal{A}}g\hat{\pi}.$$

If $\lambda_n \pi_n \to \gamma_\infty$ and $\pi_n \to \pi_\infty$ weakly, then, since g is C_1 with compact support, we have

$$(4.95) \quad \int \left[\theta + \sum_{i=1}^{d} K\epsilon_i a_i \right] \lambda_n \pi_n + \int \hat{\mathcal{A}}g\pi_n \to \int \left[\theta + \sum_{i=1}^{d} K\epsilon_i a_i \right] \gamma_\infty + \int \hat{\mathcal{A}}g\pi_\infty.$$

Since $-\sum_{i=1}^{d} K\epsilon_i b_i z_i$ is continuous and nonpositive on $\mathcal{Z}$, we have

$$(4.96) \quad \limsup_{n \to \infty} \int \left[-\sum_{i=1}^{d} K\epsilon_i b_i z_i \right] \pi_n \leq \int \left[-\sum_{i=1}^{d} K\epsilon_i b_i z_i \right] \pi_\infty.$$

Hence, we conclude that F is upper semicontinuous in the weak topology.

In order to apply the minmax theorem, we want to prove the compactness in the weak topology of the level set

$$(4.97) \quad \left\{ (\hat{\lambda}\hat{\pi}, \hat{\pi}) : \int \left[-\frac{\theta \sum_{i=1}^{d} b_i z_i}{\sum_{i=1}^{d} a_i} - \hat{\mathcal{A}}f \right] \hat{\pi} + \hat{H}(\hat{\lambda}, \hat{\pi}) \leq C \right\}.$$

For any $f = \frac{\theta \sum_{i=1}^{d} z_i}{\sum_{i=1}^{d} a_i} + \sum_{i=1}^{d} K\epsilon_i z_i + g + L \in \mathcal{F}$, where g is C_1 with compact support

etc., there exist some $C_1, C_2 > 0$ such that

$$(4.98) \quad C_1 \geq \hat{H} + \sum_{i=1}^{d} Kb_i\epsilon_i \int z_i\hat{\pi} - C_2 \int \hat{\lambda}\hat{\pi}$$

$$\geq \int_{\hat{\lambda} \geq \sum_{i=1}^{d} c_i z_i + \ell} \left[\lambda - \hat{\lambda} + \hat{\lambda} \log(\hat{\lambda}/\lambda) \right] \hat{\pi} + \sum_{i=1}^{d} Kb_i\epsilon_i \int z_i\hat{\pi}$$

$$- C_2 \int_{\hat{\lambda} \geq \sum_{i=1}^{d} c_i z_i + \ell} \hat{\lambda}\hat{\pi} - C_2 \int_{\hat{\lambda} < \sum_{i=1}^{d} c_i z_i + \ell} \hat{\lambda}\hat{\pi}$$

$$\geq \left[\min_{(z_1, \ldots, z_d) \in \mathcal{Z}} \log \frac{c_1 z_1 + \cdots + c_d z_d + \ell}{\lambda(z_1 + \cdots + z_d)} - 1 - C_2 \right] \int_{\hat{\lambda} \geq \sum_{i=1}^{d} c_i z_i + \ell} \hat{\lambda}\hat{\pi}$$

$$+ \sum_{i=1}^{d} [-c_i \cdot C_2 + Kb_i\epsilon_i] \int z_i\hat{\pi} - \ell C_2.$$

If $a_i > 0$, then $\epsilon_i > 0$, pick up $c_i > 0$ such that $-c_i \cdot C_2 + Kb_i\epsilon_i > 0$. If $a_i < 0$, then $\epsilon_i < 0$, pick up c_i such that $-c_i \cdot C_2 + Kb_i\epsilon_i < 0$. Finally, choose ℓ big enough such that the big bracket above is positive. Then

$$(4.99) \quad \int |z_i|\hat{\pi} \leq C_3, \quad \int_{\hat{\lambda} \geq \sum_{i=1}^{d} c_i z_i + \ell} \hat{\lambda}\hat{\pi} \leq C_4.$$

Hence, $\int \hat{\lambda}\hat{\pi} \leq C_5$ and $\hat{H} \leq C_6$. We can use the similar method as in the proof of Theorem 14 to show that

$$(4.100) \quad \lim_{\ell \to \infty} \sup_{n} \int_{|z_i| > \ell} \lambda_n\pi_n = 0, \quad 1 \leq i \leq d.$$

For any $(\lambda_n\pi_n, \pi_n) \in \mathcal{R}$, we can find a subsequence that converges in the weak

topology by Prokhorov's Theorem. Therefore,

(4.101)

$$
\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]
$$

$$
\geq \sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}} \inf_{f\in\mathcal{F}} \int \left[\frac{\theta \sum_{i=1}^{d} b_i z_i}{\sum_{i=1}^{d} a_i} - \lambda + \hat{\lambda} - \hat{\lambda}\log\left(\hat{\lambda}/\lambda\right) + \hat{\mathcal{A}}f \right] \hat{\pi}
$$

$$
= \inf_{f\in\mathcal{F}} \sup_{\hat{\pi}} \sup_{\hat{\lambda}} \int \left[\frac{\theta \sum_{i=1}^{d} b_i z_i}{\sum_{i=1}^{d} a_i} - \lambda + \hat{\lambda} - \hat{\lambda}\log\left(\hat{\lambda}/\lambda\right) + \hat{\mathcal{A}}f \right] \hat{\pi}
$$

$$
= \inf_{f\in\mathcal{F}} \sup_{(z_1,\ldots,z_d)\in\mathcal{Z}} \frac{\theta \sum_{i=1}^{d} b_i z_i}{\sum_{i=1}^{d} a_i} + \lambda(e^{f(z_1+a_1,\ldots,z_d+a_d)-f(z_1,\ldots,z_d)} - 1) - \sum_{i=1}^{d} b_i z_i \frac{\partial f}{\partial z_i}
$$

$$
\geq \inf_{u\in\mathcal{U}_\theta} \sup_{(z_1,\ldots,z_d)\in\mathcal{Z}} \left\{ \frac{\mathcal{A}u}{u} + \frac{\theta}{\sum_{i=1}^{d} a_i} \sum_{i=1}^{d} b_i z_i \right\}.
$$

That is because optimizing over $\hat{\lambda}$, we get $\hat{\lambda} = \lambda e^{f(z_1+a_1,\ldots,z_d+a_d)-f(z_1,\ldots,z_d)}$ and finally for each $f \in \mathcal{F}$, $u = e^f \in \mathcal{U}_\theta$. $\qquad\square$

Theorem 17. *Assume* $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$, $\lambda(\cdot)$ *is continuous and bounded below by some positive constant. Then,* $(\frac{N_t}{t} \in \cdot)$ *satisfies the large deviation principle with the rate function* $I(\cdot)$ *as the Fenchel-Legendre transform of* $\Gamma(\cdot)$,

(4.102)
$$
I(x) = \sup_{\theta\in\mathbb{R}} \left\{ \theta x - \Gamma(\theta) \right\},
$$

where

(4.103)
$$
\Gamma(\theta) = \sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e} \int \left[\theta\hat{\lambda} - \lambda + \hat{\lambda} - \hat{\lambda}\log\left(\hat{\lambda}/\lambda\right) \right] \hat{\pi}.
$$

Proof. The proof is the same as in the case of exponential $h(\cdot)$. $\qquad\square$

4.4 Large Deviations for a Special Class of Nonlinear Hawkes Processes: An Approximation Approach

We already proved in Chapter 3 a large deviation principle of $(N_t/t \in \cdot)$ for nonlinear Hawkes process by proving a level-3 large deviation first and then applying the contraction principle. In this section, we point out that there is an alternative approach, i.e. for general exciting function $h(\cdot)$, we can use sums of exponential functions to approximate $h(\cdot)$ and use the large deviations for the case when $h(\cdot)$ is a sum of exponentials to obtain the large deviations for general $h(\cdot)$. The advantage of approximating the general case by the case when h is a sum of exponentials is that the rate function for the large deviations when h is a sum of exponentials can be evaluated by an optimization problem, which should be computable by some numerical scheme.

Before we proceed, let us first prove that h can be approximated by a sum of exponentials in both L_1 and L_∞ norms.

Lemma 27. *If $h(t) > 0$, $\int_0^\infty h(t)dt < \infty$, $h(\infty) = 0$, and h is continuous, then h can be approximated by a sum of exponentials both in L^1 and L^∞ norms.*

Proof. The Stone-Weierstrass theorem says that if X is a compact Hausdorff space and suppose A is a subspace of $C(X)$ with the following properties. (i) If $f, g \in A$, then $f \times g \in A$. (ii) $1 \in A$. (iii) If $x, y \in X$ then we can find an $f \in A$ such that $f(x) \neq f(y)$. Then A is dense in $C(X)$ in L^∞ norm. Consider $X = \mathbb{R}^+ \cup \{\infty\} = [0, \infty]$ and $C[0, \infty]$ consists of continuous functions vanishing at ∞ and the constant function 1.

By Stone-Weierstrass theorem, the linear combination of 1, e^{-t}, e^{-2t} etc. is dense in $C[0, \infty]$. In other words, for any continuous function h on $C[0, \infty]$, we have

$$(4.104) \qquad \sup_{t \geq 0} \left| h(t) - \sum_{j=0}^{n} a_j e^{-jt} \right| \leq \epsilon.$$

In fact, since $h(\infty) = 0$, we get $|a_0| \leq \epsilon$. Thus

$$(4.105) \qquad \sup_{t \geq 0} \left| h(t) - \sum_{j=1}^{n} a_j e^{-jt} \right| \leq 2\epsilon.$$

However, $\sum_{j=1}^{n} a_j e^{-jt}$ may not be positive. We can approximate $\sqrt{h(t)}$ first by a sum of exponentials and then approximate $h(t)$ by the square of that sum of exponentials, which is again a sum of exponentials but positive this time.

Indeed, we can approximate $h(t)$ by the sum of exponentials in L^1 norm as well. Suppose $\|h - h_n\|_{L^\infty} \to 0$, where h_n is a sum of exponentials. Then, by dominated convergence theorem, for any $\delta > 0$, $\int |h - h_n| e^{-\delta t} dt \to 0$ as $n \to \infty$. Thus, we can find a sequence $\delta_n > 0$ such that $\delta_n \to 0$ as $n \to \infty$ and $\int |h - h_n| e^{-\delta_n t} dt \to 0$. By dominated convergence theorem again, $\int h(1 - e^{-\delta_n t}) dt \to 0$. Hence, we have $\int |h - h_n e^{-\delta_n t}| dt \to 0$ as $n \to \infty$, where $h_n e^{-\delta_n t}$ is a sum of exponentials.

We will show that $h_n e^{-\delta_n t}$ converges to h in L^∞ as well.

$$(4.106) \qquad \|h - h_n e^{-\delta_n t}\|_{L^\infty} \leq \|h - h_n\|_{L^\infty} + \|h_n - h_n e^{-\delta_n t}\|_{L^\infty}.$$

Notice that $(1 - e^{-\delta_n t}) h_n \leq (1 - e^{-\delta_n t})(h(t) + \epsilon)$. Since $h(\infty) = 0$, there exists some $M > 0$, such that for $t > M$, $h(t) \leq \epsilon$ so that $(1 - e^{-\delta_n t}) h_n \leq 2\epsilon$ for $t > M$. For $t \leq M$, $(1 - e^{-\delta_n t}) h_n \leq (1 - e^{-\delta_n M})(\|h\|_{L^\infty} + \epsilon)$ which is small if δ_n is small. $\qquad \square$

We have the following results.

Theorem 18. *Assume that $\lambda(\cdot) \geq c$ for some $c > 0$, $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$ and $\lambda(\cdot)^\alpha$ is Lipschitz with constant L_α for any $\alpha \geq 1$. We have $(N_t/t \in \cdot)$ satisfies the large deviation principle with the rate function*

$$(4.107) \qquad I(x) = \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma(\theta)\}.$$

Remark 5. *The class of nonlinear Hawkes process with general exciting function h for which we proved the large deviation principle here is unfortunately a bit too special. It works for the rate function like $\lambda(z) = [\log(c+z)]^\beta$ for example but does not work for $\lambda(\cdot)$ that has sublinear power law growth.*

We end this chapter with the proof of Theorem 18.

Let P_n denote the probability measure under which N_t follows the Hawkes process with exciting function $h_n = \sum_{i=1}^n a_i e^{-b_i t}$ such that $h_n \to h$ as $n \to \infty$ in both L^1 and L^∞ norms. We can find such a sequence h_n by Lemma 27. Let us define

$$(4.108) \qquad \Gamma_n(\theta) = \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}^{P_n}\left[e^{\theta N_t}\right].$$

We need the following lemmas to prove Theorem 18.

Lemma 28. *For any $K > 0$ and $\theta_1, \theta_2 \in [-K, K]$, there exists some constant $C(K)$ such that for any n,*

$$(4.109) \qquad |\Gamma_n(\theta_1) - \Gamma_n(\theta_2)| \leq C(K)|\theta_1 - \theta_2|.$$

Proof. Without loss of generality, take $\theta_2 > \theta_1$. Then

(4.110)
$$\Gamma_n(\theta_1) \leq \Gamma_n(\theta_2)$$

$$= \sup_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^*} \int (\theta_2 - \theta_1)\hat{\lambda}\hat{\pi} + \theta_1\hat{\lambda}\hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi})$$

$$\leq \sup_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^*} \int (\theta_2 - \theta_1)\hat{\lambda}\hat{\pi} + \Gamma_n(\theta_1),$$

where

(4.111)
$$\mathcal{Q}_e^* = \left\{ (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e : \int \theta_1\hat{\lambda}\hat{\pi} - \hat{H}(\hat{\lambda}, \hat{\pi}) \geq \Gamma_n(\theta_1) - 1 \right\}.$$

The key is to prove that $\sup_{(\hat{\lambda},\hat{\pi}) \in \mathcal{Q}_e^*} \int \hat{\lambda}\hat{\pi} \leq C(K)$ for some positive constant $C(K)$ depending only on K. Define $u = u(z_1, \ldots, z_n) = e^{\sum_{i=1}^n c_i z_i}$ where

(4.112)
$$c_i = \frac{3K}{\sum_{i=1}^n \frac{a_i}{b_i}} \cdot \frac{1}{b_i}, \quad 1 \leq i \leq n.$$

Define $V = -\frac{\mathcal{A}u}{u}$ such that

(4.113)
$$V(z_1, \ldots, z_n) = \frac{3K}{\sum_{i=1}^n \frac{a_i}{b_i}} \sum_{i=1}^n z_i - \lambda(z_1 + \cdots + z_n)(e^{3K} - 1).$$

Notice that $\int \hat{\mathcal{A}}f\hat{\pi} = 0$ for any test function f with certain regularities. If we try $f = \frac{z_i}{b_i}$, $1 \leq i \leq n$, we get

(4.114)
$$-\int z_i\hat{\pi} + \frac{a_i}{b_i}\int \hat{\lambda}\hat{\pi} = 0, \quad 1 \leq i \leq n.$$

134

Summing over $1 \leq i \leq n$, we get

$$(4.115) \qquad \int \hat{\lambda} \hat{\pi} = \frac{1}{\sum_{i=1}^{n} \frac{a_i}{b_i}} \int \sum_{i=1}^{n} z_i \hat{\pi}.$$

Notice that $\sum_{i=1}^{n} \frac{a_i}{b_i} = \|h_n\|_{L^1}$ which is approximately $\|h\|_{L^1}$ when n is large. Since $\limsup_{z \to \infty} \frac{\lambda(z)}{z} = 0$ and $\sum_{i=1}^{n} z_i \geq 0$, we have

$$(4.116) \qquad \theta_1 \int \hat{\lambda} \hat{\pi} \leq K \int \hat{\lambda} \hat{\pi} = \frac{K}{\sum_{i=1}^{n} \frac{a_i}{b_i}} \int \sum_{i=1}^{n} z_i \hat{\pi} \leq \frac{1}{2} \int V \hat{\pi} + C_{1/2}(K),$$

where $C_{1/2}(K)$ is some positive constant depending only on K.

We claim that $\int V(z) \hat{\pi} \leq \hat{H}(\hat{\pi})$ for any $\hat{\pi} \in \mathcal{Q}_e^*$. Let us prove it. By the ergodic theorem and Jensen's inequality,

$$(4.117)$$
$$\int V(z) \hat{\pi} = \lim_{t \to \infty} \mathbb{E}^{\hat{\pi}} \left[\frac{1}{t} \int_0^t V(Z_s) ds \right] \leq \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}^{\pi} \left[e^{\int_0^t V(Z_s) ds} \right] + \hat{H}(\hat{\pi}).$$

Next, we will show that $u \geq 1$. That is equivalent to proving $\sum_{i=1}^{n} \frac{z_i}{b_i} \geq 0$. Consider the process

$$(4.118) \qquad Y_t = \sum_{i=1}^{n} \frac{Z_i(t)}{b_i} = \sum_{\tau_j < t} \sum_{i=1}^{n} \frac{a_i}{b_i} e^{-b_i(t - \tau_j)} = \sum_{\tau_j < t} g(t - \tau_j),$$

where $g(t) = \sum_{i=1}^{n} \frac{a_i}{b_i} e^{-b_i t}$. Notice that $g(t) = \int_t^{\infty} h(s) ds > 0$. Therefore, $Y_t \geq 0$ almost surely and $\sum_{i=1}^{n} \frac{Z_i(t)}{b_i} \geq 0$. Since $\frac{Au}{u} + V = 0$ and $u \geq 1$, by Feynman-Kac

formula and Dynkin's formula,

$$(4.119) \quad \mathbb{E}^\pi \left[e^{\int_0^t V(Z_s)ds} \right] \leq \mathbb{E}^\pi \left[u(Z_t) e^{\int_0^t V(Z_s)ds} \right]$$

$$= u(Z_0) + \int_0^t \mathbb{E}^\pi \left[(\mathcal{A}u(Z_s) + V(Z_s)u(Z_s)) e^{\int_0^s V(Z_u)du} \right] ds$$

$$= u(Z_0),$$

and therefore $\int V(z)\hat{\pi} \leq \hat{H}(\hat{\pi})$ for any $\hat{\pi} \in \mathcal{Q}_e^*$. Hence,

$$(4.120) \qquad \theta_1 \int \hat{\lambda}\hat{\pi} \leq \frac{1}{2} \int V(z) + C_{1/2}(K) \leq \frac{1}{2}\hat{H} + C_{1/2}(K).$$

Notice that

$$(4.121) \qquad -\infty < \Gamma_n(\theta_1) - 1 \leq \theta_1 \int \hat{\lambda}\hat{\pi} - \hat{H} \leq \Gamma_n(\theta_1) < \infty.$$

Hence,

$$(4.122) \qquad \Gamma_n(\theta_1) - 1 + \frac{1}{2}\hat{H} \leq \theta_1 \int \hat{\lambda}\hat{\pi} - \frac{1}{2}\hat{H} \leq C_{1/2}(K),$$

which implies $\hat{H} \leq 2(C_{1/2}(K) - \Gamma_n(\theta_1) + 1)$ and so also,

$$(4.123) \quad \int \hat{\lambda}\hat{\pi} \leq \frac{1}{2K} \int V\hat{\pi} + \frac{1}{K}C_{1/2}(K) \leq \frac{1}{K}(C_{1/2}(K) - \Gamma_n(\theta_1) + 1) + \frac{1}{K}C_{1/2}(K).$$

