The Annals of Applied Probability
2015, Vol. 25, No. 2, 548–581
DOI: 10.1214/14-AAP1003

LARGE DEVIATIONS FOR MARKOVIAN NONLINEAR HAWKES PROCESSES

BY LINGJIONG ZHU[1]

New York University

Hawkes process is a class of simple point processes that is self-exciting and has clustering effect. The intensity of this point process depends on its entire past history. It has wide applications in finance, neuroscience and many other fields. In this paper, we study the large deviations for nonlinear Hawkes processes. The large deviations for linear Hawkes processes has been studied by Bordenave and Torrisi. In this paper, we prove first a large deviation principle for a special class of nonlinear Hawkes processes, that is, a Markovian Hawkes process with nonlinear rate and exponential exciting function, and then generalize it to get the result for sum of exponentials exciting functions. We then provide an alternative proof for the large deviation principle for a linear Hawkes process. Finally, we use an approximation approach to prove the large deviation principle for a special class of nonlinear Hawkes processes with general exciting functions.

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

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

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

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

$$\lambda_t := \lambda\left(\int_0^t h(t-s)N(ds)\right), \tag{1.2}$$

where $\lambda(\cdot):\mathbb{R}^+ \to \mathbb{R}^+$ is locally integrable and left continuous, $h(\cdot):\mathbb{R}^+ \to \mathbb{R}^+$, and we always assume that $\|h\|_{L^1} = \int_0^\infty h(t)\,dt < \infty$. The notation $\int_0^t h(t-s)N(ds)$ stands for $\int_{(0,t)} h(t-s)N(ds)$. Local integrability assumption of $\lambda(\cdot)$ ensures that the process is nonexplosive and left continuity assumption ensures that λ_t is $\mathcal{F}_t$-predictable.

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

Received June 2012; revised December 2013.

[1]Supported by NSF Grant DMS-09-04701, DARPA grant and MacCracken Fellowship at New York University.

MSC2010 subject classifications. 60G55, 60F10.

Key words and phrases. Large deviations, rare events, point processes, Hawkes processes, self-exciting processes.

Let $Z_t = \sum_{0<\tau_j<t} h(t-\tau_j)$, where τ_j is the jth arrival time of the process for $j \geq 1$. Thus we can write $\lambda_t = \lambda(Z_t)$.

This is known as the nonlinear Hawkes process; see Brémaud and Massoulié [3]. When the exciting function $h(\cdot)$ is exponential or a sum of exponentials, the process is Markovian, and we name it a Markovian nonlinear Hawkes process.

When $\lambda(\cdot)$ is linear, this is known as the (linear) Hawkes process, which was introduced in Hawkes [12]. If $\lambda(\cdot)$ is linear and $h(\cdot)$ is exponential or a sum of exponentials, the (linear) Markovian Hawkes process is sometimes referred to as Markovian self-exciting processes; see, for example, Oakes [20]. You can think of the arrival times τ_j as "bad" events, which can be the arrivals of claims in insurance literature or the time of defaults of big firms in the real world. Hawkes process captures both the self-exciting property and the clustering effect, which explains why it has wide applications in cosmology, ecology, epidemiology, seismology, neuroscience and DNA modeling. For a list of references to these applications, see Bordenave and Torrisi [2].

Hawkes process has also been applied in finance. Empirical comparisons suggest that Hawkes processes have some of the typical characteristics of a financial time series. Financial data have been analyzed using Hawkes processes. Self-exciting processes are used for the calculation of conditional risk measures, such as the value-at-risk. Another area of finance where Hawkes processes have been considered is credit default modeling. Hawkes processes have been proposed as models for the arrival of company defaults in a bond portfolio. For a list of references to the applications in finance, see Liniger [18] and Zhu [26].

For a short history of Hawkes process, we refer to Liniger [18]. For a survey on Hawkes processes and related self-exciting processes, Poisson cluster processes, marked point processes, etc., we refer to Daley and Vere-Jones [5].

When $\lambda(\cdot)$ is linear, say $\lambda(z) = \nu + z$, then one can use immigration-birth representation, also known as Galton–Watson theory to study it. Under the immigration-birth representation, if the immigrants are distributed as Poisson process with intensity ν and each immigrant generates a cluster whose number of points is denoted by S, then N_t is the total number of points generated in the clusters up to time t. If the process is ergodic, we have

$$\lim_{t\to\infty} \frac{N_t}{t} = \nu \mathbb{E}[S] \qquad \text{a.s.} \tag{1.3}$$

The central limit theorem for linear Hawkes processes was obtained in Bacry et al. [1], and it was proved for nonlinear Hawkes processes in Zhu [28]. The moderate deviations for linear Hawkes processes was obtained in Zhu [29].