Finally, notice that since $h_n \to h$ in both L^1 and L^∞ norms, we can find a function g such that $\sup_n h_n \leq g$ and $\|g\|_{L^1} < \infty$. and thus

$$(4.124) \qquad \Gamma_n(\theta_1) \geq \Gamma_n(-K) \geq \Gamma_g(-K),$$

136

where Γ_g denotes the case when the rate function is still $\lambda(\cdot)$ but the exciting function is $g(\cdot)$ instead of $h_n(\cdot)$. Notice that here $\|g\|_{L^1} < \infty$ but may not be less than 1. It is still well defined because of the assumption $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$. Indeed, we can find $\lambda(z) = \nu_\epsilon + \epsilon z$ that dominates the original $\lambda(\cdot)$ for $\nu_\epsilon > 0$ big enough and $\epsilon > 0$ small enough so that $\epsilon\|g\|_{L^1} < 1$. Now, we have $\Gamma_g(-K) \geq \Gamma_{\epsilon g}^{\nu_\epsilon}(-K)$ which is finite, where $\Gamma_{\epsilon g}^{\nu_\epsilon}(-K)$ corresponds to the case when $\lambda(z) = \nu_\epsilon + \epsilon z$. Hence,

$$(4.125) \qquad \sup_{(\hat\lambda,\hat\pi)\in\mathcal{Q}_e^*} \int \hat\lambda\hat\pi \leq C(K),$$

for some $C(K) > 0$ depending only on K. $\qquad\square$

Lemma 29. *Assume that $\lambda(\cdot) \geq c$ for some $c > 0$, $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$ and $\lambda(\cdot)^\alpha$ is Lipschitz with constant L_α for any $\alpha \geq 1$. Then for any $K > 0$, $\Gamma_n(\theta)$ is Cauchy with θ uniformly in $[-K, K]$.*

Proof. Let us write $H_n(t) = \sum_{\tau_j < t} h_n(t - \tau_j)$. Observe first, that for any q,

$$(4.126) \quad \exp\left\{ q\int_0^t \log\left(\frac{\lambda(H_m(s))}{\lambda(H_n(s))} \right) dN_s - \int_0^t \left(\frac{\lambda(H_m(s))^q}{\lambda(H_n(s))^{q-1}} - \lambda(H_n(s)) \right) ds \right\}$$

is a martingale under P_n. By Hölder's inequality, for any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$(4.127)$$

$$\mathbb{E}^{P_m}[e^{\theta N_t}] = \mathbb{E}^{P_n}\left[e^{\theta N_t} \frac{dP_m}{dP_n} \right]$$

$$= \mathbb{E}^{P_n}\left[e^{\theta N_t - \int_0^t (\lambda(H_m(s)) - \lambda(H_n(s)))ds - \int_0^t \log\left(\frac{\lambda(H_n(s))}{\lambda(H_m(s))} \right)dN_s} \right]$$

$$\leq \mathbb{E}^{P_n}\left[e^{p\theta N_t - p\int_0^t (\lambda(H_m(s)) - \lambda(H_n(s)))ds} \right]^{1/p} \mathbb{E}^{P_n}\left[e^{q\int_0^t \log\left(\frac{\lambda(H_m(s))}{\lambda(H_n(s))} \right)dN_s} \right]^{1/q}.$$

137

By the Cauchy-Schwarz inequality,

$$(4.128) \quad \mathbb{E}^{P_n}\left[e^{q\int_0^t \log\left(\frac{\lambda(H_m(s))}{\lambda(H_n(s))}\right)dN_s}\right]^{1/q} \leq \mathbb{E}^{P_n}\left[e^{\int_0^t \left(\frac{\lambda(H_m(s))^{2q}}{\lambda(H_n(s))^{2q-1}}-\lambda(H_n(s))\right)ds}\right]^{\frac{1}{2q}}$$

$$\leq \mathbb{E}^{P_n}\left[e^{\frac{1}{c^{2q-1}}L_{2q}\int_0^t \sum_{\tau<s}|h_m(s-\tau)-h_n(s-\tau)|ds}\right]^{\frac{1}{2q}}$$

$$\leq \mathbb{E}^{P_n}\left[e^{\frac{1}{c^{2q-1}}L_{2q}\|h_m-h_n\|_{L^1}N_t}\right]^{\frac{1}{2q}}.$$

We also have

$$(4.129) \quad \mathbb{E}^{P_n}\left[e^{p\theta N_t - p\int_0^t (\lambda(H_m(s))-\lambda(H_n(s)))ds}\right]^{1/p} \leq \mathbb{E}^{P_n}\left[e^{p\theta N_t + pL_1\|h_m-h_n\|_{L^1}N_t}\right]^{1/p}.$$

Therefore, by Lemma 28 and the fact $\Gamma_n(0) = 0$ for any n, we have

(4.130)

$$\Gamma_m(\theta) - \Gamma_n(\theta)$$

$$\leq \frac{1}{p}\Gamma_n\left(p\theta + pL_1\epsilon_{m,n}\right) + \frac{1}{2q}\Gamma_n\left(\frac{L_{2q}\epsilon_{m,n}}{c^{2q-1}}\right) - \Gamma_n(\theta)$$

$$\leq C(K)L_1\epsilon_{m,n} + \frac{C(K)}{2q}\cdot\frac{L_{2q}\epsilon_{m,n}}{c^{2q-1}} + \frac{1}{p}\Gamma_n(p\theta) - \frac{1}{p}\Gamma_n(\theta) + \left(1-\frac{1}{p}\right)|\Gamma_n(\theta)|,$$

$$\leq C(K)L_1\epsilon_{m,n} + \frac{C(K)}{2q}\cdot\frac{L_{2q}\epsilon_{m,n}}{c^{2q-1}} + \frac{C(K)(p-1)K}{p} + \left(1-\frac{1}{p}\right)C(K)K,$$

where $\epsilon_{m,n} = \|h_m - h_n\|_{L^1}$. Hence,

$$(4.131) \qquad \limsup_{m,n\to\infty}\{\Gamma_m(\theta) - \Gamma_n(\theta)\} \leq 2\left(1-\frac{1}{p}\right)C(K)K,$$

which is true for any $p > 1$. Letting $p \downarrow 1$, we get the desired result. $\qquad\square$

Remark 6. *If $\lambda(\cdot) \geq c > 0$ and $\lim_{z\to\infty}\frac{\lambda(z)}{z^\alpha} = 0$ for any $\alpha > 0$, then, $\lambda(\cdot)^\sigma$ is*

138

Lipschitz for any $\sigma \geq 1$. For instance, $\lambda(z) = [\log(z + c)]^\beta$ satisfies the conditions if $\beta > 0$ and $c > 1$.

Theorem 19. *Assume that $\lambda(\cdot) \geq c$ for some $c > 0$, $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$ and $\lambda(\cdot)^\alpha$ is Lipschitz with constant L_α for any $\alpha \geq 1$.*

$$(4.132) \qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \Gamma(\theta) = \lim_{n \to \infty} \Gamma_n(\theta),$$

for any $\theta \in \mathbb{R}$.

Proof. By Lemma 29, $\Gamma_n(\theta)$ tends to $\Gamma(\theta)$ uniformly on any compact set $[-K, K]$. Since $\Gamma_n(\theta)$ is Lipschitz by Lemma 28, it is continuous and the limit Γ is also continuous. Let $\epsilon_n = \|h_n - h\|_{L^1} \leq \epsilon$. As in the proof of Lemma 29, for any $\theta \in [-K, K]$, $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, we get

$$(4.133) \qquad \limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$$
$$\leq \Gamma_n(\theta) + C(K)L_1\epsilon_n + \frac{C(K)}{2q} \cdot \frac{L_{2q}\epsilon_n}{c^{2q-1}} + 2\left(1 - \frac{1}{p}\right)C(K)K.$$

Letting $n \to \infty$ first and then $p \downarrow 1$, we get $\limsup_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \leq \Gamma(\theta)$. Similarly, for any $p', q' > 1$ with $\frac{1}{p'} + \frac{1}{q'} = 1$,

$$(4.134) \qquad \Gamma_n(\theta) \leq \liminf_{t \to \infty} \frac{1}{pt} \log \mathbb{E}[e^{(p\theta + pL_1\epsilon_n)N_t}] + \liminf_{t \to \infty} \frac{1}{2qt} \log \mathbb{E}\left[e^{\frac{L_{2q}\epsilon_n}{c^{2q-1}}N_t}\right]$$
$$\leq \liminf_{t \to \infty} \frac{1}{pp't} \log \mathbb{E}[e^{pp'\theta N_t}] + \liminf_{t \to \infty} \frac{1}{pq't} \log \mathbb{E}[e^{q'pL_1\epsilon_n N_t}]$$
$$+ \liminf_{t \to \infty} \frac{1}{2qt} \log \mathbb{E}\left[e^{\frac{L_{2q}\epsilon_n}{c^{2q-1}}N_t}\right].$$

Since we can dominate $\lambda(\cdot)$ by the linear function $\lambda(z) = \nu + z$ in which case the limit of logarithmic moment generating function $\Gamma_\nu(\theta)$ is continuous in θ, we may

let $n \to \infty$ to obtain

$$(4.135) \qquad \Gamma(\theta) \leq \liminf_{t \to \infty} \frac{1}{pp't} \log \mathbb{E}[e^{pp'\theta N_t}].$$

This holds for any θ and thus

$$(4.136) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \geq pp'\Gamma\left(\frac{\theta}{pp'}\right).$$

Letting $p, p' \downarrow 1$ and using the continuity of $\Gamma(\cdot)$, we get the desired result. $\qquad \square$

Finally, let us prove Theorem 18.

Proof of Theorem 18. For the upper bound, apply the Gärtner-Ellis Theorem. Let us prove the lower bound. Let $B_\epsilon(x)$ denote the open ball centered at x with radius $\epsilon > 0$. By Hölder's inequality, for any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$(4.137) \qquad P_n\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \leq \left\|\frac{dP_n}{d\mathbb{P}}\right\|_{L^p(\mathbb{P})} \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right)^{1/q}.$$

Therefore, letting $t \to \infty$, we have

$$(4.138) \quad \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma_n(\theta)\} = \lim_{t \to \infty} \frac{1}{t} \log P_n\left(\frac{N_t}{t} \in B_\epsilon(x)\right)$$
$$\leq \frac{1}{pp'}\Gamma(pp'L_1\epsilon_n) + \frac{1}{2pq'}\Gamma\left(\frac{L_{2pq'}\epsilon_n}{c^{2pq'-1}}\right) + \frac{1}{q}\liminf_{t \to \infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right),$$

where $\epsilon_n = \|h_n - h\|_{L^1}$. Hence, letting $n \to \infty$, see that

$$(4.139) \qquad \frac{1}{q}\liminf_{t \to \infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq \limsup_{n \to \infty}\sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma_n(\theta)\}.$$

140

Since $\Gamma_n(\theta) \to \Gamma(\theta)$ uniformly on any compact set K,

(4.140)
$$\sup_{\theta \in K}\{\theta x - \Gamma_n(\theta)\} \to \sup_{\theta \in K}\{\theta x - \Gamma(\theta)\},$$

as $n \to \infty$ for any such set K. Notice that $\lambda(\cdot) \geq c > 0$ and recall that the limit for the logarithmic moment generating function with parameter θ for a Poisson process with constant rate c is $(e^\theta - 1)c$. Hence

(4.141)
$$\liminf_{\theta \to +\infty} \frac{\Gamma_n(\theta)}{\theta} \geq \liminf_{\theta \to +\infty} \frac{(e^\theta - 1)c}{\theta} = +\infty,$$

which implies that $\sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma_n(\theta)\} \to \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma(\theta)\}$. Therefore,

(4.142)
$$\frac{1}{q} \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma(\theta)\}.$$

Letting $q \downarrow 1$, we get the desired result. $\qquad\square$

Chapter 5

Asymptotics for Nonlinear

Hawkes Processes

In the existing literature of on nonlinear Hawkes processes, the usual assumption is that $\lambda(\cdot)$ is α-Lipschitz, $h(\cdot)$ is integrable and $\alpha\|h\|_{L^1} < 1$. But how about other regimes? How do the asymptotics vary in different regimes? This is the question we would try to answer in this chapter.

We divide the nonlinear Hawkes process into the following regimes.

1. $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$. This is the sublinear regime. In this regime, if we assume that $\lambda(\cdot)$ is α-Lipschitz, $\|h\|_{L^1} < \infty$ and $\alpha\|h\|_{L^1} < 1$, then there exists a unique stationary version of the nonlinear Hawkes process. The central limit theorem and large deviations for this regime are proved in Zhu [114], [112] and [113]. On the contrary, if we assume that $\|h\|_{L^1} = \infty$, then, there is no stationary version. Figure 5.1 illustrates λ_t in this case. We will obtain the time asymptotics for λ_t in Section 5.1.

2. $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 1$ and $\|h\|_{L^1} < 1$. This is the sub-critical regime. In this

regime, if we assume that $\lambda(\cdot)$ is α-Lipschitz and $\alpha \|h\|_{L^1} < 1$, then there exists a unique stationary version of the nonlinear Hawkes process, see Brémaud and Massoulié [14]. The central limit theorem is proved in Zhu [114]. Figure 5.3 illustrates λ_t in this case. We will summarize some known results about the limit theorems in Section 5.2.

3. $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 1$ and $\|h\|_{L^1} = 1$. This is the critical regime. This regime is very subtle. We will show in Section 5.3 that in some cases, there exists a stationary version of the Hawkes process. In some other cases, it does not exist. In particular, when $\lambda(z) = \nu + z$ and $\int_0^\infty th(t)dt < \infty$, we will prove that $\frac{N_{tT}}{T^2} \to \int_0^t \eta_s ds$, where η_s is a squared Bessel process. $N[T, T + \frac{t}{T}]$ will converge to a Pólya process as $T \to \infty$. Figure 5.4 illustrates the behavior of λ_t in this case. When $h(\cdot)$ has heavy tails, i.e. $\int_0^\infty th(t)dt = \infty$, we will prove that the time asymptotic behavior is different from the light tail case.

4. $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 1$ and $\|h\|_{L^1} > 1$. This is the super-critical regime. We will prove in Section 5.4 that λ_t grows exponentially in t in this regime, which is consistent with what we can see in Figure 5.5.

5. $\sum_{n=0}^\infty \frac{1}{\lambda(n)} < \infty$. This is the explosive regime. In Section 5.5, we will first provide a criterion for the explosion and non-explosion for nonlinear Hawkes process. Then, we will study the asymptotic behavior of the explosion time. Figure 5.6 illustrates the explosion of a finite time.

Notice that if $\|h\|_{L^1} = \infty$ and $\lim_{z \to \infty} \frac{\lambda(z)}{z} = \alpha > 0$, then one is in the super-critical regime and we will see that λ_t grows exponentially; this is discussed Section 5.4. If $\|h\|_{L^1} = \infty$ and $\sum_{n=0}^\infty \frac{1}{\lambda(n)} < \infty$, then one is in the explosive regime to be discussed in Section 5.5.

We will launch a systematic study of the time asymptoics for Hawkes process in different regimes. We will study the sublinear regime, sub-critical regime, critical regime and super-critical regime in Sections 5.1, 5.2, 5.3, 5.4 respectively. Finally, in Section 5.5, we will provide a criterion for explosion and non-explosion for Hawkes process and obtain some asymptotics for the explosion time.

Figure 5.1: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = (t+1)^{-\frac{1}{2}}$ and $\lambda(z) = (1+z)^{\frac{1}{2}}$.
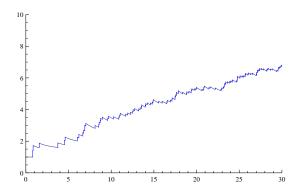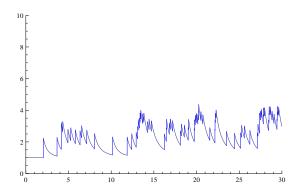

Figure 5.2: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = \frac{4}{(t+1)^3}$ and $\lambda(z) = (1+z)^{\frac{1}{2}}$. In this case, $\|h\|_{L^1} < \infty$ and $\lambda(\cdot)$ is sublinear and Lipschitz. It will converge to the unique stationary version of the Hawkes process.
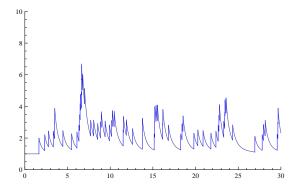
Figure 5.3: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = \frac{1}{(t+1)^3}$ and $\lambda(z) = 1 + z$. In this case, $\|h\|_{L^1} = \frac{1}{2} < 1$. It is in the sub-critical regime. This is a classical Hawkes process and it will converge to the unique stationary version of the Hawkes process.

5.1 Sublinear Regime

In this section, we are interested in the sublinear case $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 0$. If $\|h\|_{L^1} < \infty$ and $\lambda(\cdot)$ is α-Lipschitz and $\alpha\|h\|_{L^1} < 1$, then, as Brémaud and Massoulié [14] proved, there exists a unique stationary Hawkes process. Recently, Karabash [63] relaxed the Lipschitz condition and proved the stability result for a wider class of $\lambda(\cdot)$. Let $\mathbb{P}$ and $\mathbb{E}$ denote the probability measure and expectation for stationary Hawkes process. Then, by ergodic theorem, we have the law of large numbers,

$$(5.1) \qquad \frac{N_t}{t} \to \mu = \mathbb{E}[N[0,1]], \quad \text{as } t \to \infty.$$

The central limit theorem and large deviations have already been discussed in Chapter 2, Chapter 3 and Chapter 4.

If $\|h\|_{L^1} = \infty$, then there is no stationary version of Hawkes process and λ_t tends to ∞ as $t \to \infty$. This is the case we are going to study for the rest of
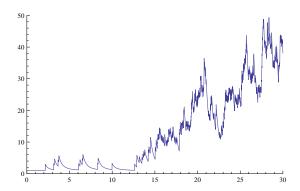
Figure 5.4: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = \frac{2}{(t+1)^3}$ and $\lambda(z) = 1 + z$. In this case, $\|h\|_{L^1} = 1$, $\int_0^\infty th(t)dt < \infty$ and $\lambda(\cdot)$ is linear. It is therefore in the critical regime. From the graph, we can see that λ_t grows linearly in t, which will be proved in this chapter. Indeed, we will prove that $\frac{N_T}{T^2}$ converges to $\int_0^\cdot \eta_s ds$ as $T \to \infty$, where η_s is a squared Bessel process.

the subsection. We are interested the time asymptotic behavior of the nonlinear

Hawkes process in this regime.

Let us first make a simple observation. Assume that $\lambda(z) \uparrow \infty$ as $z \to \infty$.

Then, assuming $\|h\|_{L^1} = \infty$, we have $\lambda_t \to \infty$ as $t \to \infty$ a.s. This can be seen by

noticing that $\int_0^t h(t-s)N(ds) \to \infty$ a.s. if $\|h\|_{L^1} = \infty$, where N_t follows from a

standard Poisson process with constant rate $\lambda(0)$.

Let us prove a special case first.

Proposition 1. *Assume that $h(\cdot) \equiv 1$ and $\lambda(z) = \gamma(\nu + z)^\beta$, where $\gamma, \nu > 0$ and $0 < \beta < 1$. Then,*

$$(5.2) \qquad\qquad \frac{\lambda_t}{t^{\frac{\beta}{1-\beta}}} \to \gamma^{\frac{1}{1-\beta}} (1 - \beta)^{\frac{\beta}{1-\beta}},$$

*in probability as $t \to \infty$.*
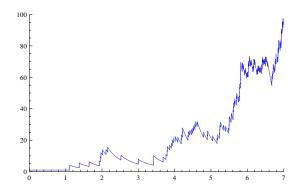
Figure 5.5: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = \frac{3}{(t+1)^3}$ and $\lambda(z) = 1 + z$. In this case, $\|h\|_{L^1} = \frac{3}{2} > 1$ and it is in the super-critical regime. We expect that λ_t would grow exponentially in this case.

Proof. For $\alpha > 0$,

$$
(5.3) \qquad d\lambda_t^{\frac{1}{\alpha}} = \left[\lambda(\nu + N_t + 1)^{\frac{1}{\alpha}} - \lambda(\nu + N_t)^{\frac{1}{\alpha}} \right] dN_t
$$

$$
= \left[\gamma^{\frac{1}{\alpha}}(\nu + N_t + 1)^{\frac{\beta}{\alpha}} - \gamma^{\frac{1}{\alpha}}(\nu + N_t)^{\frac{\beta}{\alpha}} \right] dN_t
$$

$$
= \left[\left(\lambda_t^{\frac{1}{\beta}} + \gamma^{\frac{1}{\beta}} \right)^{\frac{\beta}{\alpha}} - \lambda_t^{\frac{1}{\alpha}} \right] dN_t.
$$

Let $\alpha = \frac{\beta}{1-\beta}$. We have

(5.4)
$$
\lambda_t^{\frac{1-\beta}{\beta}} = \int_0^t \left[(\lambda_s^{\frac{1}{\beta}} + \gamma^{\frac{1}{\beta}})^{1-\beta} - (\lambda_s^{\frac{1}{\beta}})^{1-\beta} \right] \lambda_s ds + \int_0^t \left[(\lambda_s^{\frac{1}{\beta}} + \gamma^{\frac{1}{\beta}})^{1-\beta} - (\lambda_s^{\frac{1}{\beta}})^{1-\beta} \right] dM_s.
$$

Since $\lambda_t \to \infty$ a.s. as $t \to \infty$, by the bounded convergence theorem,

$$
(5.5) \qquad \frac{1}{t} \int_0^t \mathbb{E}\left\{ \left[(\lambda_s^{\frac{1}{\beta}} + \gamma^{\frac{1}{\beta}})^{1-\beta} - (\lambda_s^{\frac{1}{\beta}})^{1-\beta} \right] \lambda_s \right\} ds \to (1-\beta)\gamma^{\frac{1}{\beta}},
$$

as $t \to \infty$. It is not difficult to see that $\frac{1}{t} \int_0^t [(\lambda_s^{\frac{1}{\beta}} + \gamma^{\frac{1}{\beta}})^{1-\beta} - (\lambda_s^{\frac{1}{\beta}})^{1-\beta}] dM_s \to 0$ in probability as $t \to \infty$. Hence, $\frac{\lambda_t^{\frac{1-\beta}{\beta}}}{t} \to (1-\beta)\gamma^{\frac{1}{\beta}}$ in probability as $t \to \infty$. $\qquad\square$
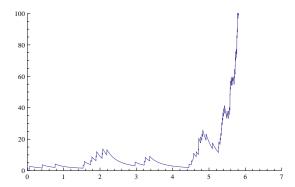
Figure 5.6: Plot of intensity λ_t for a realization of Hawkes process. Here $h(t) = \frac{1}{(t+1)^3}$ and $\lambda(z) = (1+z)^{\frac{3}{2}}$. This is in the explosive regime. The plot is a little bit cheating because it is impossible to "plot" explosion. Nevertheless, you can think it as an illustration. It "appears" that the process explodes near time $t = 6$.

Remark 7. *Assume that* $h(t) = (t+1)^\delta$, $\delta > -1$ *and* $\lambda(z) = \gamma(\nu + z)^\beta$, *where* $\gamma, \nu > 0$ *and* $0 < \beta < 1$. *We conjecture that*

$$(5.6) \qquad \frac{\lambda_t}{t^\alpha} \to \gamma^{\frac{1}{1-\beta}} B(\delta, \alpha)^{\frac{\beta}{1-\beta}},$$

as $t \to \infty$ *a.s., where* $\alpha = \frac{(1+\delta)\beta}{1-\beta}$ *and* $B(\delta, \alpha) = \int_0^1 u^\delta (1-u)^\alpha du$.

5.2   Sub-Critical Regime

In this section, we review some known results about the limit theorems in the sub-critical regime. We say the Hawkes process is in the sub-critical regime if $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 1$ and $\|h\|_{L^1} < 1$. If we further assume that $\lambda(\cdot)$ is α-Lipschitz and $\alpha\|h\|_{L^1} < 1$, then Brémaud and Massoulié [14] proved that there exists a unique stationary Hawkes process. In this regime, we also have the law of large numbers and the central limit theorem just as in Section 5.1. For the case when $\lambda(\cdot)$ is nonlinear, we refer to the review in Section 5.1 for the law of large numbers and

148

central limit theorem.

In particular, when $\lambda(z) = \nu + z$ and $\nu > 0$, we have explict expressions for the law of large numbers, central limit theorem and large deviation principle. They are well known in the literature.

The ergodic theorem implies the following law of large numbers,

$$(5.7) \qquad \frac{N_t}{t} \to \frac{\nu}{1 - \|h\|_{L^1}}, \qquad \text{as } t \to \infty \text{ a.s.}$$

Bordenave and Torrisi [11] proved a large deviation principle for $(\frac{N_t}{t} \in \cdot)$ with the rate function

$$(5.8) \qquad I(x) = \begin{cases} x \log\left(\frac{x}{\nu + x\|h\|_{L^1}}\right) - x + x\|h\|_{L^1} + \nu & \text{if } x \in [0, \infty) \\ +\infty & \text{otherwise} \end{cases}.$$

Bacry et al. [2] proved a functional central limit theorem, stating that

$$(5.9) \qquad \frac{N_{\cdot t} - \cdot \mu t}{\sqrt{t}} \to \sigma B(\cdot), \qquad \text{as } t \to \infty,$$

on $D[0,1]$ with Skorokhod topology, where

$$(5.10) \qquad \mu = \frac{\nu}{1 - \|h\|_{L^1}} \quad \text{and} \quad \sigma^2 = \frac{\nu}{(1 - \|h\|_{L^1})^3}.$$

When $\lambda(\cdot)$ is nonlinear and sub-critical, the central limit theorem has been obtained in Chapter 2.

5.3 Critical Regime

In this section, we are interested in the critical regime, i.e. $\lim_{z \to \infty} \frac{\lambda(z)}{z} = 1$ and $\|h\|_{L^1} = 1$. This regime is very subtle. In some cases, there exists a stationary version of Hawkes process whilst in some cases there does not. For example, Brémaud and Massoulié [15] proved that

Proposition 2 (Brémaud and Massoulié)**.** *Assume* $\lambda(z) = z$, $\|h\|_{L^1} = 1$ *and*

$$(5.11) \qquad \sup_{t \geq 0} t^{1+\alpha} h(t) \leq R, \quad \lim_{t \to \infty} t^{1+\alpha} h(t) = r,$$

for some finite constants $r, R > 0$ *and* $0 < \alpha < \frac{1}{2}$. *Then, there exists a non-trivial stationary Hawkes process with finite intensity.*

Brémaud and Massoulié considered only the linear Hawkes process in their paper [15]. If you allow nonlinear rate function, you get a much richer class of Hawkes processes and in some cases, there still exists a stationary Hawkes process. It is much easier to work with the exponential case, i.e. when $h(t) = ae^{-at}$ and $\|h\|_{L^1} = 1$.

The lecture notes by Hairer [47] provides a sufficient condition for which there exists an invariant probability measure. Let $\mathcal{L}$ be the generator of a Markov process. If there exists $V \geq 1$, continuous, with precompact sublevel sets and some function $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ strictly concave, increasing, with $\phi(0) = 0$, and $\phi(x) \to \infty$ as $x \to \infty$ and $\mathcal{L}V \leq K - \phi(V)$ for some $K > 0$, then there exists an invariant probability measure.

Proposition 3. *Assume* $h(t) = ae^{-at}$, $a > 0$ *and* $\lambda(z) = z - \psi(z) + \nu$, *where* $\psi(z)$ *is positive, increasing, strictly concave and* $\psi(z) \to \infty$ *and* $\frac{\psi(z)}{z} \to 0$ *as* $z \to \infty$. *If*

also $\lambda(z)$ is strictly positive. Then there exists an invarint probability measure.

Proof. Let $V(z) = z + 1$ and $\phi(V) = a(\psi(V) - \psi(0))$. Then $\phi : \mathbb{R}^+ \to \mathbb{R}^+$ is increasing and strictly concave, $\phi(z) \to \infty$, and $\phi(0) = 0$. Recall that the generator is given by

$$(5.12) \qquad \mathcal{A}f(z) = -az\frac{\partial f}{\partial z} + \lambda(z)[f(z + a) - f(z)].$$

Hence, we have

$$(5.13) \quad \mathcal{A}V + \phi(V) = -\psi(z)a + a\psi(z + 1) - a\psi(0) + a\nu \le a\psi(1) - 2a\psi(0) + a\nu.$$

$\square$

We can generalize our result to the much wider class of $h(\cdot)$ when $h(\cdot)$ is a sum of exponentials: $h(t) = \sum_{i=1}^{d} a_i e^{-b_i t}$, where $b_i > 0$ and $a_i > 0$, $1 \le i \le d$. Write $Z_i(t) = \sum_{\tau < t} a_i e^{-b_i(t-\tau)}$. Then $Z_t = \sum_{i=1}^{d} Z_i(t)$ and $(Z_1(t), \dots, Z_d(t))$ is Markovian with the generator

$$(5.14) \quad \mathcal{A}f = -\sum_{i=1}^{d} b_i z_i \frac{\partial f}{\partial z_i} + \lambda\left(\sum_{i=1}^{d} z_i\right) \cdot [f(z_1 + a_1, \dots, z_d + a_d) - f(z_1, \dots, z_d)].$$

We have the following result.

Proposition 4. *Assume $h(t) = \sum_{i=1}^{d} a_i e^{-b_i t}$, $b_i > 0$ and $a_i > 0$, $1 \le i \le d$ and $\|h\|_{L^1} = \sum_{i=1}^{d} \frac{a_i}{b_i} = 1$. Also assume that $\lambda(z) = z - \psi(z) + \nu$, where $\psi(z)$ is positive, increasing, strictly concave and $\psi(z) \to \infty$ and $\frac{\psi(z)}{z} \to 0$ as $z \to \infty$ and $\lambda(z)$ is strictly positive. Then, there exists an invariant probability measure.*

Proof. Let $V = \sum_{i=1}^{d} \frac{z_i}{b_i} + 1$ and $\phi(V) = \psi(\min_{1 \le i \le d} b_i V) - \psi(0)$. Then $\phi : \mathbb{R}^+ \to$

151

$\mathbb{R}^+$ is increasing and strictly concave, $\phi(z) \to \infty$ as $z \to \infty$ and $\phi(0) = 0$. Using the concavity and monotonicity of $\psi(\cdot)$, we have

$$(5.15) \quad \mathcal{A}V + \phi(V)$$

$$= -\psi\left(\sum_{i=1}^{d} z_i\right) + \psi\left(\min_{1\leq i\leq d} b_i \sum_{i=1}^{d} b_i \frac{z_i}{b_i} + \min_{1\leq i\leq d} b_i\right) - \psi(0) + \nu$$

$$\leq -\psi\left(\min_{1\leq i\leq d} b_i \sum_{i=1}^{d} \frac{z_i}{b_i}\right) + \psi\left(\min_{1\leq i\leq d} b_i \sum_{i=1}^{d} b_i \frac{z_i}{b_i} + \min_{1\leq i\leq d} b_i\right) - \psi(0) + \nu$$

$$\leq \psi\left(\min_{1\leq i\leq d} b_i\right) - 2\psi(0) + \nu.$$

$\square$

Remark 8. *The following $\psi(z)$ satisfies the assumptions in Proposition 4 for sufficiently large $\nu > 0$.*

(i) $\psi(z) = (c_1 + c_2 z)^\alpha$, where $c_1, c_2 > 0$ and $0 < \alpha < 1$.

(ii) $\psi(z) = \log(c_3 + z)$, where $c_3 > 1$.

Remark 9. *Let μ be the invariant probability measure for $(Z_1(t), \ldots, Z_d(t))$ in Prosposition 4. Then, we have $\int \psi\left(\min_{1\leq i\leq d} b_i \sum_{i=1}^{d} \frac{z_i}{b_i} + 1\right) \mu(dz) < \infty$.*

Indeed, when $h(\cdot)$ may not be exponential or a sum of exponentials, we have the following result.

Theorem 20. *Assume $\lambda(z) = \nu + z - \psi(z)$, where $\psi(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+$ satisfies $\lim_{z\to\infty} \psi(z) = \infty$ and $\lim_{z\to\infty} \frac{\psi(z)}{z} = 0$ and also $\lambda(z)$ is increasing. Also assume that $\|h\|_{L^1} = 1$. Then there exists a stationary Hawkes process satisfying the dynamics (1.2).*

Proof. The proof uses Poisson embedding and follows the ideas in Brémaud and Massoulié [14]. Consider the canonical space of a point process on $\mathbb{R}^2$ in which

152

$\overline{N}$ is Poisson with intensity 1. Let $\lambda_t^0 = Z_t^0 = 0$, $t \in \mathbb{R}$ and let N^0 be the point process counting the points of $\overline{N}$ below the curve $t \mapsto \lambda_t^0$, i.e. $N^0 = \varnothing$. Define recursively the processes λ_t^n, Z_t^n and N^n, $n \geq 0$ as follows.

(5.16)
$$\lambda_t^{n+1} = \lambda \left(\int_{-\infty}^t h(t-s)N^n(ds) \right), \quad Z_t^{n+1} = \int_{-\infty}^t h(t-s)N^n(ds), \quad t \in \mathbb{R},$$
$$N^{n+1}(C) = \int_C \overline{N}(dt \times [0, \lambda_t^{n+1}]), \quad C \in \mathcal{B}(\mathbb{R}).$$

By our construction, λ_t^n is an $\mathcal{F}_t^{\overline{N}}$-intensity of N^n (see Brémaud and Massoulié [14]). Since $\lambda(\cdot)$ is increasing, the processes λ_t^n, Z_t^n and N^n are increasing in n Thus, the limit processes λ_t, Z_t, N exist. Since λ_t^n, Z_t^n are stationary in t and increasing in n, we have

(5.17) $$\mathbb{E}\lambda_0^{n+1} = \nu + \mathbb{E}[\lambda_0^n] \int_0^\infty h(t)dt - \mathbb{E}\psi(Z_0^{n+1}) \leq \nu + \mathbb{E}\lambda_0^{n+1} - \mathbb{E}\psi(Z_0^{n+1}).$$

Therefore, by Fatou's lemma, $\mathbb{E}[\psi(Z_0)] \leq \nu < \infty$. Thus, $\psi(Z_t)$ is finite a.s. Since $\lim_{z \to \infty} \psi(z) = \infty$, Z_t is finite a.s. and thus λ_t is finite a.s. N, which counts the number of points of $\overline{N}$ below the curve $t \mapsto \lambda_t$, admits λ_t as an $\mathcal{F}_t^{\overline{N}}$-intensity. The monotonicity implies

(5.18) $$\lambda_t^n \leq \lambda \left(\int_{-\infty}^t h(t-s)N(ds) \right), \quad \lambda_t \geq \lambda \left(\int_{-\infty}^t h(t-s)N^n(ds) \right).$$

Letting $n \to \infty$, we complete the proof. $\qquad\qquad\qquad\qquad\qquad\qquad \square$

Remark 10. *The following $\psi(z)$ satisfies the assumptions in Theorem 20.*

(i) $\psi(z) = (c_1 + c_2 z)^\alpha$, where $c_1, c_2 > 0$, $0 < \alpha < 1$, $\nu > c_1^\alpha$ and $\alpha c_1^{\alpha-1} c_2 < 1$.

(ii) $\psi(z) = \log(c_3 + z)$, where $1 < c_3 < e^\nu$.

Next, let us consider the critical linear case, i.e. $\lambda(z) = \nu + z$, $\nu > 0$ and $\|h\|_{L^1} = 1$. We also assume that $m := \int_0^\infty th(t)dt < \infty$. There is no stationary Hawkes process in this regime and in the rest of this subsection, we will try to understand its time asymptotics.

First, let us prove a lemma concerning the expectations of λ_t and N_t.

Lemma 30. *Assume* $\lambda(z) = \nu + z$, $\nu > 0$ *and* $\|h\|_{L^1} = 1$ *and* $m = \int_0^\infty th(t)dt < \infty$. *We have*

$$(5.19) \qquad \lim_{t\to\infty} \frac{\mathbb{E}[\lambda_t]}{t} = \frac{\nu}{m}, \quad \lim_{t\to\infty} \frac{\mathbb{E}[N_t]}{t^2} = \frac{\nu}{2m}.$$

Proof. Since

$$(5.20) \qquad \lambda_t = \nu + \int_0^t h(t-s)dN_s,$$

taking $f(t) = \mathbb{E}[\lambda_t]$, we get

$$(5.21) \qquad f(t) = \nu + \int_0^t h(t-s)f(s)ds = \nu + \int_0^t h(s)f(t-s)ds.$$

Taking the Laplace transform on both sides of the equation, it is easy to see that the Laplace transform $\hat{f}$ of f is given by

$$(5.22) \qquad \hat{f}(\sigma) = \frac{\nu}{\sigma(1-\hat{h}(\sigma))} \sim \frac{\nu}{m}\frac{1}{\sigma^2}, \quad \text{as } \sigma \downarrow 0,$$

since $\hat{h}(0) = 1$ by $\|h\|_{L^1} = 1$ and $\frac{1-\hat{h}(\sigma)}{\sigma} \sim -\hat{h}'(0) = m$. By a Tauberian theorem, (see Chapter XIII of Feller [38]), we get $\frac{f(t)}{t} \to \frac{\nu}{m}$ as $t \to \infty$. Using the simple fact that $\mathbb{E}[N_t] = \int_0^t f(s)ds$, we complete the proof. $\qquad\square$

Theorem 21. *Assume $\lambda(z) = \nu + z$, $\nu > 0$ and $\|h\|_{L^1} = 1$, $m = \int_0^\infty t h(t) dt < \infty$ and $h(\cdot)$ Lipschitz. We have the following asymptotics.*

(i) As $T \to \infty$, on $D[0,1]$,

$$(5.23) \qquad \frac{N_{tT}}{T^2} \to \int_0^t \eta_s ds,$$

where η_t is a squared Bessel process, i.e.

$$(5.24) \qquad d\eta_t = \frac{\nu}{m} dt + \frac{1}{m}\sqrt{\eta_t} dB_t, \quad \eta_0 = 0.$$

(ii) $\lim_{T \to \infty} N\left[T, T + \frac{t}{T}\right] = P(t)$, where $P(t)$ is a Pólya process with parameters $\frac{1}{2m^2}$ and $2\nu m$.