Bordenave and Torrisi [2] proves that if $0 < \mu = \int_0^\infty h(t)\,dt < 1$ and $\int_0^\infty th(t)\,dt < \infty$, then $(\frac{N_t}{t} \in \cdot)$ satisfies the large deviation principle (LDP) with the good rate function $I(\cdot)$, that is, for any closed set $C \subset \mathbb{R}$,

$$\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{P}(N_t/t \in C) \leq -\inf_{x\in C} I(x), \tag{1.4}$$

and for any open set $G \subset \mathbb{R}$,

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{P}(N_t/t \in G) \geq -\inf_{x\in G} I(x), \tag{1.5}$$

where

$$I(x) = \begin{cases} x\theta_x + \nu - \dfrac{\nu x}{\nu + \mu x}, & \text{if } x \in [0,\infty), \\ +\infty, & \text{otherwise}, \end{cases} \tag{1.6}$$

where $\theta = \theta_x$ is the unique solution in $(-\infty, \mu - 1 - \log\mu]$ of $\mathbb{E}[e^{\theta S}] = \frac{x}{\nu + x\mu}$, $x > 0$. It is well known that (e.g., see page 39 of Jagers [14]), for all $\theta \in (-\infty, \mu - 1 - \log\mu]$, $\mathbb{E}[e^{\theta S}]$ satisfies

$$\mathbb{E}[e^{\theta S}] = e^{\theta} \exp\{\mu(\mathbb{E}[e^{\theta S}] - 1)\}. \tag{1.7}$$

See Dembo and Zeitouni [7] for general background regarding large deviations and the applications. Also Varadhan [23] has an excellent survey article on this subject.

In a recent paper, Zhu [24] studied the limit theorems for a Cox–Ingersoll–Ross process with Hawkes jumps, an extension of the linear Hawkes processes. Karabash and Zhu [16] obtained to the limit theorems for linear marked Hawkes processes, another extension of the classical Hawkes processes.

The large deviations result for $(N_t/t \in \cdot)$ is helpful to study the ruin probabilities of a risk process when the claims arrivals follow a Hawkes process. Stabile and Torrisi [21] considered risk processes with nonstationary Hawkes claims arrivals and studied the asymptotic behavior of infinite and finite horizon ruin probabilities under light-tailed conditions on the claims. The corresponding result for heavy-tailed claims was obtained by Zhu [27].

In this paper, we are interested in Hawkes processes with general nonlinear $\lambda(\cdot)$. If $\lambda(\cdot)$ is nonlinear, then the usual Galton–Watson theory approach no longer works. If the exciting function h is exponential or a sum of exponentials, the process is Markovian, and there exists a generator of the process. The difficulty arises when h is not exponential or a sum of exponentials in which case the process is non-Markovian. Another possible generalization is to consider h to be random. Then, we will get a marked point process. For a discussion on marked point processes, see Cox and Isham [4].

When $\lambda(\cdot)$ is nonlinear, Brémaud and Massoulié [3] proves that under certain conditions, there exists a unique stationary version of the nonlinear Hawkes process and Brémaud and Massoulié [3] also proves the convergence to equilibrium of a nonstationary version, both in distribution and in variation.

In this paper, we will prove the large deviation when h is exponential, and λ is nonlinear first. Then, we will generalize the proof to the case when h is a sum of exponentials. We will use that to recover the result proved in Bordenave and Torrisi [2]. Finally, we will prove the result for a special class of nonlinear λ and general h.

2. 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 $Z_i(t) := \sum_{\tau_j<t} a_i e^{-b_i(t-\tau_j)}$, $1 \leq i \leq d$, where $b_i > 0$, $a_i \neq 0$ (might be negative), and τ_j's are the arrivals of the simple point process with intensity $\lambda(Z_1(t), \dots, Z_d(t))$ at time t, where $\lambda : \mathcal{Z} \to \mathbb{R}^+$ and $\mathcal{Z} := \mathbb{R}^{\varepsilon_1} \times \cdots \times \mathbb{R}^{\varepsilon_d}$ is the domain for $(Z_1(t), \dots, Z_d(t))$, where $\mathbb{R}^{\varepsilon_i} := \mathbb{R}^+$ or $\mathbb{R}^-$ depending on whether $\varepsilon_i = +1$ or -1, where $\varepsilon_i = +1$ if $a_i > 0$ and $\varepsilon_i = -1$ 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 $(Z_1(t), \dots, Z_d(t))$ is given by

$$\begin{aligned} \mathcal{A}f = &-\sum_{i=1}^d b_i z_i \frac{\partial f}{\partial z_i} \\ &+ \lambda(z_1, \dots, z_d)\big[f(z_1 + a_1, \dots, z_d + a_d) - f(z_1, \dots, z_d)\big]. \end{aligned} \tag{2.1}$$

For a reference to generators for Markov processes with jumps, see Davis [6].

We want to prove the existence and uniqueness of the invariant probability measure for $(Z_1(t), \dots, Z_d(t))$. Here the invariance is in time.