Remark 11. *The fact that a squared Bessel process arises in the limit of a critical linear Hawkes process is not a surprise. It is well known that a critical branching process after certain scalings will converge to a squared Bessel process in the limit. This was discovered by Wei and Winnicki [105].*

Remark 12. *Before we proceed to the proof of Theorem 21, let us recall that a Pólya process with parameters α and β is a point process defined as the following. Generate a positive random variable ξ, with Gamma distribution of parameters α (shape) and β (scale). Conditional on ξ, $P(t)$ is a Poisson process with intensity ξ. The marginal distribution of $P(t)$ is negative binomial and unlike the usual Poisson process, Pólya process has dependent increments. The covariance of the increments can be computed explicitly as $Cov(P(t + \delta t) - P(t), P(t)) = t \cdot \delta t \cdot \alpha\beta^2$. Peng and Kou [92] used Pólya process to model clustering effects in the credit markets.*

Proof of Theorem 21. (i) Let $H(t) := \int_t^\infty h(s)ds$. Then, we have $H(0) = 1$ and

$\int_0^\infty H(t)dt = \int_0^\infty th(t)dt = m$. Let $M_t := N_t - \int_0^t \lambda_s ds$. Let us integrate $\lambda_s = \int_0^s h(s-u)N(du) + \nu$ over $0 \leq s \leq tT$. We get

$$(5.25) \qquad \int_0^{tT} \lambda_s ds = \int_0^{tT} \int_0^s h(s-u)dM_u ds + \int_0^{tT} \int_0^s h(s-u)\lambda_u du ds + \nu tT.$$

Rearranging the equation and dividing by T, we get

$$(5.26) \qquad \frac{1}{T}\left[\int_0^{tT} \lambda_s ds - \int_0^{tT}\int_0^s h(s-u)\lambda_u du ds\right] = \frac{1}{T}\int_0^{tT}\int_0^s h(s-u)dM_u ds + \nu t.$$

Fubini's theorem implies that

$$(5.27)$$
$$\frac{1}{T}\left[\int_0^{tT} \lambda_u du - \int_0^{tT}\left(\int_0^{tT-u} h(s)ds\right)\lambda_u du\right] = \frac{1}{T}\int_0^{tT}\left(\int_0^{tT-u} h(s)ds\right)dM_u + \nu t.$$

By the definition of $H(\cdot)$, this is equivalent to

$$(5.28) \qquad \int_0^t TH(tT-uT)\frac{\lambda_{uT}}{T}du = \frac{M_{tT}}{T} + \nu t + \frac{1}{T}\int_0^t TH(tT-uT)d\left(\frac{M_{uT}}{T}\right).$$

$\frac{M_{tT}}{T}$ is a martingale and the tightness can be easily established. Furthermore, we have

$$(5.29) \qquad \sup_{T>0} \mathbb{E}\left[\left(\frac{M_{tT}}{T}\right)^2\right] = \sup_{T>0}\frac{1}{T^2}\mathbb{E}\left[\int_0^{tT}\lambda_s ds\right] < \infty,$$

since $\mathbb{E}[\lambda_t] \leq Ct$ for some $C > 0$ by Lemma 30. This implies that the limit of $\frac{M_{tT}}{T}$ is also a martinagle.

Moreover, $\frac{N_{tT}}{T^2}$ and $\int_0^t \frac{\lambda_{sT}}{T}ds$ are both tight. To see this, since N_t and λ_t are nonnegative, we can think of $\left(d\left(\frac{N_{tT}}{T^2}\right), 0 \leq t \leq 1\right)$ and $\left(\frac{\lambda_{tT}}{T}dt, 0 \leq t \leq 1\right)$ as two

measures. But by Lemma 30, we know that there exist some positive contant $C > 0$, such that

$$(5.30) \qquad \mathbb{E}\left[\frac{N_T}{T^2}\right] \le C \quad \text{and} \quad \mathbb{E}\left[\int_0^1 \frac{\lambda_{sT}}{T}ds\right] \le C,$$

uniformly in $T > 0$. Therefore, $\left(d\left(\frac{N_{tT}}{T^2}\right), 0 \le t \le 1\right)$ and $\left(\frac{\lambda_{tT}}{T}dt, 0 \le t \le 1\right)$ are tight in the weak topology. Hence, their distribution functions $\frac{N_{tT}}{T^2}$ and $\int_0^t \frac{\lambda_{sT}}{T}ds$ are tight in $D[0,1]$ equipped with the Skorohod topology. Let us say that $\frac{M_{tT}}{T} \to \beta_t$, $\frac{N_{tT}}{T^2} \to \psi_t$ and $\int_0^t \frac{\lambda_{sT}}{T}ds \to \phi_t$ as $T \to \infty$. Since the jumps of $\frac{N_{tT}}{T^2}$ are uniformly bounded by $\frac{1}{T^2}$ which goes to zero as $T \to \infty$, we conclude that ψ_t is continuous. Similarly, β_t and ϕ_t are continuous. Moreover, the difference

$$(5.31) \qquad \frac{N_{tT}}{T^2} - \int_0^t \frac{\lambda_{sT}}{T}ds = \frac{M_{tT}}{T^2},$$

is a martingale and by Doob's martingale inequality, for any $\epsilon > 0$,

$$(5.32) \qquad \mathbb{P}\left(\sup_{0 \le t \le 1}\left|\frac{M_{tT}}{T^2}\right| \ge \epsilon\right) \le \frac{4}{T^4}\mathbb{E}\left[\int_0^{tT}\lambda_s ds\right] \to 0,$$

as $T \to \infty$. Therefore, $\psi_t = \phi_t$. Let us denote $TH(\cdot T)$ by H_T, $\frac{M_{\cdot T}}{T}$ by M_T and $\int_0^{\cdot} \frac{\lambda_{sT}}{T}ds$ by Λ_T. For any smooth function $K(\cdot)$ supported on $\mathbb{R}^+$, taking the convolutions of the both sides of (5.28), we get

$$(5.33) \qquad K * H_T * \Lambda_T = K * M_T + K * (\nu \cdot) + \frac{1}{T}K * H_T * M_T.$$

Letting $T \to \infty$, using the fact that $\int_0^\infty H(t)dt = \int_0^\infty th(t)dt = m$, we get

$$(5.34) \qquad m \int_0^t K(t-s)d\phi_s = \int_0^t K(t-s)(\beta_s + \nu s)ds.$$

Since this is true for any K, we get $\frac{d\phi_t}{dt} = \frac{\beta_t}{m} + \frac{\nu}{t}$. Finally,

$$(5.35) \qquad \left(\frac{M_{tT}}{T}\right)^2 - \int_0^t \frac{\lambda_{sT}}{T}ds$$

is a martingale and if we let $T \to \infty$, we conclude that $\beta_t^2 - \phi_t$ is a martingale. Let $\eta_t := \frac{d\phi_t}{dt}$. We have proved that $\frac{N_{tT}}{T^2} \to \int_0^t \eta_s ds$ weakly on $D[0,1]$ equipped with Skorohod topology and η_t is a squared Bessel process,

$$(5.36) \qquad d\eta_t = \frac{\nu}{m}dt + \frac{1}{m}\sqrt{\eta_t}dB_t, \quad \eta_0 = 0.$$

(ii) $N[T, T + \frac{t}{T}]$ has the compensator $\int_T^{T+\frac{t}{T}} \lambda_s ds$. Observe that $\int_T^{T+\frac{t}{T}} \lambda_s ds = T^2 \int_1^{1+\frac{t}{T^2}} \frac{\lambda_{sT}}{T}ds \to \eta_1 t$ as $T \to \infty$, where η_1 has a Gamma distribution with shape $\frac{1}{2m^2}$ and scale $2\nu m$. $\qquad \square$

Now let us consider the case when $h(\cdot)$ has heavy tail, i.e. $\int_0^\infty th(t)dt = \infty$. Let us first prove the following lemma.

Lemma 31. *Assume that*

$$(5.37) \qquad 1 - \int_0^t h(s)ds = \int_t^\infty h(s)ds \sim t^{-\alpha}, \quad 0 < \alpha < 1.$$

Then,

$$(5.38) \qquad \lim_{t \to \infty} \frac{\mathbb{E}[\lambda_t]}{t^\alpha} = \nu \cdot \frac{\sin \pi\alpha}{\pi\alpha}, \quad \lim_{t \to \infty} \frac{\mathbb{E}[N_t]}{t^{1+\alpha}} = \frac{\nu}{\Gamma(1-\alpha)\Gamma(2+\alpha)}.$$

158

Proof. The Tauberian theorem of Chapter XIII of Feller [38] says that

$$(5.39) \qquad 1 - \hat{h}(\sigma) \sim \Gamma(1 - \alpha)\sigma^\alpha, \quad \sigma \to 0^+.$$

Let $\mathbb{E}[\lambda_t] = f(t)$. This implies that

$$(5.40) \qquad \hat{f}(\sigma) = \frac{\nu}{\sigma(1 - \hat{h}(\sigma))} \sim \frac{\nu\sigma^{-1-\alpha}}{\Gamma(1 - \alpha)}, \quad \sigma \to 0^+,$$

which again by a Tauberian theorem (Theorem 2 of Chapter XIII.5 of Feller [38]) implies

$$(5.41) \qquad \int_0^t f(s)ds \sim \frac{\nu}{\Gamma(1 - \alpha)\Gamma(2 + \alpha)} t^{1+\alpha}, \quad t \to \infty.$$

Hence,

$$(5.42) \qquad \mathbb{E}[N_t] = \int_0^t \mathbb{E}[\lambda_s]ds = \int_0^t f(s)ds \sim \frac{\nu}{\Gamma(1 - \alpha)\Gamma(2 + \alpha)} t^{1+\alpha}, \quad t \to \infty.$$

Since $\mathbb{E}[\lambda_t] = \nu + \int_0^t h(t - s)d\mathbb{E}[N_s]$, it is easy to check that

$$(5.43) \qquad \mathbb{E}[\lambda_t] = f(t) \sim \frac{\nu}{\Gamma(1 - \alpha)\Gamma(1 + \alpha)} t^\alpha = \nu \cdot \frac{\sin \pi\alpha}{\pi\alpha} \cdot t^\alpha, \quad t \to \infty.$$

□

We obtain the following law of large numbers.

Theorem 22. *Assume that $\int_t^\infty h(s)ds \sim \frac{1}{t^\alpha}$, $0 < \alpha < 1$. Then,*

$$(5.44) \qquad \frac{N_t}{t^{1+\alpha}} \to \frac{\nu}{\Gamma(1 - \alpha)\Gamma(2 + \alpha)} \quad and \quad \frac{\lambda_t}{t^\alpha} \to \nu \cdot \frac{\sin \pi\alpha}{\pi\alpha}, \quad a.s. \ as \ t \to \infty.$$

Proof. Let $X_t = N_t - \mathbb{E}[N_t]$. Then, X_t satisfies (see Bacry et al. [2])

$$(5.45) \qquad X_t = M_t + \int_0^t \psi(t-s)M_s ds,$$

where $M_t = N_t - \int_0^t \lambda_s ds$ and $\psi = \sum_n h^{*n}$. Then, by Doob's maximal inequality, it is not hard to see that

$$(5.46) \qquad \mathbb{E}\left[\left(\frac{N_t - \mathbb{E}[N_t]}{t^{1+\alpha}}\right)^2\right] \leq \frac{1}{t^{2+2\alpha}} \mathbb{E}\left[\sup_{s \leq t} M_s^2\right] \left(1 + \int_0^t \psi(t-s)ds\right)^2$$

$$(5.47) \qquad \leq \frac{C}{t^{2+2\alpha}} t^{1+\alpha}(t^\alpha)^2 \to 0,$$

as $t \to \infty$ since $0 < \alpha < 1$. Hence, as $t \to \infty$,

$$(5.48) \qquad \frac{N_t}{t^{1+\alpha}} \to \frac{\nu}{\Gamma(1-\alpha)\Gamma(2+\alpha)}, \quad \text{in } L^2 \text{ as } t \to \infty.$$

To show the almost sure convergence, we need only to show that $\frac{1}{t}\sup_{s \leq t} M_s \to 0$ a.s. as $t \to \infty$. Define $Y_t = \int_0^t \frac{1}{1+s} dM_s$. Then by Lemma 31,

$$(5.49) \qquad \sup_{t>0} \mathbb{E}[Y_t^2] = \int_0^\infty \frac{\mathbb{E}[\lambda_s]}{(1+s)^2} ds < \infty.$$

By the martingale convergence theorem, $Y_t \to Y_\infty$ a.s. as $t \to \infty$. It follows that

$$(5.50) \qquad \frac{M_t}{t+1} = Y_t - \frac{1}{t+1}\int_0^t Y_s ds \to 0,$$

a.s. as $t \to \infty$. From here, it is easy to show that $\frac{1}{t}\sup_{s \leq t} M_s \to 0$ a.s. Finally,

160

since $\lambda_t = \nu + \int_0^t h(t-s)N(ds)$, we conclude that

$$(5.51) \qquad \frac{\lambda_t}{t^\alpha} \to \nu \cdot \frac{\sin \pi\alpha}{\pi\alpha}, \qquad \text{a.s. as } t \to \infty.$$

$\square$

5.4 Super-Critical Regime

In this section, we are interested in the super-critical regime, i.e. $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 1$ and $\|h\|_{L^1} > 1$. First, let us compute the asymptotics for the expectations. Let $\theta > 0$ be the unique positive number such that $\int_0^\infty e^{-\theta t} h(t) dt = 1$. θ is sometimes referred to as the Malthusian parameter in the literature. Let us also define

$$(5.52) \qquad \overline{h}(t) = h(t)e^{-\theta t}, \quad \overline{m} = \int_0^\infty t h(t) e^{-\theta t} dt.$$

Clearly under our assumptions $0 < \overline{m} < \infty$ and $\|\overline{h}\|_{L^1} = 1$.

Lemma 32. *(i) Assume $\lambda(z) = \nu + z$, $\nu > 0$ being a constant. Then,*

$$(5.53) \qquad \lim_{t\to\infty} \frac{\mathbb{E}[\lambda_t]}{e^{\theta t}} = \frac{\nu}{\theta\overline{m}}.$$

(ii) Assume $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 1$ and let $\lambda(\cdot)$ be bounded below by a positive constant. Then,

$$(5.54) \qquad \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[\lambda_t] = \lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[N_t] = \theta.$$

161

Proof. (i) Let $f(t) = \mathbb{E}[\lambda_t]$. We have

$$(5.55) \qquad \frac{f(t)}{e^{\theta t}} = \frac{\nu}{e^{\theta t}} + \int_0^t h(t-s) e^{-\theta(t-s)} \frac{f(s)}{e^{\theta s}} ds = \frac{\nu}{e^{\theta t}} + \int_0^t \bar{h}(t-s) \frac{f(s)}{e^{\theta s}} ds$$

Taking Laplace transform, we get

$$(5.56) \qquad \widehat{f(t)e^{-\theta t}}(\sigma) = \frac{\nu}{\theta(1 - \hat{\bar{h}}(\sigma))} \sim \frac{\nu}{\theta \overline{m}} \cdot \frac{1}{\sigma},$$

as $\sigma \downarrow 0$. By the Tauberian theorem, we have

$$(5.57) \qquad \lim_{t \to \infty} \frac{\mathbb{E}[\lambda_t]}{e^{\theta t}} = \frac{\nu}{\theta \overline{m}}.$$

(ii) is a direct consequence of (i). $\qquad \qquad \square$

This is consistent with the exponential case when $h(t) = ae^{-bt}$ and $a > b$. We have

$$(5.58) \quad \mathbb{E}[\lambda_t] = -\frac{\nu b}{a-b} + \frac{\nu a}{a-b} e^{(a-b)t}, \quad \mathbb{E}[N_t] = -\frac{\nu bt}{a-b} + \frac{\nu a}{(a-b)^2}(e^{(a-b)t} - 1).$$

Indeed, in the exponential case, $\theta = a - b$ and

$$(5.59) \quad d(Z_t e^{-(a-b)t}) = e^{-(a-b)t} dZ_t + Z_t de^{-(a-b)t} = -aZ_t e^{-(a-b)t} dt + ae^{-(a-b)t} dN_t.$$

Let $Y_t = Z_t e^{-(a-b)t}$. We have

$$(5.60) \qquad dY_t = -aY_t dt + ae^{-(a-b)t} dN_t = \nu a e^{-(a-b)t} dt + ae^{-(a-b)t} dM_t.$$

If we assume that $N(-\infty, 0] = 0$, then $Z_0 = 0$ and

$$(5.61) \qquad Y_t = \int_0^t \nu a e^{-(a-b)s} ds + a \int_0^t e^{-(a-b)s} dM_s.$$

Clearly, $\int_0^t \nu e^{-(a-b)s} ds \to \frac{\nu a}{a-b}$ and $\int_0^t a e^{-(a-b)s} dM_s$ is a martingale and

$$(5.62) \quad \sup_{t>0} \mathbb{E}\left[\left(\int_0^t e^{-(a-b)s} dM_s\right)^2\right] = \int_0^\infty e^{-2(a-b)s} \mathbb{E}[\lambda_s] ds = \frac{\nu(2a-b)}{2(a-b)^2} < \infty.$$

Therefore, by the martingale convergence theorem, there exists some W in $L^2(\mathbb{P})$ such that

$$(5.63) \qquad \frac{\lambda_t}{e^{(a-b)t}} \to \frac{\nu a}{a-b} + aW,$$

as $t \to \infty$. The convergence is a.s. and also in $L^2(\mathbb{P})$.

For the general $h(\cdot)$ such that $\|h\|_{L^1} > 1$, we may even consider the case when $\|h\|_{L^1} = \infty$. For instance, if we assume that $h(\cdot)$ is decreasing and continuous and then $h(\cdot)$ is bounded and all the arguments for the case $1 < \|h\|_{L^1} < \infty$ would work for the case $\|h\|_{L^1} = \infty$ as well.

Theorem 23. *Assume $\lambda(z) = \nu + z$, $\nu > 0$. We have,*

$$(5.64) \qquad \frac{\lambda_t}{e^{\theta t}} \to \frac{\nu}{\theta \overline{m}} + \frac{W}{\overline{m}}, \quad a.s. \ as \ t \to \infty,$$

where $W = \int_0^\infty e^{-\theta t} dM_t$.

Proof. It is not very hard to observe that

$$(5.65) \qquad \frac{\lambda_t}{e^{\theta t}} = \frac{\nu}{e^{\theta t}} + \int_0^t \frac{h(t-s)}{e^{\theta t}} dM_s + \int_0^t \frac{h(t-s)}{e^{\theta t}} \lambda_s ds$$

$$= \frac{\nu}{e^{\theta t}} + \int_0^t h(t-s)e^{-\theta(t-s)} \left(\frac{dM_s}{e^{\theta s}}\right) + \int_0^t h(t-s)e^{-\theta(t-s)} \frac{\lambda_s}{e^{\theta s}} ds$$

$$= \frac{\nu}{e^{\theta t}} + \int_0^t \overline{h}(t-s)d\overline{M}_s + \int_0^t \overline{h}(t-s)\frac{\lambda_s}{e^{\theta s}} ds.$$

Taking Laplace transform, we get

$$(5.66) \qquad \widehat{\lambda_t e^{-\theta t}}(\sigma) = \frac{\frac{\nu}{\theta+\sigma} + \widehat{\overline{h}}(\sigma) \int_0^\infty e^{-\sigma t} d\overline{M}_t}{1 - \widehat{\overline{h}}(\sigma)} = \frac{\frac{\nu}{\theta+\sigma} + \widehat{\overline{h}}(\sigma) \int_0^\infty e^{-(\sigma+\theta)t} dM_t}{1 - \widehat{\overline{h}}(\sigma)}$$

$$\sim \frac{\frac{\nu}{\theta} + W}{\overline{m}} \cdot \frac{1}{\sigma},$$

as $\sigma \downarrow 0$, where $W = \int_0^\infty e^{-\theta t} dM_t$. Notice that W is well defined a.s. because $\int_0^t e^{-\theta s} dM_s$ is a martingale and

$$(5.67) \qquad \sup_{t>0} \mathbb{E}\left[\left(\int_0^t e^{-\theta s} dM_s\right)^2\right] = \int_0^\infty e^{-2\theta s} \mathbb{E}[\lambda_s] ds < \infty$$

by Lemma 32. Hence, by the Tauberian theorem, we conclude that, as $t \to \infty$,

$$(5.68) \qquad \frac{\lambda_t}{e^{\theta t}} \to \frac{\nu}{\theta \overline{m}} + \frac{W}{\overline{m}}. \quad \text{a.s.}$$

□

Corollary 1. $\frac{N_t}{e^{\theta t}} \to \frac{\nu}{\theta^2 \overline{m}} + \frac{W}{\theta \overline{m}}$ *a.s. as* $t \to \infty$.

Proof. Let $M_t = N_t - \int_0^t \lambda_s ds$. Then, since M_t is a martingale and $\mathbb{E}[M_t^2] = \int_0^t \mathbb{E}\lambda_s ds \leq Ce^{\theta t}$ for some $C > 0$, it is easy to see that $\frac{M_t}{e^{\theta t}} \to 0$ a.s. as $t \to \infty$. On

164

the other hand,

$$(5.69) \qquad \frac{1}{e^{\theta t}} \int_0^t \lambda_s ds = \int_0^t e^{-\theta(t-s)} \frac{\lambda_s}{e^{\theta s}} ds \to \frac{1}{\theta} \left[\frac{\nu}{\theta \overline{m}} + \frac{W}{\overline{m}} \right],$$

by Theorem 23. Hence, we get the desired result. $\qquad\square$

Remark 13. *It would be interesting to study the properties of W defined in Theorem 23. Observe that*

$$(5.70) \qquad \mathbb{E}\left[e^{-\sigma \int_0^t e^{-\theta s} dM_s} \right] = \mathbb{E}\left[e^{-\sigma \left(\int_0^t e^{-\theta s} dN_s - \int_0^t \int_0^s h(s-u) N(du) e^{-\theta s} ds \right)} \right] e^{\frac{\sigma}{\theta} \nu (1 - e^{-\theta t})}$$

$$= \mathbb{E}\left[e^{-\sigma \int_0^t \left(e^{-\theta s} - \int_s^t h(u-s) e^{-\theta u} du \right) N(ds)} \right] e^{\frac{\sigma}{\theta} \nu (1 - e^{-\theta t})}$$

$$= \mathbb{E}\left[e^{-\sigma \int_0^t e^{-\theta s} \overline{H}(t-s) N(ds)} \right] e^{\frac{\sigma}{\theta} \nu (1 - e^{-\theta t})},$$

where $\overline{H}(t) = \int_t^\infty \overline{h}(s) ds$. Hence,

$$(5.71) \qquad \mathbb{E}[e^{-\sigma W}] = e^{\frac{\sigma \nu}{\theta}} \lim_{t \to \infty} e^{\nu \int_0^t g_t(s) ds},$$

where $g_t(s) = \exp\left\{ -\frac{\sigma}{e^{\theta t}} e^{\theta s} \overline{H}(s) + \int_0^s h(s-u) g_t(u) du \right\} - 1$.

5.5 Explosive Regime

In this section, we will provide an explosion, non-explosion criterion for nonlinear Hawkes processes, together with some asymptotics for the explosion time in the explosive regime. Let $\tau_\ell = \inf\{t > 0 : \lambda_t \geq \ell\}$. The quantity

$$(5.72) \qquad \lim_{\ell \to \infty} \mathbb{P}(\tau_\ell \leq t) = F(t) = \mathbb{P}(\tau \leq t),$$

is defined as the distribution function of the explosion time τ. We say there is no explosion if $F \equiv 0$, otherwise there is explosion. For a short introduction to explosion, non-explosion, we refer to Varadhan [109].

Next, we provide an explosion, non-explosion criterion for nonlinear Hawkes processes. The proof is based on a well known result for the explosion, non-explosion criterion for a class of point processes which can be found in the book by Kallenberg [61].

Theorem 24 (Explosion, Non-Explosion Criterion). *Assume that $\lambda(\cdot)$ is increasing and that $h(\cdot)$ is integrable and decreasing, then there is explosion if and only if*

$$(5.73) \qquad \sum_{n=0}^{\infty} \frac{1}{\lambda(n)} < \infty.$$

Proof. Observe that, for any $T > 0$,

$$(5.74) \qquad \mathbb{P}^{h(T)}(\tau \leq T) \leq \mathbb{P}(\tau \leq T) \leq \mathbb{P}^{h(0)}(\tau \leq T),$$

where $\mathbb{P}^{h(0)}$ denotes the probability measure for the point process such that initially the rate function is $\lambda(0)$ and after nth jumps, the rate function becomes $\lambda(nh(0))$; $\mathbb{P}^{h(T)}$ is defined similarly. It is well known that the point process with intensity $\lambda(N_{t-})$ is explosive if and only if

$$(5.75) \qquad \sum_{n=0}^{\infty} \frac{1}{\lambda(n)} < \infty.$$

For the details and proof of the above result, we refer to Kallenberg [61]. But it is clear under our assumptions that $\sum_{n=0}^{\infty} \frac{1}{\lambda(n)} < \infty$ if and only if $\sum_{n=0}^{\infty} \frac{1}{\lambda(cn)} < \infty$,

166

where $c > 0$ is any positive constant. Therefore, there is explosion if and only if

$$(5.76) \qquad \sum_{n=0}^{\infty} \frac{1}{\lambda(n)} < \infty.$$

$\square$

Evaluating the exact probability distribution of the explosion time τ, i.e. $\mathbb{P}(\tau \leq t)$, is hard and almost impossible. Nevertheless, one can still study its asymptotic behavior, i.e.

(i) $\mathbb{P}(\tau \geq t)$ for large time t;

(ii) $\mathbb{P}(\tau \leq \epsilon)$ for small time ϵ.

In the rest of this section, we will use Proposition 5 to answer (i) and Proposition 6 to answer (ii).

Proposition 5. *Under the assumptions of Theorem 24 satisfying the explosion criterion, we have*

$$(5.77) \qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}^{\varnothing}(\tau \geq t) = \inf_{t > 0} \frac{1}{t} \log \mathbb{P}^{\varnothing}(\tau \geq t) = -\sigma,$$

where $0 < \sigma < \infty$.

Proof. For a nonlinear Hawkes process with empty history, i.e. $N(-\infty, 0] = 0$, we have

$$(5.78) \quad \mathbb{P}^{\varnothing}(\tau \geq t + s) = \mathbb{P}^{\varnothing}(\tau \geq t + s | \tau \geq s)\mathbb{P}^{\varnothing}(\tau \geq s) \leq \mathbb{P}^{\varnothing}(\tau \geq t)\mathbb{P}^{\varnothing}(\tau \geq s).$$

Therefore, $\log \mathbb{P}^{\varnothing}(\tau \geq t)$ is sub-additive and we know that

$$(5.79) \qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}^{\varnothing}(\tau \geq t) = \inf_{t > 0} \frac{1}{t} \log \mathbb{P}^{\varnothing}(\tau \geq t) = -\sigma$$

exists. And we also know that $0 < \sigma < \infty$. For example, it is easy to see that $\sigma \leq \lambda(0)$. That is because $\mathbb{P}^{\varnothing}(\tau \geq t) \geq \mathbb{P}^{\varnothing}(N[0, t] = 0) = e^{-\lambda(0)t}$. To see that $\sigma > 0$, choose M large enough so that $\mathbb{P}(\tau \geq M) < 1$ and then $\sigma \geq -\frac{1}{M} \log \mathbb{P}(\tau \geq M) > 0$. $\qquad \square$

Remark 14. *Indeed, in the Markovian case, we can say something more about σ defined in Proposition 5. When $h(t) = ae^{-bt}$, $Z_t = \sum_{\tau < t} ae^{-b(t - \tau)}$ is Markovian and by noticing that*

$$(5.80) \qquad \exp\left\{ f(Z_t) - f(Z_0) - \int_0^t \frac{\mathcal{A}e^f}{e^f}(Z_s)ds \right\}$$

is a martingale and that N_t explodes if and only if Z_t explodes, we have

$$(5.81) \qquad \lim_{t \to \infty} \frac{1}{t} \log \mathbb{P}^{\varnothing}(\tau \geq t) = -\sigma,$$

where σ is the principal eigenvalue for

$$(5.82) \qquad \mathcal{A}u = -\sigma u, \quad u \geq 1.$$

Note that here you have to choose the test function $u \geq 1$ rather than $u \geq 0$.

Proposition 6. *Assume that $\lambda(z) = \gamma z^k + \delta$, where $\gamma, \delta > 0$ and $k > 1$. According to Theorem 24, it is in the explosive regime. We have the following asymptotics*

for small time ϵ.

$$(5.83) \qquad \lim_{\epsilon \to 0} \epsilon^{\frac{1}{k-1}} \log \mathbb{P}(\tau \le \epsilon) = C_k^{\frac{k}{k-1}} (k^{-\frac{1}{k-1}} - k^{-\frac{k}{k-1}}),$$

where $C_k = \int_0^\infty \log \left(\frac{\gamma y^k h(0)^k}{\gamma y^k h(0)^k + 1} \right) dy.$

Before we proceed, let us first quote de Bruijn's Tauberian theorem from the book by Bingham, Goldie and Teugels [9], which will be used in the proof of Proposition 6.

Theorem 25 (de Bruijn's Tauberian theorem). *Let μ be a measure on $(0, \infty)$ whose Laplace transform $M(\lambda) := \int_0^\infty e^{-\lambda x} d\mu(x)$ converges for all $\lambda > 0$. If $\alpha < 0$, $\phi \in \mathcal{R}_\alpha(0+)$, i.e. $\phi(\lambda t)/\phi(t) \sim \lambda^\alpha$ as $t \sim 0+$, put $\psi(\lambda) := \phi(\lambda)/\lambda \in \mathcal{R}_{\alpha-1}(0+)$, then, for $B > 0$,*

$$(5.84) \qquad -\log \mu(0, x] \sim \frac{B}{\bar{\phi}(1/x)}, \quad x \to 0+,$$

if and only if

$$(5.85) \qquad -\log M(\lambda) \sim (1 - \alpha) \left(\frac{B}{-\alpha} \right)^{\frac{\alpha}{\alpha-1}} \frac{1}{\bar{\psi}(\lambda)}, \quad \lambda \to \infty.$$

Here, $\bar{\phi}(\lambda) := \sup\{t : \phi(t) > \lambda\}$ and similarly for $\bar{\psi}$.

Proof of Proposition 6. First, let us observe that since we are considering the event $\{\tau \le \epsilon\}$ for $\epsilon > 0$ very small. It is sufficient to consider the point process with intensity $\lambda(h(0)N_{t-})$ at time t.

To apply de Bruijin's Tauberian theorem, notice that

$$(5.86) \qquad -\log M(\sigma) = -\sum_{i=0}^{\infty} \log\left(\frac{\lambda(ih(0))}{\lambda(ih(0)) + \sigma}\right).$$

Recall that $\lambda(z) = \gamma z^k + \delta$, where $\gamma, \delta > 0$ and $k > 1$. Then,

$$(5.87) \qquad -\log M(\sigma) = -\sum_{i=0}^{\infty} \log\left(\frac{\gamma i^k h(0)^k + \delta}{\gamma i^k h(0)^k + \delta + \sigma}\right)$$

$$\geq -\int_1^{\infty} \log\left(\frac{\gamma x^k h(0)^k + \delta}{\gamma x^k h(0)^k + \delta + \sigma}\right) dx$$

$$= -\sigma^{1/k} \int_{1/\sigma^{1/k}}^{\infty} \log\left(\frac{\gamma \sigma y^k h(0)^k + \delta}{\gamma \sigma y^k h(0)^k + \delta + \sigma}\right) dy$$

$$\sim -\sigma^{1/k} \int_0^{\infty} \log\left(\frac{\gamma y^k h(0)^k}{\gamma y^k h(0)^k + 1}\right) dy, \quad \text{as } \sigma \to \infty.$$

Similarly,

$$(5.88) \qquad -\log M(\sigma) \leq -\int_0^{\infty} \log\left(\frac{\gamma x^k h(0)^k + \delta}{\gamma x^k h(0)^k + \delta + \sigma}\right) dx$$

$$\sim -\sigma^{1/k} \int_0^{\infty} \log\left(\frac{\gamma y^k h(0)^k}{\gamma y^k h(0)^k + 1}\right) dy \quad \text{as } \sigma \to \infty.$$

Now let $C_k = \int_0^{\infty} \log\left(\frac{\gamma y^k h(0)^k}{\gamma y^k h(0)^k + 1}\right) dy$, $\phi(t) = t^{1-k}$, $\psi(t) = t^{-k}$ and $\alpha = 1 - k < 0$. Then $\bar{\phi}(1/\epsilon) = (1/\epsilon)^{-\frac{1}{k-1}}$ and $\bar{\psi}(\sigma) = \sigma^{-\frac{1}{k}}$. To apply the theorem, we need to solve B such that

$$(5.89) \qquad (1 - \alpha)\left(\frac{B}{-\alpha}\right)^{\frac{\alpha}{\alpha-1}} = k\left(\frac{B}{k-1}\right)^{\frac{k-1}{k}} = C_k,$$

170

for $B = (k-1)(C_k/k)^{\frac{k}{k-1}}$. Therefore,

$$(5.90) \qquad \lim_{\epsilon \to 0} \epsilon^{\frac{1}{k-1}} \log \mathbb{P}(\tau \le \epsilon) = C_k'^{\frac{k}{k-1}}(k^{-\frac{1}{k-1}} - k^{-\frac{k}{k-1}}).$$

$\square$

Chapter 6

Limit Theorems for Marked

Hawkes Processes

6.1 Introduction and Main Results

6.1.1 Introduction

We consider in this chapter a linear Hawkes process with random marks. Let N_t be a simple point process. N_t denotes the number of points in the interval $[0, t)$. Let $\mathcal{F}_t$ be the natural filtration up to time t. We assume that $N(-\infty, 0] = 0$. At time t, the point process has $\mathcal{F}_t$-predictable intensity

$$(6.1) \qquad \lambda_t := \nu + Z_t, \quad Z_t := \sum_{\tau_i < t} h(t - \tau_i, a_i),$$

where $\nu > 0$, the $(\tau_i)_{i \geq 1}$ are arrival times of the points, and the $(a_i)_{i \geq 1}$ are i.i.d. random marks, a_i being independent of previous arrival times τ_j, $j \leq i$. Let us assume that a_i has a common distribution $q(da)$ on a metric space $\mathbb{X}$. Here,

$h(\cdot, \cdot) : \mathbb{R}^+ \times \mathbb{X} \to \mathbb{R}^+$ is integrable, i.e. $\int_0^\infty \int_{\mathbb{X}} h(t, a) q(da) dt < \infty$. Let $H(a) := \int_0^\infty h(t, a) dt$ for any $a \in \mathbb{X}$. We also assume that

$$(6.2) \qquad \int_{\mathbb{X}} H(a) q(da) < 1.$$

Let $\mathbb{P}^q$ denote the probability measure for the a_i's with the common law $q(da)$. Under assumption (6.2), it is well known that there exists a unique stationary version of the linear marked Hawkes process satisfying the dynamics (6.1) and that by ergodic theorem, a law of large numbers holds,

$$(6.3) \qquad \lim_{t \to \infty} \frac{N_t}{t} = \frac{\nu}{1 - \mathbb{E}^q[H(a)]}.$$

This chapter is organized as follows. In Section 6.1.2, we will introduce the main results of this paper, i.e. the central limit theorem and the large deviation principle for linear marked Hawkes processes. The proof of the central limit theorem will be given in Section 6.2 and the proof of the large deviation principle will be given in Section 6.3. Finally, we will discuss an application of our results to a risk model in finance in Section 6.4.

6.1.2 Main Results

For a linear marked Hawkes process satisfying the dynamics (6.1), we have the following large deviation principle.

Theorem 26 (Large Deviation Principle)**.** *Assume the conditions* (6.2) *and*

$$(6.4) \qquad \lim_{x \to \infty} \left\{ \int_{\mathbb{X}} e^{H(a)x} q(da) - x \right\} = \infty.$$

Then, $(N_t/t \in \cdot)$ satisfies a large deviation principle with rate function,

$$\Lambda(x) := \begin{cases} \inf_{\hat{q}} \left\{ x\mathbb{E}^{\hat{q}}[H(a)] + \nu - x + x \log \left(\frac{x}{x\mathbb{E}^{\hat{q}}[H(a)]+\nu} \right) + x\mathbb{E}^{\hat{q}} \left[\log \frac{d\hat{q}}{dq} \right] \right\} & x \geq 0 \\ +\infty & x < 0 \end{cases}$$

$$= \begin{cases} \theta_* x - \nu(x_* - 1) & x \geq 0 \\ +\infty & x < 0 \end{cases},$$

where the infimum of $\hat{q}$ is taken over $\mathcal{M}(\mathbb{X})$, the space of probability measures on $\mathbb{X}$ such that $\hat{q}$ is absolutely continuous w.r.t. q. Here, θ_* and x_* satisfy the following equations

$$(6.5) \qquad \begin{cases} x_* = \mathbb{E}^q \left[e^{\theta_* + (x_* - 1)H(a)} \right] \\ \frac{x}{\nu} = x_* + \frac{x}{\nu}\mathbb{E}^q \left[H(a)e^{\theta_* + (x_* - 1)H(a)} \right] \end{cases}.$$

Theorem 27 (Central Limit Theorem). *Assume* $\lim_{t \to \infty} t^{1/2} \int_t^\infty \mathbb{E}^q[h(s,a)]ds = 0$ *and that* (6.2) *holds. Then,*

$$(6.6) \qquad \frac{N_t - \frac{\nu t}{1 - \mathbb{E}^q[H(a)]}}{\sqrt{t}} \to N \left(0, \frac{\nu(1 + Var^q[H(a)])}{(1 - \mathbb{E}^q[H(a)])^3} \right),$$

in distribution as $t \to \infty$.