LEMMA 1. *Consider* $h(t) = \sum_{i=1}^d a_i e^{-b_i t} > 0$. *Assume* $\lambda(z_1, \dots, z_n) \leq \sum_{i=1}^d \alpha_i |z_i| + \beta$, *where* $\beta > 0$ *and* $\alpha_i > 0$, $1 \leq i \leq d$, *satisfies* $\sum_{i=1}^d \frac{|a_i|}{b_i} \alpha_i < 1$. *Then, there exists a unique invariant probability measure for* $(Z_1(t), \dots, Z_d(t))$.

PROOF. The lecture notes [11] 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 $u(z_1, \dots, z_d) = \sum_{i=1}^d \varepsilon_i c_i z_i \geq 0$, where $c_i > 0$, $1 \leq i \leq d$. Then

$$\begin{aligned} \mathcal{A}u &= -\sum_{i=1}^d b_i \varepsilon_i c_i z_i + \lambda(z_1, \dots, z_d) \sum_{i=1}^d a_i \varepsilon_i c_i \\ &\leq -\sum_{i=1}^d b_i c_i |z_i| + \sum_{i=1}^d \alpha_i |z_i| \sum_{i=1}^d |a_i| c_i + \beta \sum_{i=1}^d |a_i| c_i. \end{aligned} \tag{2.2}$$

Taking $c_i = \frac{\alpha_i}{b_i} > 0$, we get

$$\begin{aligned} \mathcal{A}u &\leq -\left(1 - \sum_{i=1}^d \frac{|a_i|\alpha_i}{b_i}\right) \sum_{i=1}^d \alpha_i |z_i| + \beta \sum_{i=1}^d \frac{|a_i|\alpha_i}{b_i} \\ &\leq -\min_{1\leq i\leq d} b_i \cdot \left(1 - \sum_{i=1}^d \frac{|a_i|\alpha_i}{b_i}\right) u + \beta \sum_{i=1}^d \frac{|a_i|\alpha_i}{b_i}. \end{aligned} \tag{2.3}$$

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 $Z_0 = x$, that is, $Z_T^x = xe^{-bT} + \sum_{\tau_j < T} ae^{-b(T-\tau_j)}$. 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 [22]. 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

$$\big((a+x)e^{-bT}, xe^{-bT} + a\big) \quad \text{and} \quad \big((a+y)e^{-bT}, ye^{-bT} + a\big), \tag{2.4}$$

respectively. Choosing $T > \frac{1}{b}\log(\frac{x-y+a}{a})$, we have

$$\big((a+x)e^{-bT}, xe^{-bT} + a\big) \cap \big((a+y)e^{-bT}, ye^{-bT} + a\big) \neq \varnothing, \tag{2.5}$$

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 $Z_t^x := (Z_t^{x_1}, \dots, Z_t^{x_d}) \in \mathcal{Z}_d$ and $Z_t^y := (Z_t^{y_1}, \dots, Z_t^{y_d}) \in \mathcal{Z}_d$, where $Z_t^{x_i} = x_i e^{-b_i t} + \sum_{\tau_j < t} a_i e^{-b_i(t-\tau_j)}$, $1 \le i \le d$. 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. $\square$

3. Large deviations for Markovian nonlinear Hawkes processes with exponential exciting function.

We assume first that $h(t) = ae^{-bt}$, where $a, b > 0$, that is, the process Z_t jumps upward 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

$$dZ_t = -bZ_t\,dt + a\,dN_t, \tag{3.1}$$

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 2. *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*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}\big[e^{\theta N_t}\big] = \Gamma(\theta), \tag{3.2}$$

where

$$\Gamma(\theta)=\sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e}\left\{\int\frac{\theta b}{a}z\hat{\pi}(dz)+\int(\hat{\lambda}-\lambda)\hat{\pi}(dz)-\int\big(\log(\hat{\lambda}/\lambda)\big)\hat{\lambda}\hat{\pi}(dz)\right\}, \tag{3.3}$$

where $\mathcal{Q}_e$ *is defined as*

$$\mathcal{Q}_e=\{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}:\widehat{\mathcal{A}}\text{ has unique invariant probability measure }\hat{\pi}\}, \tag{3.4}$$

where

$$\mathcal{Q}=\left\{(\hat{\lambda},\hat{\pi}):\hat{\pi}\in\mathcal{M}(\mathbb{R}^+),\int z\hat{\pi}(dz)<\infty,\hat{\lambda}\in L^1(\hat{\pi}),\hat{\lambda}>0\right\}, \tag{3.5}$$

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* $\widehat{\mathcal{A}}$ *as*

$$\widehat{\mathcal{A}}f(z)=-bz\frac{\partial f}{\partial z}+\hat{\lambda}(z)[f(z+a)-f(z)], \tag{3.6}$$

for any $f:\mathbb{R}^+\to\mathbb{R}$ *that is* C^1, *that is, continuously differentiable.*