6.2 Proof of Central Limit Theorem

Proof of Theorem 27. First, let us observe that

$$(6.7) \qquad \int_0^t \lambda_s ds = \nu t + \sum_{\tau_i < t} \int_{\tau_i}^t h(s - \tau_i, a_i) ds$$

$$= \nu t + \sum_{\tau_i < t} H(a_i) - \mathcal{E}_t,$$

where the error term $\mathcal{E}_t$ is given by

$$(6.8) \qquad \mathcal{E}_t := \sum_{\tau_i < t} \int_t^\infty h(s - \tau_i, a_i) ds.$$

Therefore,

(6.9)

$$\frac{N_t - \int_0^t \lambda_s ds}{\sqrt{t}} = \frac{N_t - \nu t - \sum_{\tau_i < t} H(a_i)}{\sqrt{t}} + \frac{\mathcal{E}_t}{\sqrt{t}}$$

$$= (1 - \mathbb{E}^q[H(a)]) \frac{N_t - \mu t}{\sqrt{t}} + \frac{\mathbb{E}^q[H(a)]N_t - \sum_{\tau_i < t} H(a_i)}{\sqrt{t}} + \frac{\mathcal{E}_t}{\sqrt{t}},$$

where $\mu := \frac{\nu}{1 - \mathbb{E}^q[H(a)]}$. Rearranging the terms in (6.9), we get

(6.10)

$$\frac{N_t - \mu t}{\sqrt{t}} = \frac{1}{1 - \mathbb{E}^q[H(a)]} \left[\frac{N_t - \int_0^t \lambda_s ds}{\sqrt{t}} + \frac{\sum_{\tau_i < t}(H(a_i) - \mathbb{E}^q[H(a)])}{\sqrt{t}} - \frac{\mathcal{E}_t}{\sqrt{t}} \right].$$

It is easy to check that $\frac{\mathcal{E}_t}{\sqrt{t}} \to 0$ in probability as $t \to \infty$. To see this, first notice that $\mathbb{E}[\lambda_t] \leq \frac{\nu}{1 - \mathbb{E}^q[H(a)]}$ uniformly in t. Let $g(t, a) := \int_t^\infty h(s, a) ds$. We have

$\mathcal{E}_t = \sum_{\tau_i < t} g(t - \tau_i, a_i)$ and thus

$$(6.11) \qquad \mathbb{E}[\mathcal{E}_t] = \int_0^t \int_{\mathbb{X}} g(t - s, a) q(da) \mathbb{E}[\lambda_s] ds$$

$$\leq \frac{\nu}{1 - \mathbb{E}^q[H(a)]} \int_0^t \int_{\mathbb{X}} g(t - s, a) q(da) ds$$

$$= \frac{\nu}{1 - \mathbb{E}^q[H(a)]} \int_0^t \mathbb{E}^q[g(s, a)] ds.$$

Hence, by L'Hôpital's rule,

$$(6.12) \qquad \lim_{t \to \infty} \frac{1}{t^{1/2}} \int_0^t \mathbb{E}^q[g(s, a)] ds = \lim_{t \to \infty} \frac{\mathbb{E}^q[g(t, a)]}{\frac{1}{2} t^{-1/2}}$$

$$= \lim_{t \to \infty} 2t^{1/2} \int_t^\infty \mathbb{E}^q[h(s, a)] ds = 0.$$

Hence, $\frac{\mathcal{E}_t}{\sqrt{t}} \to 0$ in probability as $t \to \infty$.

Furthermore, $M_1(t) := N_t - \int_0^t \lambda_s ds$ and $M_2(t) := \sum_{\tau_i < t} (H(a_i) - \mathbb{E}^q[H(a)])$ are both martingales.

Moreover, since $\int_0^t \lambda_s ds$ is of finite variation, the quadratic variation of $M_1(t) + M_2(t)$ is the same as the quadratic variation of $N_t + M_2(t)$. And notice that $N_t + M_2(t) = \sum_{\tau_i < t} (1 + H(a_i) - \mathbb{E}^q[H(a)])$ which has quadratic variation

$$(6.13) \qquad \sum_{\tau_i < t} (1 + H(a_i) - \mathbb{E}^q[H(a)])^2.$$

By the standard law of large numbers, we have

(6.14)
$$\frac{1}{t}\sum_{\tau_i<t}(1+H(a_i)-\mathbb{E}^q[H(a)]) = \frac{N_t}{t}\cdot\frac{1}{N_t}\sum_{\tau_i<t}(1+H(a_i)-\mathbb{E}^q[H(a)])^2$$
$$\to \frac{\nu}{1-\mathbb{E}^q[H(a)]}\cdot\mathbb{E}^q\left[(1+H(a)-\mathbb{E}^q[H(a)])^2\right]$$
$$= \frac{\nu(1+\mathrm{Var}^q[H(a)])}{1-\mathbb{E}^q[H(a)]},$$

a.s. as $t\to\infty$. By a standard martingale central limit theorem, we conclude that

(6.15)
$$\frac{N_t - \frac{\nu t}{1-\mathbb{E}^q[H(a)]}}{\sqrt{t}} \to N\left(0, \frac{\nu(1+\mathrm{Var}^q[H(a)])}{(1-\mathbb{E}^q[H(a)])^3}\right),$$

in distribution as $t\to\infty$. $\qquad\square$

6.3 Proof of Large Deviation Principle

6.3.1 Limit of a Logarithmic Moment Generating Function

In this subsection, we prove the existence of the limit of the logarithmic moment generating function $\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]$ and give a variational formula and a more explicit formula for this limit.

Theorem 28. *The limit $\Gamma(\theta)$ of the logarithmic moment generating function is*

(6.16)
$$\Gamma(\theta) = \lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}] = \begin{cases} \nu(f(\theta)-1) & \text{if } \theta\in(-\infty,\theta_c] \\ +\infty & \text{otherwise} \end{cases},$$

where $f(\theta)$ is the minimal solution to $x = \int_{\mathbb{X}} e^{\theta + H(a)(x-1)} q(da)$ and

$$(6.17) \qquad \theta_c = -\log \int_{\mathbb{X}} H(a) e^{H(a)(x_c - 1)} q(da) > 0,$$

where $x_c > 1$ satisfies the equation $x \int_{\mathbb{X}} H(a) e^{H(a)(x-1)} q(da) = \int_{\mathbb{X}} e^{H(a)(x-1)} q(da)$.

We will break the proof of Theorem 28 into the proof of the lower bound, i.e. Lemma 34 and the proof of the upper bound, i.e. Lemma 35.

Before we proceed, let us first prove Lemma 33, which will be repeatedly used.

Lemma 33. *Consider a linear marked Hawkes process with intensity*

$$(6.18) \qquad \lambda_t := \alpha + \beta Z_t := \alpha + \beta \sum_{\tau_i < t} h(t - \tau_i, a_i),$$

and $\beta \mathbb{E}^q[H(a)] < 1$, where the a_i are i.i.d. random marks with the common law $q(da)$ independent of the previous arrival times, then there exists a unique invariant measure π for Z_t such that

$$(6.19) \qquad \int \lambda(z) \pi(dz) = \frac{\alpha}{1 - \beta \mathbb{E}^q[H(a)]}.$$

Proof. The ergodicity of Z_t is well known. Let π be the invariant probability measure for Z_t. Then

$$(6.20) \qquad \int \lambda(z) \pi(dz) = \alpha + \beta \int_{\mathbb{X}} \int_0^\infty h(t, a) dt q(da) \int \lambda(z) \pi(dz).$$

$\square$

Lemma 34 (Lower Bound).

$$
(6.21) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \geq \begin{cases} \nu(f(\theta) - 1) & \text{if } \theta \in (-\infty, \theta_c] \\ +\infty & \text{otherwise} \end{cases},
$$

where $f(\theta)$ is the minimal solution to $x = \int e^{\theta + H(a)(x-1)} q(da)$ and θ_c is defined in (6.17).

Proof. The intensity at time t is $\lambda_t := \lambda(Z_t)$ where $\lambda(z) = \nu + z$ and $Z_t = \sum_{\tau_i < t} h(t - \tau_i, a_i)$. We tilt λ to $\hat{\lambda}$ and q to $\hat{q}$ such that by Girsanov formula the tilted probability measure $\hat{\mathbb{P}}$ is given by

$$
(6.22)
$$
$$
\frac{d\hat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = \exp\left\{ \int_0^t (\lambda(Z_s) - \hat{\lambda}(Z_s))ds + \int_0^t \log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right) + \log\left(\frac{d\hat{q}}{dq}\right) dN_s \right\}.
$$

Let $\mathcal{Q}_e$ be the set of $(\hat{\lambda}, \hat{q}, \hat{\pi})$ such that the marked Hawkes process with intensity $\hat{\lambda}(Z_t)$ and random marks distributed as $\hat{q}$ is ergodic with $\hat{\pi}$ as the invariant measure of Z_t.

By the ergodic theorem and Jensen's inequality, for any $(\hat{\lambda}, \hat{q}, \hat{\pi}) \in \mathcal{Q}_e$,

$$
(6.23) \quad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]
$$
$$
\geq \liminf_{t \to \infty} \hat{\mathbb{E}}\left[\frac{1}{t} \theta N_t - \frac{1}{t} \int_0^t (\lambda - \hat{\lambda}) ds - \frac{1}{t} \int_0^t \left[\log(\hat{\lambda}/\lambda) + \log(d\hat{q}/dq) \right] \hat{\lambda} ds \right]
$$
$$
= \int \theta \hat{\lambda} \hat{\pi}(dz) + \int (\hat{\lambda} - \lambda) \hat{\pi}(dz) - \iint \left[\log(\hat{\lambda}/\lambda) + \log(d\hat{q}/dq) \right] \hat{\lambda} \hat{q} \hat{\pi}(dz).
$$

Hence,

(6.24)

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$$

$$\geq \sup_{(\hat{\lambda},\hat{q},\hat{\pi})\in\mathcal{Q}_e} \left\{ \int \theta \hat{\lambda}\hat{\pi} + \int (\hat{\lambda}-\lambda)\hat{\pi} - \iint \left[\log(\hat{\lambda}/\lambda) + \log(d\hat{q}/dq) \right] \hat{\lambda}\hat{q}\hat{\pi} \right\}.$$

$$\geq \sup_{(K\lambda,\hat{q},\hat{\pi})\in\mathcal{Q}_e} \int \left[\left(\theta - \mathbb{E}^{\hat{q}}[\log(d\hat{q}/dq)] \right) \hat{\lambda} + \hat{\lambda} - \lambda - \hat{\lambda} \log\left(\hat{\lambda}/\lambda\right) \right] \hat{\pi}$$

$$\geq \sup_{0<K<\mathbb{E}^{\hat{q}}[H(a)]^{-1},(K\lambda,\hat{q},\hat{\pi})\in\mathcal{Q}_e} \int \left[\left(\theta - \mathbb{E}^{\hat{q}}[\log(d\hat{q}/dq)] \right) + 1 - \frac{1}{K} - \log K \right] \hat{\lambda}\hat{\pi}$$

$$= \sup_{\hat{q}} \sup_{0<K<\mathbb{E}^{\hat{q}}[H(a)]^{-1}} \left[\left(\theta - \mathbb{E}^{\hat{q}}[\log(d\hat{q}/dq)] \right) + 1 - \frac{1}{K} - \log K \right] \cdot \frac{K\nu}{1 - K\mathbb{E}^{\hat{q}}[H(a)]},$$

where the last equality is obtained by applying Lemma 33. The supremum of $\hat{q}$ is taken over $\mathcal{M}(\mathbb{X})$ such that $\hat{q}$ is absolutely continuous w.r.t. q. Optimizing over $K > 0$, we get

(6.25)

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}]$$

$$\geq \begin{cases} \sup_{\hat{q}} \nu(\hat{f}(\theta) - 1) & \text{if } \theta \in \left(-\infty, \mathbb{E}^{\hat{q}} \left[\log \frac{d\hat{q}}{dq} \right] + \mathbb{E}^{\hat{q}}[H(a)] - 1 - \log \mathbb{E}^{\hat{q}}[H(a)] \right] \\ +\infty & \text{otherwise} \end{cases},$$

where $\hat{f}(\theta)$ is the minimal solution to the equation

(6.26)
$$x = e^{\theta + \mathbb{E}^{\hat{q}}[\log(dq/d\hat{q})] + \mathbb{E}^{\hat{q}}[H(a)](x-1)}$$

$$\leq \mathbb{E}^{\hat{q}} \left[e^{\theta + H(a)(x-1)} \frac{dq}{d\hat{q}} \right] = \int e^{\theta + H(a)(x-1)} q(da).$$

180

The last inequality is satisfied by Jensen's inequality; the equality holds if and only if

$$(6.27) \qquad \frac{d\hat{q}}{dq} = \frac{e^{H(a)(x-1)}}{\mathbb{E}^q[e^{H(a)(x-1)}]}.$$

Optimizing over $\hat{q}$, we get

$$(6.28) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \geq \begin{cases} \nu(f(\theta) - 1) & \text{if } \theta \in (-\infty, \theta_c] \\ +\infty & \text{otherwise,} \end{cases}$$

where θ_c is some critical value to be determined. Let

$$(6.29) \qquad G(x) = e^\theta \int e^{H(a)(x-1)} q(da) - x.$$

If $\theta = 0$, then $G(x) = \int e^{H(a)(x-1)} q(da) - x$ satisfies $G(1) = 0$, $G(\infty) = \infty$ (by (6.4)) and $G'(1) = \mathbb{E}^q[H(a)] - 1 < 0$ which implies $\min_{x>1} G(x) < 0$. Hence, there exists some critical $\theta_c > 0$ such that $\min_{x>1} G(x) = 0$. The critical values x_c and θ_c satisfy $G(x_c) = G'(x_c) = 0$, which implies

$$(6.30) \qquad \theta_c = -\log \int H(a) e^{H(a)(x_c - 1)} q(da),$$

where $x_c > 1$ satisfies the equation $x \int H(a) e^{H(a)(x-1)} q(da) = \int e^{H(a)(x-1)} q(da)$.

It is easy to check that indeed, for $dq_* = \frac{e^{H(a)(x_* - 1)}}{\mathbb{E}^q[e^{H(a)(x_* - 1)}]} dq$,

$$(6.31)$$
$$\mathbb{E}^{q_*}\left[\log \frac{dq_*}{dq}\right] + \mathbb{E}^{q_*}[H(a)] - 1 - \log \mathbb{E}^{q_*}[H(a)]) = -\log \int H(a) e^{H(a)(x_* - 1)} q(da).$$

□

Lemma 35 (Upper Bound).

$$(6.32) \qquad \limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \leq \begin{cases} \nu(f(\theta) - 1) & \textit{if } \theta \in (-\infty, \theta_c] \\ +\infty & \textit{otherwise} \end{cases},$$

where $f(\theta)$ is the minimal solution to $x = \int e^{\theta + H(a)(x-1)} q(da)$ and θ_c is defined in (6.17).

Proof. It is well known that a linear Hawkes process has an immigration-birth representation. The immigrants (roots) arrive via a standard Poisson process with constant intensity $\nu > 0$. Each immigrant generates children according to a Galton-Watson tree. (See for example Hawkes and Oakes [54] and Karabash [63].) Consider a random, rooted tree (with root, i.e. immigrant, at time 0) associated to the Hawkes process via the Galton-Watson interpretation. Note the root is unmarked at the start of the process so the marking goes into the expectation calculation later. Let K be the number of children of the root node, and let $S_t^{(1)}, S_t^{(2)}, \ldots, S_t^{(K)}$ be the number of descendants of root's k-th child that were born before time t (including k-th child if an only if it was born before time t). Let S_t be the total number of children in tree before time t including root node.

Then

(6.33) $F_S(t) := \mathbb{E}[\exp(\theta S_t)]$

$$= \sum_{k=0}^{\infty} \mathbb{E}[\exp(\theta S_t)|K = k]\mathbb{P}(K = k)$$

$$= \exp(\theta) \sum_{k=0}^{\infty} \mathbb{P}(K = k) \prod_{i=1}^{k} \mathbb{E}\left[\exp\left(\theta S_t^{(i)}\right)\right]$$

$$= \exp(\theta) \sum_{k=0}^{\infty} \mathbb{E}\left[\exp\left(\theta S_t^{(1)}\right)\right]^k \mathbb{P}(K = k)$$

$$= \exp(\theta) \sum_{k=0}^{\infty} \int_{\mathbb{X}} \left[\left(\int_0^t \frac{h(s,a)}{H(a)} F_S(t-s)ds\right)^k e^{-H(a)} \frac{H(a)^k}{k!}\right] q(da)$$

$$= \int_{\mathbb{X}} \exp\left(\theta + \int_0^t h(s,a)(F_S(t-s) - 1)ds\right) q(da).$$

Now observe that $F_S(t)$ is strictly increasing and hence must approach to the smaller solution x_* of the following equation

(6.34) $$x = \int_{\mathbb{X}} \exp\left[\theta + H(a)(x - 1)\right] q(da).$$

Finally, since random roots arrive according to a Poisson process with constant intensity $\nu > 0$, we have

(6.35) $$F_N(t) := \mathbb{E}[\exp(\theta N_t)] = \exp\left[\nu \int_0^t (F_S(t-s) - 1)ds\right].$$

But since $F_S(s) \uparrow x_*$ as $s \to \infty$ we obtain the main result

(6.36) $$\frac{1}{t} \log F_N(t) = \nu \frac{1}{t} \left[\int_0^t (F_S(s) - 1)\, ds\right] \xrightarrow{t \to \infty} \nu(x_* - 1),$$

which proves the desired formula. Note that $x_* = \infty$ when there is no solution to

183

(6.34). The proof is complete. □

6.3.2 Large Deviation Principle

In this section, we prove the main result, i.e. Theorem 26 by using the Gärtner-Ellis theorem for the upper bound and tilting method for the lower bound.

Proof of Theorem 26. For the upper bound, since we have Theorem 28, we can simply apply Gärtner-Ellis theorem. To prove the lower bound, it suffices to show that for any $x > 0$, $\epsilon > 0$, we have

$$(6.37) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq -\sup_{\theta \in \mathbb{R}} \{\theta x - \Gamma(\theta)\},$$

where $B_\epsilon(x)$ denotes the open ball centered at x with radius ϵ. Let $\hat{\mathbb{P}}$ denote the tilted probability measure with rate $\hat{\lambda}$ and marks distributed by $\hat{q}(da)$ as defined in Lemma 34. By Jensen's inequality,

$$
\begin{aligned}
(6.38) \quad & \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \\
& \geq \frac{1}{t} \log \int_{\frac{N_t}{t} \in B_\epsilon(x)} \frac{d\mathbb{P}}{d\hat{\mathbb{P}}} d\hat{\mathbb{P}} \\
& = \frac{1}{t} \log \hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) - \frac{1}{t} \log \left[\frac{1}{\hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right)} \int_{\frac{N_t}{t} \in B_\epsilon(x)} \frac{d\hat{\mathbb{P}}}{d\mathbb{P}} d\hat{\mathbb{P}} \right] \\
& \geq \frac{1}{t} \log \hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) - \frac{1}{\hat{\mathbb{P}}\left(\frac{N_t}{t} \in B_\epsilon(x)\right)} \cdot \frac{1}{t} \cdot \hat{\mathbb{E}}\left[1_{\frac{N_t}{t} \in B_\epsilon(x)} \log \frac{d\hat{\mathbb{P}}}{d\mathbb{P}} \right].
\end{aligned}
$$

By the ergodic theorem,

$$(6.39) \qquad \liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P}\left(\frac{N_t}{t} \in B_\epsilon(x)\right) \geq - \inf_{\substack{0 < K < \mathbb{E}^{\hat{q}}[H(a)]^{-1} \\ (K\lambda, \hat{q}, \hat{\pi}) \in \mathcal{Q}_e^x}} \mathcal{H}(\hat{\lambda}, \hat{q}, \hat{\pi}).$$

where $\mathcal{Q}_e^x$ is defined by

$$(6.40) \qquad \mathcal{Q}_e^x = \left\{ (\hat{\lambda}, \hat{q}, \hat{\pi}) \in \mathcal{Q}_e : \int \hat{\lambda}(z) \hat{\pi}(dz) = x \right\}.$$

and the relative entropy $\mathcal{H}$ is

$$(6.41) \qquad \mathcal{H}(\hat{\lambda}, \hat{q}, \hat{\pi}) = \int (\lambda - \hat{\lambda}) \hat{\pi} + \int \log(\hat{\lambda}/\lambda) \hat{\lambda} \hat{\pi} + \iint \log(d\hat{q}/dq) \hat{q} \hat{\lambda} \hat{\pi}.$$

By Lemma 33,

$$(6.42) \qquad \inf_{0 < K < \mathbb{E}^{\hat{q}}[H(a)]^{-1}, x = \frac{\nu K}{1 - K \mathbb{E}^{\hat{q}}[H(a)]}, (K\lambda, \hat{q}, \hat{\pi}) \in \mathcal{Q}_e} \mathcal{H}(\hat{\lambda}, \hat{q}, \hat{\pi})$$

$$= \inf_{K = \frac{x}{x \mathbb{E}^{\hat{q}}[H(a)] + \nu}, (K\lambda, \hat{q}, \hat{\pi}) \in \mathcal{Q}_e} \left\{ \frac{1}{K} - 1 + \log K + \mathbb{E}^{\hat{q}} \left[\log \frac{d\hat{q}}{dq} \right] \right\} \int \hat{\lambda} \hat{\pi}$$

$$= \inf_{\hat{q}} \left\{ \mathbb{E}^{\hat{q}}[H(a)] + \frac{\nu}{x} - 1 + \log \left(\frac{x}{x \mathbb{E}^{\hat{q}}[H(a)] + \nu} \right) + \mathbb{E}^{\hat{q}} \left[\log \frac{d\hat{q}}{dq} \right] \right\} x$$

$$= \inf_{\hat{q}} \left\{ x \mathbb{E}^{\hat{q}}[H(a)] + \nu - x + x \log \left(\frac{x}{x \mathbb{E}^{\hat{q}}[H(a)] + \nu} \right) + x \mathbb{E}^{\hat{q}} \left[\log \frac{d\hat{q}}{dq} \right] \right\}.$$

Next, let us find a more explict form for the Legendre-Fenchel transform of $\Gamma(\theta)$.

$$(6.43) \qquad \sup_{\theta \in \mathbb{R}} \{ \theta x - \Gamma(\theta) \} = \sup_{\theta \in \mathbb{R}} \{ \theta x - \nu(f(\theta) - 1) \},$$

where $f(\theta) = \mathbb{E}^q[e^{\theta + (f(\theta) - 1)H(a)}]$. Here,

$$(6.44) \qquad f'(\theta) = \mathbb{E}^q \left[(1 + f'(\theta) H(a)) e^{\theta + (f(\theta) - 1) H(a)} \right].$$

So the optimal θ_* for (6.43) would satisfy $f'(\theta_*) = \frac{x}{\nu}$ and θ_* and $x_* = f(\theta_*)$ satisfy

185

the following equations

$$(6.45) \qquad \begin{cases} x_* = \mathbb{E}^q \left[e^{\theta_* + (x_* - 1)H(a)} \right] \\[2mm] \frac{x}{\nu} = x_* + \frac{x}{\nu}\mathbb{E}^q \left[H(a)e^{\theta_* + (x_* - 1)H(a)} \right] \end{cases},$$

and $\sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma(\theta)\} = \theta_* x - \nu(x_* - 1)$.