PROOF. By Lemma 3, $\mathbb{E}[e^{\theta N_t}]<\infty$ for any $\theta\in\mathbb{R}$, also

$$\mathbb{E}[e^{\theta N_t}]=\mathbb{E}[e^{(\theta/a)(Z_t+b\int_0^t Z_s\,ds)}]. \tag{3.7}$$

Define the set

$$\mathcal{U}_\theta=\{u\in C^1(\mathbb{R}^+,\mathbb{R}^+):u(z)=e^{f(z)},\ \text{where } f\in\mathcal{F}\}, \tag{3.8}$$

where

$$\mathcal{F}=\Big\{f:f(z)=Kz+g(z)+L,\quad K>\frac{\theta}{a},K,L\in\mathbb{R},g\text{ is }C_1\text{ with compact support}\Big\}. \tag{3.9}$$

Now for any $u\in\mathcal{U}_\theta$, define

$$M:=\sup_{z\geq 0}\frac{\mathcal{A}u(z)+((\theta b)/a)zu(z)}{u(z)}. \tag{3.10}$$

By Dynkin's formula if $M<\infty$, for $V(z):=\frac{\theta b}{a}z$, we have

$$\begin{aligned}&\mathbb{E}[u(Z_t)e^{\int_0^t V(Z_s)\,ds}]\\&\quad=u(Z_0)+\int_0^t\mathbb{E}[(\mathcal{A}u(Z_s)+V(Z_s)u(Z_s))e^{\int_0^s V(Z_v)\,dv}]\,ds\\&\quad\leq u(Z_0)+M\int_0^t\mathbb{E}[u(Z_s)e^{\int_0^s V(Z_v)\,dv}]\,ds,\end{aligned} \tag{3.11}$$

which implies by Gronwall's lemma that

$$\mathbb{E}\big[u(Z_t)e^{\int_0^t V(Z_s)\,ds}\big] \le u(Z_0)e^{Mt} = u(0)e^{Mt}. \tag{3.12}$$

Observe that by the definition of $\mathcal{U}_\theta$, for any $u \in \mathcal{U}_\theta$, we have $u(z) \ge c_1 e^{(\theta/a)z}$ for some constant $c_1 > 0$ and therefore by (3.7) and (3.12),

$$\mathbb{E}\big[e^{\theta N_t}\big] \le \frac{1}{c_1}\mathbb{E}\big[u(Z_t)e^{\int_0^t((\theta b)/a)Z_s\,ds}\big] \le \frac{1}{c_1}u(0)e^{Mt}. \tag{3.13}$$

Hence

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}\big[e^{\theta N_t}\big] \le M = \sup_{z\ge 0}\frac{\mathcal{A}u(z) + ((\theta b)/a)zu(z)}{u(z)}, \tag{3.14}$$

which is still true even if $M = \infty$. Since this holds for any $u \in \mathcal{U}_\theta$,

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}\big[e^{\theta N_t}\big] \le \inf_{u\in\mathcal{U}_\theta}\sup_{z\ge 0}\frac{\mathcal{A}u(z) + ((\theta b)/a)zu(z)}{u(z)}. \tag{3.15}$$

Define the tilted probability measure $\widehat{\mathbb{P}}$ by

$$\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = \exp\left\{\int_0^t\big(\lambda(Z_s) - \hat{\lambda}(Z_s)\big)\,ds + \int_0^t \log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right)dN_s\right\}. \tag{3.16}$$

Notice that $\widehat{\mathbb{P}}$ defined in (3.16) 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 [19].)

Now by Jensen's inequality

$$\begin{aligned}
&\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{E}\big[e^{\theta N_t}\big] \\
&\quad = \liminf_{t\to\infty}\frac{1}{t}\log\widehat{\mathbb{E}}\left[\exp\left\{\theta N_t - \log\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t}\right\}\right] \\
&\quad \ge \liminf_{t\to\infty}\widehat{\mathbb{E}}\left[\frac{1}{t}\theta N_t - \frac{1}{t}\log\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t}\right] \\
&\quad = \liminf_{t\to\infty}\widehat{\mathbb{E}}\left[\frac{1}{t}\theta N_t - \frac{1}{t}\int_0^t\big(\lambda(Z_s) - \hat{\lambda}(Z_s)\big)\,ds - \int_0^t\log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right)dN_s\right].
\end{aligned} \tag{3.17}$$

Since $N_t - \int_0^t \hat{\lambda}(Z_s)\,ds$ is a martingale under $\widehat{\mathbb{P}}$, we have

$$\widehat{\mathbb{E}}\left[\int_0^t\log\left(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\right)\big(dN_s - \hat{\lambda}(Z_s)\,ds\big)\right] = 0. \tag{3.18}$$