On the other hand, letting $dq_* = \frac{e^{(x_*-1)H(a)}}{\mathbb{E}^q[e^{(x_*-1)H(a)}]}dq$, we have

$$(6.46) \qquad \mathbb{E}^{q_*}[H(a)] = \frac{\mathbb{E}^q \left[e^{\theta_* + (x_* - 1)H(a)} \right]}{\mathbb{E}^q \left[e^{(x_* - 1)H(a)} \right]} = \frac{1}{x_*} - \frac{\nu}{x},$$

and $\mathbb{E}^{q_*}[\log \frac{dq_*}{dq}] = (x_* - 1)\mathbb{E}^{q_*}[H(a)] - \log \mathbb{E}^q[e^{(x_*-1)H(a)}]$, which imply

$$
\begin{aligned}
(6.47) \quad &\liminf_{t \to \infty} \frac{1}{t} \log \mathbb{P} \left(\frac{N_t}{t} \in B_\epsilon(x) \right) \\
&\geq -\inf_{\hat{q}} \left\{ x\mathbb{E}^{\hat{q}}[H(a)] + \nu - x + x\log \left(\frac{x}{x\mathbb{E}^{\hat{q}}[H(a)] + \nu} \right) + x\mathbb{E}^{\hat{q}} \left[\log \frac{d\hat{q}}{dq} \right] \right\} \\
&\geq -\left\{ x\mathbb{E}^{q_*}[H(a)] + \nu - x + x\log \left(\frac{x}{x\mathbb{E}^{q_*}[H(a)] + \nu} \right) + x\mathbb{E}^{q_*} \left[\log \frac{dq_*}{dq} \right] \right\} \\
&= \theta_* x - \nu(x_* - 1) = \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma(\theta)\}.
\end{aligned}
$$

$\square$

6.4   Risk Model with Marked Hawkes Claims Arrivals

We consider the following risk model for the surplus process R_t of an insurance portfolio,

$$
(6.48) \qquad R_t = u + \rho t - \sum_{i=1}^{N_t} C_i,
$$

where $u > 0$ is the initial reserve, $\rho > 0$ is the constant premium and the C_i's are i.i.d. positive random variables with the common distribution $\mu(dC)$. C_i represents the claim size at the ith arrival time, these being independent of N_t, a marked Hawkes process.

For $u > 0$, let

$$
(6.49) \qquad \tau_u = \inf\{t > 0 : R_t \leq 0\},
$$

and denote the infinite and finite horizon ruin probabilities by

$$
(6.50) \qquad \psi(u) = \mathbb{P}(\tau_u < \infty), \quad \psi(u, uz) = \mathbb{P}(\tau_u \leq uz), \quad u, z > 0.
$$

By the law of large numbers,

$$
(6.51) \qquad \lim_{t \to \infty} \frac{1}{t} \sum_{i=1}^{N_t} C_i = \frac{\mathbb{E}^\mu[C]\nu}{1 - \mathbb{E}^q[H(a)]}.
$$

Therefore, to exclude the trivial case, we need to assume that

$$(6.52) \qquad \frac{\mathbb{E}^{\mu}[C]\nu}{1 - \mathbb{E}^q[H(a)]} < \rho < \frac{\nu(x_c) - 1}{\theta_c},$$

where the critical values θ_c and $x_c = f(\theta_c)$ satisfy

$$(6.53) \qquad \begin{cases} x_c = \int_{\mathbb{R}^+} \int_{\mathbb{X}} e^{\theta_c C + H(a)(x_c - 1)} q(da)\mu(dC) \\ 1 = \int_{\mathbb{R}^+} \int_{\mathbb{X}} H(a) e^{H(a)(x_c - 1) + \theta_c C} q(da)\mu(dC) \end{cases}.$$

Let us first assume that the claim sizes following light tails, i.e. there exists some $\theta > 0$ such that $\int_{\mathbb{R}^+} e^{\theta C} \mu(dC) < \infty$.

Following the proofs of large deviation results in Section 6.3, we have

$$(6.54) \qquad \Gamma_C(\theta) := \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}\left[e^{\theta \sum_{i=1}^{N_t} C_i} \right] = \begin{cases} \nu(x - 1) & \text{if } \theta \in (-\infty, \theta_c] \\ +\infty & \text{otherwise} \end{cases},$$

where x is the minimal solution to the equation

$$(6.55) \qquad x = \int_{\mathbb{R}^+} \int_{\mathbb{X}} e^{\theta C + (x - 1)H(a)} q(da)\mu(dC).$$

Before we proceed, let us quote a result from Glynn and Whitt [43], which will be used in our proof Theorem 30.

Theorem 29 (Glynn and Whitt [43]). *Let S_n be random variables. $\tau_u = \inf\{n : S_n > u\}$ and $\psi(u) = \mathbb{P}(\tau_u < \infty)$. Assume that there exist $\gamma, \epsilon > 0$ such that*

(i) $\kappa_n(\theta) = \log \mathbb{E}[e^{\theta S_n}]$ is well defined and finite for $\gamma - \epsilon < \theta < \gamma + \epsilon$.

(ii) $\limsup_{n \to \infty} \mathbb{E}[e^{\theta(S_n - S_{n-1})}] < \infty$ for $-\epsilon < \theta < \epsilon$.

188

(iii) $\kappa(\theta) = \lim_{n \to \infty} \frac{1}{n} \kappa_n(\theta)$ *exists and is finite for* $\gamma - \epsilon < \theta < \gamma + \epsilon$.

(iv) $\kappa(\gamma) = 0$ *and* κ *is differentiable at* γ *with* $0 < \kappa'(\gamma) < \infty$.

Then, $\lim_{u \to \infty} \frac{1}{u} \log \psi(u) = -\gamma$.

Remark 15. *We claim that* $\Gamma_C(\theta) = \rho\theta$ *has a unique positive solution* $\theta^\dagger < \theta_c$. *Let* $G(\theta) = \Gamma_C(\theta) - \rho\theta$. *Notice that* $G(0) = 0$, $G(\infty) = \infty$, *and that* G *is convex. We also have* $G'(0) = \frac{\mathbb{E}^\mu[C]\nu}{1 - \mathbb{E}^q[H(a)]} - \rho < 0$ *and* $\Gamma_C(\theta_c) - \rho\theta_c > 0$ *since we assume that* $\rho < \frac{\nu(f(\theta_c) - 1)}{\theta_c}$. *Therefore, there exists only one solution* $\theta^\dagger \in (0, \theta_c)$ *of* $\Gamma_C(\theta^\dagger) = \rho\theta^\dagger$.

Theorem 30 (Infinite Horizon)**.** *Assume all the assumptions in Theorem 26 and in addition* (6.52), *we have* $\lim_{u \to \infty} \frac{1}{u} \log \psi(u) = -\theta^\dagger$, *where* $\theta^\dagger \in (0, \theta_c)$ *is the unique positive solution of* $\Gamma_C(\theta) = \rho\theta$.

Proof. Take $S_t = \sum_{i=1}^{N_t} C_i - \rho t$ and $\kappa_t(\theta) = \log \mathbb{E}[e^{\theta S_t}]$. Then $\lim_{t \to \infty} \frac{1}{t} \kappa_t(\theta) = \Gamma_C(\theta) - \rho\theta$. Consider $\{S_{nh}\}_{n \in \mathbb{N}}$. We have $\lim_{n \to \infty} \frac{1}{n} \kappa_{nh}(\theta) = h\Gamma_C(\theta) - h\rho\theta$. Checking the conditions in Theorem 29 and applying it, we get

$$(6.56) \qquad \lim_{u \to \infty} \frac{1}{u} \log \mathbb{P} \left(\sup_{n \in \mathbb{N}} S_{nh} > u \right) = -\theta^\dagger.$$

Finally, notice that

$$(6.57) \qquad \sup_{t \in \mathbb{R}^+} S_t \geq \sup_{n \in \mathbb{N}} S_{nh} \geq \sup_{t \in \mathbb{R}^+} S_t - \rho h.$$

Hence, $\lim_{u \to \infty} \frac{1}{u} \log \psi(u) = -\theta^\dagger$. $\qquad \square$

Theorem 31 (Finite Horizon)**.** *Under the same assumptions as in Theorem 30, we have*

$$(6.58) \qquad \lim_{u \to \infty} \frac{1}{u} \log \psi(u, uz) = -w(z), \quad \text{for any } z > 0.$$

Here

$$(6.59) \qquad w(z) = \begin{cases} z\Lambda_C\left(\frac{1}{z} + \rho\right) & \text{if } 0 < z < \frac{1}{\Gamma'(\theta^\dagger) - \rho} \\ \theta^\dagger & \text{if } z \geq \frac{1}{\Gamma'(\theta^\dagger) - \rho} \end{cases},$$

$\Lambda_C(x) = \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma_C(\theta)\}$ *and* $\theta^\dagger \in (0, \theta_c)$ *is the unique positive solution of* $\Gamma_C(\theta) = \rho\theta$, *as before.*

Proof. The proof is similar to that in Stabile and Torrisi [102] and we omit it here. $\qquad\qquad\square$

Next, we are interested to study the case when the claim sizes have heavy tails, i.e. $\int_{\mathbb{R}^+} e^{\theta C}\mu(dC) = +\infty$ for any $\theta > 0$.

A distribution function B is subexponential, i.e. $B \in \mathcal{S}$ if

$$(6.60) \qquad \lim_{x \to \infty} \frac{\mathbb{P}(C_1 + C_2 > x)}{\mathbb{P}(C_1 > x)} = 2,$$

where C_1, C_2 are i.i.d. random variables with distribution function B. Let us denote $B(x) := \mathbb{P}(C_1 \geq x)$ and let us assume that $\mathbb{E}[C_1] < \infty$ and define $B_0(x) := \frac{1}{\mathbb{E}[C]}\int_0^x \overline{B}(y)dy$, where $\overline{F}(x) = 1 - F(x)$ is the complement of any distribution function $F(x)$.

Goldie and Resnick [44] showed that if $B \in \mathcal{S}$ and satisfies some smoothness conditions, then B belongs to the maximum domain of attraction of either the Frechet distribution or the Gumbel distribution. In the former case, $\overline{B}$ is regularly varying, i.e. $\overline{B}(x) = L(x)/x^{\alpha+1}$, for some $\alpha > 0$ and we write it as $\overline{B} \in \mathcal{R}(-\alpha - 1)$, $\alpha > 0$.

We assume that $B_0 \in \mathcal{S}$ and either $\overline{B} \in \mathcal{R}(-\alpha - 1)$ or $B \in \mathcal{G}$, i.e. the maximum

190

domain of attraction of Gumbel distribution. $\mathcal{G}$ includes Weibull and lognormal distributions.

When the arrival process N_t satisfies a large deviation result, the probability that it deviates away from its mean is exponentially small, which is dominated by subexonential distributions. By using the techniques for the asymptotics of ruin probabilities for risk processes with non-stationary, non-renewal arrivals and subexponential claims from Zhu [118], we have the following infinite-horizon and finite-horizon ruin probability estimates when the claim sizes are subexponential.

Theorem 32. *Assume the net profit condition* $\rho > \mathbb{E}[C_1] \frac{\nu}{1 - \mathbb{E}^q[H(a)]}$.

(i) (Infinite-Horizon)

$$(6.61) \qquad \lim_{u \to \infty} \frac{\psi(u)}{\overline{B}_0(u)} = \frac{\nu \mathbb{E}[C_1]}{\rho(1 - \mathbb{E}^q[H(a)]) - \nu \mathbb{E}[C_1]}.$$

(ii) (Finite-Horizon) For any $T > 0$,

(6.62)

$$\lim_{u \to \infty} \frac{\psi(u, uz)}{\overline{B}_0(u)}$$

$$= \begin{cases} \frac{\nu \mathbb{E}[C_1]}{\rho(1-\mathbb{E}^q[H(a)])-\nu\mathbb{E}[C_1]} \left[1 - \left(1 + \left(\frac{\rho(1-\mathbb{E}^q[H(a)])-\nu\mathbb{E}[C_1]}{\rho(1-\mathbb{E}^q[H(a)])} \right) \frac{T}{\alpha} \right)^{-\alpha} \right] & \textit{if } \overline{B} \in \mathcal{R}(-\alpha - 1) \\ \frac{\nu \mathbb{E}[C_1]}{\rho(1-\mathbb{E}^q[H(a)])-\nu\mathbb{E}[C_1]} \left[1 - e^{-\frac{\rho(1-\mathbb{E}^q[H(a)])-\nu\mathbb{E}[C_1]}{\rho(1-\mathbb{E}^q[H(a)])}T} \right] & \textit{if } B \in \mathcal{G} \end{cases}.$$

6.5 Examples with Explicit Formulas

In this section, we discuss two examples where an explicit formula exists.

Example 1 is about the exponential asymptotics of the infinite-horizon ruin

probability when $H(a)$ and the claim size C are exponentially distributed. Example 2 gives an explicit expression for the rate function of the large deviation principle when $H(a)$ is exponentially distributed.

Example 1. *Recall that x is the minimal solution of*

$$(6.63) \qquad x = \int_{\mathbb{R}^+} \int_{\mathbb{X}} e^{\theta C + (x-1)H(a)} q(da)\mu(dC).$$

Now, assume that $H(a)$ is exponentially distributed with parameter $\lambda > 0$, then, we have

$$(6.64) \qquad x = \mathbb{E}^{\mu}[e^{\theta C}]\frac{\lambda}{\lambda - (x-1)},$$

which implies that

$$(6.65) \qquad x = \frac{1}{2}\left\{\lambda + 1 - \sqrt{(\lambda+1)^2 - 4\lambda\mathbb{E}^{\mu}[e^{\theta C}]}\right\}.$$

Now, assume that C is exponentially distributed with parameter $\gamma > 0$. Then,

$$(6.66) \qquad x = \frac{1}{2}\left\{\lambda + 1 - \sqrt{(\lambda+1)^2 - 4\lambda\frac{\gamma}{\gamma - \theta}}\right\}.$$

The infinite horizon probability satisfies $\lim_{u\to\infty} \frac{1}{u}\log\psi(u) = -\theta^{\dagger}$, where $\theta^{\dagger}$ satisfies

$$(6.67) \qquad \rho\theta^{\dagger} = \nu\left(\frac{1}{2}\left\{\lambda + 1 - \sqrt{(\lambda+1)^2 - 4\lambda\frac{\gamma}{\gamma - \theta^{\dagger}}}\right\} - 1\right),$$

192

which implies

$$(6.68) \qquad \frac{2\rho\theta^\dagger}{\nu} + 1 - \lambda = -\sqrt{(\lambda+1)^2 - \frac{4\lambda\gamma}{\gamma - \theta^\dagger}},$$

and thus

$$(6.69) \qquad \frac{\rho^2}{\nu^2}(\theta^\dagger)^2 + \frac{\rho\theta^\dagger}{\nu}(1-\lambda) = \lambda - \frac{\lambda\gamma}{\gamma - \theta^\dagger} = \frac{-\lambda\theta^\dagger}{\gamma - \theta^\dagger}.$$

Since we are looking for positive $\theta^\dagger$, we get the quadratic equation,

$$(6.70) \qquad \rho^2(\theta^\dagger)^2 - (\rho^2\gamma - \rho\nu(1-\lambda))\theta^\dagger - (\rho\nu\gamma(1-\lambda) + \lambda\nu^2) = 0.$$

Since $\rho > \frac{\mathbb{E}^\mu[C]\nu}{1 - \mathbb{E}^q[H(a)]} = \frac{\nu\lambda}{\gamma(\lambda-1)}$, we have $\rho\nu\gamma(1-\lambda) + \lambda\nu^2 > 0$. Therefore,

$$(6.71) \quad \theta^\dagger = \frac{(\rho^2\gamma - \rho\nu(1-\lambda)) + \sqrt{(\rho^2\gamma - \rho\nu(1-\lambda))^2 + 4\rho^2(\rho\nu\gamma(1-\lambda) + \lambda\nu^2)}}{2\rho^2}.$$

Example 2. *Now, let $H(a)$ be exponentially distributed with parameter $\lambda > 0$. We want an explicit expression for the rate function of the large deviation principle for $(N_t/t \in \cdot)$. Notice that,*

$$(6.72) \quad \Gamma(\theta) = \begin{cases} \nu\left(\frac{1}{2}\left\{\lambda + 1 - \sqrt{(\lambda+1)^2 - 4\lambda e^\theta}\right\} - 1\right) & \text{for } \theta \le \log\left(\frac{(\lambda+1)^2}{4\lambda}\right) \\ +\infty & \text{otherwise} \end{cases}.$$

To get $I(x) = \sup_{\theta \in \mathbb{R}}\{\theta x - \Gamma(\theta)\}$, we optimize over θ and consider $x = \Gamma'(\theta)$.