Therefore, by the ergodic theorem, (for a reference, see Chapter 16.4 of Koralov and Sinai [17]), for any $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$,

$$
\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \\
&\quad \geq \liminf_{t\to\infty} \widehat{\mathbb{E}}\biggl[\frac{1}{t}\theta N_t - \frac{1}{t}\int_0^t \bigl(\lambda(Z_s) - \hat{\lambda}(Z_s)\bigr)\,ds \\
&\qquad\qquad - \int_0^t \log\biggl(\frac{\hat{\lambda}(Z_s)}{\lambda(Z_s)}\biggr)\hat{\lambda}(Z_s)\,ds\biggr] \\
&\quad = \int \frac{\theta b}{a} z\hat{\pi}(dz) + \int (\hat{\lambda} - \lambda)\hat{\pi}(dz) - \int \bigl(\log(\hat{\lambda}) - \log(\lambda)\bigr)\hat{\lambda}\hat{\pi}(dz).
\end{aligned}
\tag{3.19}
$$

Hence

$$
\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \\
&\quad \geq \sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e} \biggl\{\int \frac{\theta b}{a} z\hat{\pi} + \int (\hat{\lambda} - \lambda)\hat{\pi} - \int \bigl(\log(\hat{\lambda}) - \log(\lambda)\bigr)\hat{\lambda}\hat{\pi}\biggr\}.
\end{aligned}
\tag{3.20}
$$

Recall that

$$
\begin{aligned}
\mathcal{F} = \biggl\{ f : f(z) = Kz + g(z) + L, K > \frac{\theta}{a}, \\
K, L \in \mathbb{R}, g \text{ is } C_1 \text{ with compact support}\biggr\}.
\end{aligned}
\tag{3.21}
$$

We claim that

$$
\inf_{f\in\mathcal{F}} \biggl\{\int \widehat{\mathcal{A}} f(z)\hat{\pi}(dz)\biggr\} = \begin{cases} 0, & \text{if } (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e, \\ -\infty, & \text{if } (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} \setminus \mathcal{Q}_e. \end{cases}
\tag{3.22}
$$

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, $\int \widehat{\mathcal{A}} z\hat{\pi} = 0$. 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 $\int \widehat{\mathcal{A}} g\hat{\pi} = 0$, then $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$. We can easily check the conditions in Echevrría [8]. (E.g., $\mathcal{G}$ is dense in $C(\mathbb{R}^+)$, the set of continuous and bounded functions on $\mathbb{R}^+$ with limit that exists at infinity and $\widehat{\mathcal{A}}$ satisfies the minimum principle, that is, $\widehat{\mathcal{A}} f(z_0) \geq 0$ 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 [8] can also be easily verified.) Thus, Echevrría [8] implies that $\hat{\pi}$ is an invariant measure. Now, our proof in Lemma 1 shows that $\hat{\pi}$ has to be unique as well. Therefore,

$(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e$. This implies that if $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} \setminus \mathcal{Q}_e$, there exists some $g \in \mathcal{G}$, such that $\int \widehat{\mathcal{A}} g \hat{\pi} \neq 0$. Now any constant multiplier of g still belongs to $\mathcal{G}$ and thus $\inf_{g \in \mathcal{G}} \int \widehat{\mathcal{A}} g \hat{\pi} = -\infty$ and hence $\inf_{f \in \mathcal{F}} \int \widehat{\mathcal{A}} f \hat{\pi} = -\infty$ if $(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q} \setminus \mathcal{Q}_e$.

Therefore,

$$\liminf_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \geq \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}} \inf_{f \in \mathcal{F}} \left\{ \int \frac{\theta b}{a} z \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) + \int \widehat{\mathcal{A}} f \hat{\pi} \right\} \tag{3.23}$$

$$\geq \sup_{(\hat{\lambda}\hat{\pi}, \hat{\pi}) \in \mathcal{R}} \inf_{f \in \mathcal{F}} \left\{ \int \frac{\theta b}{a} z \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) + \int \widehat{\mathcal{A}} f \hat{\pi} \right\}, \tag{3.24}$$

where $\mathcal{R} = \{(\hat{\lambda}\hat{\pi}, \hat{\pi}) : (\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}\}$ and

$$\widehat{H}(\hat{\lambda}, \hat{\pi}) = \int [(\lambda - \hat{\lambda}) + \log(\hat{\lambda}/\lambda)\hat{\lambda}] \hat{\pi}. \tag{3.25}$$

Define

$$\begin{aligned} F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) &= \int \frac{\theta b}{a} z \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) + \int \widehat{\mathcal{A}} f \hat{\pi} \\ &= \int \frac{\theta b}{a} z \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) - \int bz \frac{\partial f}{\partial z} \hat{\pi} \\ &\quad + \int \big(f(z+a) - f(z)\big) \hat{\lambda} \hat{\pi}. \end{aligned} \tag{3.26}$$

Notice that F is linear in f and hence convex in f and also

$$\widehat{H}(\hat{\lambda}, \hat{\pi}) = \sup_{f \in C_b(\mathbb{R}^+)} \left\{ \int [\hat{\lambda} f + \lambda(1 - e^f)] \hat{\pi} \right\}, \tag{3.27}$$

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 $\widehat{H}$ 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 $f = Kz + g + L \in \mathcal{F}$,

$$\begin{aligned} F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) &= \int \left(\frac{\theta}{a} - K\right) bz \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) - \int bz \frac{\partial g}{\partial z} \hat{\pi} \\ &\quad + \int \big(g(z+a) - g(z)\big) \hat{\lambda} \hat{\pi} + Ka \int \hat{\lambda} \hat{\pi}. \end{aligned} \tag{3.28}$$

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

$$\begin{aligned} &- \int bz \frac{\partial g}{\partial z} \pi_n + \int \big(g(z+a) - g(z)\big) \lambda_n \pi_n + Ka \int \lambda_n \pi_n \\ &\quad \to - \int bz \frac{\partial g}{\partial z} \pi_\infty + \int \big(g(z+a) - g(z)\big) \gamma_\infty + Ka \int \gamma_\infty, \end{aligned} \tag{3.29}$$

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 $(\frac{\theta}{a} - K)bz$ is continuous and nonpositive on $\mathbb{R}^+$, we have

$$\limsup_{n\to\infty} \int \left(\frac{\theta}{a} - K\right) bz\pi_n \leq \int \left(\frac{\theta}{a} - K\right) bz\pi_\infty. \tag{3.30}$$

Hence, we conclude that F is upper semicontinuous in the weak topology.

In order to switch the supremum and infimum in (3.24), 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 [9]. Indeed, Joó developed some level set method and proved that it is sufficient to show the compactness of the level set; see Joó [15] and Frenk and Kassay [10]. In other words, it suffices to prove that, for any $C \in \mathbb{R}$ and $f \in \mathcal{F}$, the level set

$$\left\{(\hat{\lambda}\hat{\pi}, \hat{\pi}) \in \mathcal{R} : \widehat{H} + \int bz\frac{\partial f}{\partial z}\hat{\pi} - \frac{\theta b}{a}z\hat{\pi} - \hat{\lambda}[f(z+a) - f(z)]\hat{\pi} \leq C\right\} \tag{3.31}$$

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 (3.31), there exists some $C_1, C_2 > 0$ such that

$$\begin{aligned}
C_1 &\geq \widehat{H} + \left(K - \frac{\theta}{a}\right) b \int z\hat{\pi} - C_2 \int \hat{\lambda}\hat{\pi} \\
&\geq \int_{\hat{\lambda} \geq cz+\ell} [\lambda - \hat{\lambda} + \hat{\lambda}\log(\hat{\lambda}/\lambda)]\hat{\pi} + \left(K - \frac{\theta}{a}\right) b \int z\hat{\pi} \\
&\quad - C_2 \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} - C_2 \int_{\hat{\lambda} < cz+\ell} \hat{\lambda}\hat{\pi} \\
&\geq \left[\min_{z\geq 0} \log \frac{cz+\ell}{\lambda(z)} - 1 - C_2\right] \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} \\
&\quad + \left[-c \cdot C_2 + \left(K - \frac{\theta}{a}\right) b\right] \int z\hat{\pi} - \ell C_2.
\end{aligned} \tag{3.32}$$

We choose $0 < c < (K - \frac{\theta}{a})\frac{b}{C_2}$ 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,

$$\int z\hat{\pi} \leq C_3, \qquad \int_{\hat{\lambda} \geq cz+\ell} \hat{\lambda}\hat{\pi} \leq C_4, \tag{3.33}$$

where

$$\begin{aligned}
C_3 &= \frac{C_1 + \ell C_2}{-c \cdot C_2 + (K - (\theta/a))b}, \\
C_4 &= \frac{C_1 + \ell C_2}{\min_{z\geq 0} \log((cz+\ell)/\lambda(z)) - 1 - C_2}.
\end{aligned} \tag{3.34}$$

Therefore, we have

$$\int \hat{\lambda}\hat{\pi} = \int_{\hat{\lambda}\geq cz+\ell} \hat{\lambda}\hat{\pi} + \int_{\hat{\lambda}<cz+\ell} \hat{\lambda}\hat{\pi} \leq C_4 + c \cdot C_3 + \ell, \tag{3.35}$$

and hence

$$\widehat{H}(\hat{\lambda}, \hat{\pi}) \leq C_1 + C_2[C_4 + c \cdot C_3 + \ell] < \infty. \tag{3.36}$$

Therefore, for any $(\lambda_n \pi_n, \pi_n) \in \mathcal{R}$, we get

$$\lim_{\ell\to\infty} \sup_n \int_{z\geq\ell} \pi_n \leq \lim_{\ell\to\infty} \sup_n \frac{1}{\ell} \int z\pi_n \leq \lim_{\ell\to\infty} \frac{C_3}{\ell} = 0, \tag{3.37}$$

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) $\lim_{\ell\to\infty} \sup_n \int_{z\geq\ell} \lambda_n\pi_n = 0$.