Evidently,

$$(6.73) \qquad x + \frac{1}{2}\nu(-4\lambda)e^{\theta}\frac{1}{2\sqrt{(\lambda+1)^2 - 4\lambda e^{\theta}}} = 0,$$

which gives us

$$(6.74) \qquad \theta = \log\left(\frac{-2x^2 + x\sqrt{4x^2 + \nu^2(\lambda+1)^2}}{\lambda\nu^2}\right),$$

whence,

(6.75)

$$I(x) = \begin{cases} x\log\left(\frac{-2x^2 + x\sqrt{4x^2 + \nu^2(\lambda+1)^2}}{\lambda\nu^2}\right) \\ \qquad\qquad -\nu\left(\frac{1}{2}\left\{\lambda + 1 - \frac{-2x + \sqrt{4x^2 + \nu^2(\lambda+1)^2}}{\nu}\right\} - 1\right) & \textit{if } x \geq 0 \\ +\infty & \textit{otherwise} \end{cases} \cdot$$

Bibliography

[1] Bacry, E., Dayri, K. and J. F. Muzy. (2012). Non-parametric kernel estimation for symmetric Hawkes processes. Applications to high frequency financial data. *Eur. Phys. J.* B **85**:157

[2] Bacry, E., Delattre, S., Hoffmann, M. and J. F. Muzy. Scaling limits for Hawkes processes and application to financial statistics. Preprint. arXiv:1202.0842.

[3] Bacry, E., Delattre, S., Hoffmann, M. and J. F. Muzy. Modeling microstructure noise with mutually exciting point processes. To appear in *Quantitative Finance*. arXiv:1101.3422.

[4] Bacry, E. and J. F. Muzy. Hawkes model for price and trades high-frequency dynamics. Preprint. arXiv:1301.1135.

[5] Bartlett, M. S. (1963). Statistical estimation of density functions. *Sankhyā* A. **25**, 245-254.

[6] Bartlett, M. S. (1963). The spectral analysis of point processes. *J. R. Statist. Soc.* B **25**, 264-296.

[7] Bauwens, L. and N. Hautsch. Modelling financial high frequency data using point processes. *Handbook of Financial Time Series*. 953-979, 2009.

[8] Billingsley, P. *Convergence of Probability Measures*, 2nd edition. Wiley-Interscience, New York, 1999.

[9] Bingham, N. H., Goldie, C. M. and J. L. Teugels, *Regular Variation*, Cambridge University Press, 1989

[10] Blundell, C., Heller, K. A. and J. M. Beck. Modelling reciprocating relationships with Hawkes processes. Preprint, 2012.

[11] Bordenave, C. and G. L. Torrisi. (2007). Large deviations of Poisson cluster processes. *Stochastic Models*. **23**, 593-625.

[12] Bormetti, G., Calcagnile, L. M., Treccani, M., Corsi, F., Marmi, S. and F. Lillo. Modelling systemic cojumps with Hawkes factor models. Preprint. arXiv:1301.6141.

[13] Bowsher, C. G. (2007). Modelling security market events in continuous time: intensity based, multivariate point process models. *Journal of Econometrics*. **141**, 876-912.

[14] Brémaud, P. and L. Massoulié. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.*. **24**, 1563-1588.

[15] Brémaud, P., Nappo, G. and G. L. Torrisi. (2002). Rate of convergence to equilibrium of marked Hawkes processes. *J. Appl. Prob.* **39**, 123-136.

[16] Brémaud, P. and L. Massoulié. (2001). Hawkes branching point processes without ancestors. *J. Appl. Prob.* **38**, 122-135.

[17] Brémaud, P. and S. Foss. (2010). Ergodicity of a stress release point process seismic model with aftershocks. *Markov Processes Relat. Fields.* **16**, 389-408.

[18] Brémaud, P. and L. Massoulié. (2002). Power spectra of general shot noises and Hawkes point processes with a random excitation. *Advances in Applied Probability.* **34**, 205-222.

[19] Brémaud, P., L. Massoulié and A. Ridolfi. (2005). Power spectra of random spike fields and related processes. *Advances in Applied Probability.* **37**, 1116-1146.

[20] Brix, A. and W. S. Kendall. (2002). Simulation of cluster point processes without edge effects. *Advances in Applied Probability.* **34**, 267-280.

[21] Carstensen, L., Sandelin, A., Winther, O. and N. R. Hansen. (2010). Multivariate Hawkes process models of the occurrence of regulatory elements. *BMC Bioinformatics.* **11**:456.

[22] Cartea, Á., Jaimungal, S. and J. Ricci. Buy low sell high: a high frequency trading perspective. *SSRN eLibrary*, 2011.

[23] Chavez-Demoulin, V., Davison, A. C. and A. J. McNeil. (2005). Estimating value-at-risk: a point process approach. *Quantitative Finance.* **5**, 227-234.

[24] Chavez-Demoulin, V. and J. A. McGill. (2012). High-frequency financial data modeling using Hawkes processes. *Journal of Banking & Finance.* **36**, 3415-3426.

[25] Chornoboy, E. S., Schramm, L. P. and A. F. Karr. (1988). Maximum likelihood identification of neural point process systems. *Biol. Cybern.* **59**, 265-275.

[26] Crane, R. and D. Sornette. (2008). Robust dynamic classes revealed by measuring the response function of a social system. Proc. Nat. Acad. Sci. USA **105**, 15649.

[27] Daley, D. J. and D. Vere-Jones. *An Introduction to the Theory of Point Processes.* Volume I and II, 2nd edition. Springer-Verlag, New York, 2003.

[28] Dassios, A. and H. Zhao. (2011). A dynamic contagion process. *Advances in Applied Probability.* **43**, 814-846.

[29] Dassios, A. and H. Zhao. (2012). Ruin by dynamic contagion claims. *Insurance: Mathematics and Economics.* **51**, 93-106.

[30] Dembo, A. and O. Zeitouni. *Large Deviations Techniques and Applications*, 2nd edition. Springer, New York, 1998.

[31] Donsker, M. D. and S. R. S. Varadhan. (1983). Asymptotic evaluation of certain Markov process expectations for large time. IV. *Communications of Pure and Applied Mathematics.* **36**, 183-212.

[32] Echeverría, P. (1982). A criterion for invariant measures of Markov processes. *Probability Theory and Related Fields.* **61**, 1-16.

[33] Egami, M., Kato, Y. and T. Sawaki. An analysis of CDS market liquidity by the Hawkes process. *SSRN eLibrary*, 2013.

[34] Egesdal, M., Fathauer, C., Louie, K. and J. Neuman. Statistical and stochastic modelling of gang rivalries in Los Angeles. *SIAM Undergraduate Research Online.* 2010.

[35] Embrechts, P., Liniger, T. and L. Lin. (2011). Multivariate Hawkes processes: an application to financial data. *J. Appl. Prob. Spec. Vol.* **48A**, 367-378.

[36] Errais, E., Giesecke, K. and L. Goldberg. (2010). Affine point processes and portfolio credit risk. *SIAM J. Financial Math.* **1**, 642-665.

[37] Fan, K. (1953). Minimax theorems. *Proc. Natl. Acad. Sci. USA.* **39**, 42-47.

[38] Feller, W., *An Introduction to Probability Theory and Its Applications*, Volume I and Volume II, 2nd edition, New York, 1971.

[39] Filimonov, V. and D. Sornette. (2012). Quantifying reflexivity in financial markets: Toward a prediction of flash crashes. *Physical Review E* **85** 056108

[40] Frenk, J. B. G. and G. Kassay. *The Level Set Method of Joó and Its Use in Minimax Theory.* Technical Report E.I 2003-03, Econometric Institute, Erasmus University, Rotterdam, 2003.

[41] Giesecke, K. and L. Goldberg, L. and X. Ding (2011). A top-down approach to multi-name credit. *Operations Research.* **59**, 283-300.

[42] Giesecke, K. and P. Tomecek. Dependent events and changes of time. Working paper, Cornell University, July 2005.

[43] Glynn, P. W. and W. Whitt. (1994). Logarithmic asymptotics for steady-state tail probabilities in a single-server queue. *J. Appl. Probab.* **31**, 131-156.

[44] Goldie, C. M. and S. Resnick. (1988). Distributions that are both subexponential and in the domain of attraction of an extreme value distribution. *Adv. Appl. Probab.* **20**, 706-718.

[45] Grandell, J. (1977) Point processes and random measures. *Advances in Applied Probability.* **9**, 502-526.

[46] Gusto, G. and S. Schbath. (2005). F.A.D.O.: a statistical method to detect favored or avoided distances between occurrences of motifs using the Hawkes model. *Stat. Appl. Genet. Mol. Biol.*, **4**, Article 24.

[47] Hairer, M. *Convergence of Markov Processes*, Lecture Notes, University of Warwick, available at `http://www.hairer.org/notes/Convergence.pdf`, 2010.

[48] Halpin, P. F. and P. De Boeck. Modelling dyadic interaction with Hawkes process. Preprint, 2012. To appear in *Psychometrika*.

[49] Hansen, N. R., Reynaud-Bouret, P. and V. Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. Preprint. arXiv:1208.0570.

[50] Hardiman, S. J., Bercot, N. and J-P. Bouchaud. Critical reflexivity in financial markets: a Hawkes process analysis. Preprint. arXiv:1302.1405.

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

[52] Hawkes, A. G. (1971). Point spectra of some mutually exciting point processes. *J. Roy. Statist. Soc. Ser. B* **33**, 438-443.

[53] Hawkes, A. G. and L. Adamopoulos. (1973). Cluster models for earthquakes-regional comparisons. *Bull. Int. Statist. Inst.* **45**, 454-461.

[54] Hawkes, A. G. and D. Oakes. (1974). A cluster process representation of a self-exciting process. *J. Appl. Prob.* **11**, 93-503.

[55] Hegemann, R. A., Lewis, E. A. and A. L. Bertozzi. (2013). An "Estimate & Score Algorithm" for simultaneous parameter estimation and reconstruction of incomplete data on social networks. *Security Informatics.* **2**:1.

[56] Hewlett, P. Clustering of order arrivals, price impact and trade path optimisation. Workshop on Financial Modeling with Jump processes, École Polytechnique, 2006.

[57] Heyde, C. C. and Scott, D. J. (1973). Invariance principle for the law of the iterated logarithm for martingales and processes with stationary increments. *Ann. Probab.* **1**, 428-436.

[58] Jagers, P. *Branching Processes with Biological Applications.* John Wiley, London, 1975.

[59] Johnson, D. H. (1996). Point process models of single-neuron discharges. *J. Computational Neuroscience.* **3**, 275-299.

[60] Joó, I. (1984). Note on my paper "A simple proof for von Neumann's minmax theorem". *Acta. Math. Hung.* **44**, 363-365.

[61] Kallenberg, O. *Foundations of Modern Probability*, Springer, 2nd edition, 2002.

[62] Karabash, D. and L. Zhu. Limit theorems for marked Hawkes processes with application to a risk model. Preprint. arXiv:1211.4039.

[63] Karabash, D. On stability of Hawkes process. Preprint. arXiv:1201.1573.

[64] Koralov, L. B. and Ya. G. Sinai. *Theory of Probability and Random Processes.* Springer, 2nd edition, 2012.

[65] Krumin, M., Reutsky I. and S. Shoham. (2010). Correlation-based analysis and generation of multiple spike trains using Hawkes models with an exogenous input. *Frontiers in Computational Neuroscience.* **4**, article 147.

[66] Kwieciński, A. and R. Szekli. (1996). Some monotonicity and dependence properties of self-exciting point processes. *Annals of Applied Probability.* **6**, 1211-1231.

[67] Large, J. (2007). Measuring the resiliency of an electronic limit order book. *Journal of Financial Markets.* **10**, 1-25.

[68] Lewis, E. and G. Mohler. A nonparametric EM algorithm for multiscale Hawkes processes. Preprint, 2011.

[69] Lewis, E., Mohler, Brantingham, P. J. and A. Bertozzi. (2011). Self-exciting point process of insergency in Iraq. *Security Journal.* **25**, 0955-1662.

[70] Lewis, P. A. W. and G. S. Shedler. (1979). Simulation of nonhomogeneous Poisson processes by thinning. *Naval Research Logistics Quarterly.* **26**, 403-413.

[71] Liniger, T. *Multivariate Hawkes Processes.* PhD thesis, ETH, 2009.

[72] Lipster, R. S. and A. N. Shiryaev. *Statistics of Random Processes II. Applications*, 2nd edition. Springer, 2001.

[73] Marsan, D. and O. Lengline. (2008). Extending earthquakes' reach through cascading. *Science.* **319** (5866), 1076.

[74] Massoulié, L. (1998). Stability results for a general class of interacting point processes dynamics, and applications. *Stochastic Processes and their Applications.* **75**, 1-30.

[75] Mehrdad, B., Sen, S. and L. Zhu. The speed of a biased walk on a Galton-Watson tree is monotonic with respect to progeny distributions for high values of bias. Preprint. arXiv:1212.3004.

[76] Meyer, S., Elias J. and M. Höhle. (2012). A space-time conditional intensity model for invasive meningococcal disease occurence. *Biometrics.* **68**, 607-616.

[77] Mitchell, L. and M. E. Cates. (2010). Hawkes process as a model of social interactions: a view on video dynamics. *Journal of Physics A: Mathematical and Theoretical.* **43**, 045101

[78] Mohler, G. O., Short, M. B., Brantingham, P. J., Schoenberg F. P. and G. E. Tita. (2011). Self-exciting point process modelling of crime. *Journal of the American Statistical Association.* **106**, 100-108.

[79] Møller, J. and J. G. Rasmussen. (2005). Perfect simulation of Hawkes processes. *Advances in Applied Probability.* **37**, 629-646.

[80] Møller, J. and J. G. Rasmussen. (2006). Approximate simulation of Hawkes processes. *Methodology and Computing in Applied Probability* 8, 53-64.

[81] Møller, J. and G. L. Torrisi. (2005). Second order analysis for spatial Hawkes processes. Technical Report R-2005-20, Department of Mathematical Sciences, Aalborg University.

[82] Møller, J. and G. L. Torrisi. (2007). The pair correlation function of spatial Hawkes processes. *Statistics & Probability Letters.* **77**, 995-1003.

[83] Muni Toke, I. and F. Pomponio. Modelling trades-through in a limited order book using Hawkes processes. *SSRN eLibrary*, 2011.

[84] Musmeci, F. and D. Vere-Jones. (1992). A space-time clustering model for historical earthquakes. *Annals of the Institute of Statistical Mathematics.* **44**, 1-11.

[85] Oakes, D. (1975). The Markovian self-exciting process. *J. Appl. Prob.* **12**, 69-77.

[86] Ogata, Y. (1978). The asymptotic behavior of maximum likelihood estimates for stationary point processes. *Ann. Inst. Statist. Math.* **30**, 243-261.

[87] Ogata, Y. (1988). Statistical models for earthquake occurrences and residual analysis for point processes. *J. Amer. Statist. Assoc.* **83**, 9-27.

[88] Ogata, Y. (1998). Space-time point-process models for earthquake occurrences. *Ann. Inst. Statist. Math.* **50**, 379-402.

[89] Ogata, Y. (1981). On Lewis' simulation method for point processes. *IEEE Transactions on Information Theory.* **27**, 23-31.

[90] Ogata, Y., Akaike, H. and K. Katsura. (1982). The application of linear intensity models to the investigation of causal relations between a point process

and another stochastic process. *Annals of the Institute of Statistical Mathematics.* **34**, 373-387.

[91] Ozaki, T. (1979). Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Statist. Math.* **31**, 145-155.

[92] Peng, X. and S. Kou. *Default Clustering and Valuation of Collateralized Debt Obligations.* Working Paper, Columbia University, January 2009.

[93] Pernice, V., Staude B., Carndanobile, S. and S. Rotter. (2012). How structure determines correlations in neuronal networks. *PLoS Computational Biology.* **85**:031916.

[94] Pernice, V., Staude B., Carndanobile, S. and S. Rotter. (2011). Recurrent interactions in spiking networks with arbitrary topology. *Physical Review E.* **7**:e1002059.

[95] Porter, M. and G. White. (2012). Self-exciting hurdle models for terrorist activity. *Annals of Applied Statistics* **6**, 106-124.

[96] Reynaud-Bouret, P. and S. Schbath. (2010). Adaptive estimation for Hawkes processes; application to genome analysis. *Ann. Statist.* **38**, 2781-2822.

[97] Reynaud-Bouret, P. and E. Roy. (2007). Some non asymptotic tail estimates for Hawkes processes. *Bull. Belg. Math. Soc. Simon Stevin* **13**, 883-896.

[98] Reynaud-Bouret, P., Tuleau-Malot, C., Rivoirard, V. and F. Grammont. Spike trains as (in)homogeneous Poisson processes or Hawkes processes: non-parametric adaptive estimation and goodness-of-fit tests. Preprint.

[99] Rubin, I. (1972). Regular point processes and their detection. *IEEE Transactions on Information Theory.* **18**, 547-557.

[100] Sen, S. and L. Zhu. Large deviations for self-correcting point processes. In Preparation.

[101] Sornette, D. and S. Utkin. (2009). Limits of declustering methods for disentangling exogenous from endogenous events in time series with foreshocks, main shocks, and aftershocks. *Physical Review E.* **79** (6), 61110.

[102] Stabile, G. and G. L. Torrisi. (2010). Risk processes with non-stationary Hawkes arrivals. *Methodol. Comput. Appl. Prob.* **12**, 415-429.

[103] Torrisi, G. L. (2002). A class of interacting marked point processes: Rate of convergence to equilibrium. *Journal of Applied Probability* **39**, 137-160.

[104] Wang, T., Bebbington, M. and D. Harte. (2012). Markov-modulated Hawkes process with stepwise decay. *Annals of the Institute of Statistical Mathematics* **64** 521-544.

[105] Wei, C. Z. and J. Winnicki. (1989). Some asymptotic results for the branching process with immigration. *Stoch. Proc. Appl.* **31**, 261-282.

[106] Varadhan, S. R. S. (2008). Large deviations. *Annals of Probability.* **36**, 397-419.

[107] Varadhan, S. R. S. *Large Deviations and Applications*, SIAM, Philadelphia, 1984.

[108] Varadhan, S. R. S. *Probability Theory*, Courant Lecture Notes, American Mathematical Society, 2007.

[109] Varadhan, S. R. S. *Stochastic Processes*, Courant Lecture Notes, American Mathematical Society, 2007.

[110] Vere-Jones, D. (1978). Earthquake prediction: A statistician's view. *Journal of Physics of the Earth* **26**, 129-146.

[111] Zheng, B., Roueff, F. and F. Abergel. Ergodicity and scaling limit of a constrained multivariate Hawkes process. *SSRN eLibrary*, 2013.

[112] Zhu, L. Large deviations for Markovian nonlinear Hawkes processes. Preprint. arXiv:1108.2432.

[113] Zhu, L. Process-level large deviations for nonlinear Hawkes point processes. To appear in *Annales de l'Institut Henri Poincaré*. arXiv:1108.2431.

[114] Zhu, L. Central limit theorem for nonlinear Hawkes processes. To appear in *Journal of Applied Probability*. arXiv:1204.1067.

[115] Zhu, L. (2013). Moderate deviations for Hawkes processes. *Statistics & Probability Letters* **83**, 885-890.

[116] Zhu, L. Limit theorems for a Cox-Ingersoll-Ross process with Hawkes jumps. Preprint.

[117] Zhu, L. Asymptotics for nonlinear Hawkes processes. In Preparation.

[118] Zhu, L. Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. Preprint. arXiv:1304.1940.