(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

$$\sup_n \int \lambda_n\pi_n = \sup_n \int \frac{b}{a} z\pi_n \leq \frac{b}{a}[C_4 + c \cdot C_3 + \ell] < \infty. \tag{3.38}$$

To prove (ii), notice that $(\lambda - \lambda_n) + \lambda_n \log(\lambda_n/\lambda) \geq 0$. That is because $x - 1 - \log x \geq 0$ for any $x > 0$ and hence

$$\lambda - \hat{\lambda} + \hat{\lambda}\log(\hat{\lambda}/\lambda) = \hat{\lambda}\big[(\lambda/\hat{\lambda}) - 1 - \log(\lambda/\hat{\lambda})\big] \geq 0. \tag{3.39}$$

Notice that

$$\begin{aligned} &\lim_{\ell\to\infty} \sup_n \int_{z\geq\ell} \lambda_n\pi_n \\ &\quad\leq \lim_{\ell\to\infty} \sup_n \int_{\lambda_n<\sqrt{\lambda z}, z\geq\ell} \lambda_n\pi_n + \lim_{\ell\to\infty} \sup_n \int_{\lambda_n\geq\sqrt{\lambda z}, z\geq\ell} \lambda_n\pi_n. \end{aligned} \tag{3.40}$$

For the first term, since $\sup_n \int z\pi_n < \infty$ and $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$,

$$\lim_{\ell\to\infty} \sup_n \int_{\lambda_n<\sqrt{\lambda z}, z\geq\ell} \lambda_n\pi_n \leq \lim_{\ell\to\infty} \sup_n \int_{z\geq\ell} \sqrt{\lambda z}\,\pi_n = 0. \tag{3.41}$$

For the second term, since $\limsup_{z\to\infty} \frac{\lambda(z)}{z} = 0$,

$$\begin{aligned} &\lim_{\ell\to\infty} \sup_n \int_{\lambda_n\geq\sqrt{\lambda z}, z\geq\ell} \lambda_n\pi_n \\ &\quad\leq \lim_{\ell\to\infty} \sup_n \widehat{H}(\lambda_n, \pi_n) \sup_{\lambda_n\geq\sqrt{\lambda z}, z\geq\ell} \frac{\lambda_n}{\lambda - \lambda_n + \lambda_n\log(\lambda_n/\lambda)} = 0. \end{aligned} \tag{3.42}$$

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 the supremum and infimum in (3.24) and get

$$\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \tag{3.43}$$

$$\geq \inf_{f\in\mathcal{F}} \sup_{\hat{\pi} : \int z\hat{\pi}<\infty} \sup_{\hat{\lambda}\in L^1(\hat{\pi})} \left\{ \int \frac{\theta b}{a} z\hat{\pi} + (\hat{\lambda} - \lambda)\hat{\pi} - \log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi} + \widehat{\mathcal{A}} f\hat{\pi} \right\} \tag{3.44}$$

$$= \inf_{f\in\mathcal{F}} \sup_{\hat{\pi} : \int z\hat{\pi}<\infty} \int \left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} \right] \hat{\pi}(dz) \tag{3.45}$$

$$= \inf_{f\in\mathcal{F}} \sup_{z\geq 0} \left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} \right] \tag{3.46}$$

$$= \inf_{f\in\mathcal{F}} \sup_{z\geq 0} \left[\frac{\theta bze^{f(z)}}{ae^{f(z)}} + \frac{\lambda(z)}{e^{f(z)}}\big(e^{f(z+a)} - e^{f(z)}\big) - \frac{bz}{e^{f(z)}} \frac{\partial e^{f(z)}}{\partial z} \right] \tag{3.47}$$

$$\geq \inf_{u\in\mathcal{U}_\theta} \sup_{z\geq 0} \left\{ \frac{\mathcal{A}u}{u} + \frac{\theta b}{a} z \right\}. \tag{3.48}$$

We need some justifications. Define $G(\hat{\lambda}) = \hat{\lambda} - \log(\hat{\lambda}/\lambda)\hat{\lambda} + \widehat{\mathcal{A}} f$. 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 (3.45). Next, let us explain (3.46). For any probability measure $\hat{\pi}$,

$$\int \left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} \right] \hat{\pi}(dz) \leq \sup_{z\geq 0} \left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} \right], \tag{3.49}$$

which implies the right-hand side of (3.45) is less or equal to the right-hand side of (3.46). To prove the other direction. For any $f = Kz + g + L \in \mathcal{F}$, we have

$$\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} = \left(\frac{\theta b}{a} - Kb \right) z + \lambda(z)\big(e^{Ka+g(z+a)-g(z)} - 1\big) - bz\frac{\partial g}{\partial z}, \tag{3.50}$$

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

$$\begin{aligned}\frac{\theta bz}{a} &+ \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} \\ &= \frac{\theta bz^*}{a} + \lambda(z^*)\big(e^{f(z^*+a)-f(z^*)} - 1\big) - bz^*\frac{\partial f}{\partial z}\bigg|_{z=z^*}.\end{aligned} \tag{3.51}$$

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

$$\begin{aligned}\lim_{n\to\infty} \int &\left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z}\right]\hat{\pi}_n(dz) \\ &= \lim_{n\to\infty} n\int_{z^*-(1/(2n))}^{z^*+(1/(2n))} \left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z}\right] dz \\ &= \frac{\theta bz^*}{a} + \lambda(z^*)\big(e^{f(z^*+a)-f(z^*)} - 1\big) - bz^*\frac{\partial f}{\partial z}\bigg|_{z=z^*} \\ &= \sup_{z\geq 0}\left[\frac{\theta bz}{a} + \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z}\right].\end{aligned} \tag{3.52}$$

We conclude that the right-hand side of (3.45) is greater or equal to the right-hand side of (3.46).

Notice that for any $f = Kz + g + L \in \mathcal{F}$,

$$\begin{aligned}\frac{\theta bz}{a} &+ \lambda(z)\big(e^{f(z+a)-f(z)} - 1\big) - bz\frac{\partial f}{\partial z} \\ &= \frac{b(\theta - Ka)}{a}z + \lambda(z)\big(e^{Ka+g(z+a)-g(z)} - 1\big) - bz\frac{\partial g}{\partial z},\end{aligned} \tag{3.53}$$

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 (3.46). Finally, for any $f \in \mathcal{F}$, $u = e^f \in \mathcal{U}_\theta$, which implies (3.48). □

LEMMA 3. *Assume* $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$, *and we have* $\mathbb{E}[e^{\theta N_t}] < \infty$ *for any* $\theta \in \mathbb{R}$.

PROOF. Observe that for any $\gamma \in \mathbb{R}$,

$$\exp\left\{\gamma N_t - \int_0^t (e^\gamma - 1)\lambda(Z_s)\,ds\right\} \tag{3.54}$$

is a martinagle. Since $\lim_{z\to\infty}\frac{\lambda(z)}{z}=0$, for any $\varepsilon>0$, there exists a constant $C_\varepsilon>0$ such that $\lambda(z)\leq C_\varepsilon+\varepsilon z$ for any $z\geq 0$. Also,

$$\begin{aligned}\int_0^t Z_s\,ds&=\int_0^t\int_0^s h(s-u)N(du)\,ds=\int_0^t\left[\int_u^t h(s-u)\,ds\right]N(du)\\&\leq\int_0^t\left[\int_u^\infty h(s-u)\,ds\right]N(du)=\|h\|_{L^1}N_t.\end{aligned}\tag{3.55}$$

Therefore, for any $\gamma>0$,

$$\begin{aligned}1&=\mathbb{E}\big[e^{\gamma N_t-\int_0^t(e^\gamma-1)\lambda(Z_s)\,ds}\big]\\&\geq\mathbb{E}\big[e^{\gamma N_t-(e^\gamma-1)\int_0^t(C_\varepsilon+\varepsilon Z_s)\,ds}\big]\\&\geq\mathbb{E}\big[e^{\gamma N_t-(e^\gamma-1)C_\varepsilon t-(e^\gamma-1)\varepsilon\|h\|_{L^1}N_t}\big].\end{aligned}\tag{3.56}$$

For any $\theta>0$, choose $\gamma>\theta$ and ε small enough so that $\gamma-(e^\gamma-1)\varepsilon\|h\|_{L^1}\geq\theta$. Then

$$\mathbb{E}\big[e^{\theta N_t}\big]\leq e^{(e^\gamma-1)C_\varepsilon t}<\infty.\tag{3.57}$$ □

Now we are ready to prove the large deviations result.

THEOREM 4. *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* $\Gamma(\cdot)$,

$$I(x)=\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma(\theta)\}.\tag{3.58}$$

PROOF. If $\limsup_{z\to\infty}\frac{\lambda(z)}{z}=0$, then the forthcoming Lemma 6 implies that $\Gamma(\theta)<\infty$ for any θ. Thus, by the Gärtner–Ellis theorem, we have the upper bound. For the Gärtner–Ellis theorem and a general theory of large deviations, see, for example, [7]. To prove the lower bound, it suffices to show that for any $x>0$, $\varepsilon>0$, we have

$$\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t}\in B_\varepsilon(x)\right)\geq-\sup_\theta\{\theta x-\Gamma(\theta)\},\tag{3.59}$$

where $B_\varepsilon(x)$ denotes the open ball centered at x with radius ε. Let $\widehat{\mathbb{P}}$ denote the tilted probability measure with rate $\hat{\lambda}$ defined in Theorem 2. By Jensen's inequality,

$$\begin{aligned}&\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t}\in B_\varepsilon(x)\right)\\&\quad=\frac{1}{t}\log\int_{(N_t/t)\in B_\varepsilon(x)}\frac{d\mathbb{P}}{d\widehat{\mathbb{P}}}\,d\widehat{\mathbb{P}}\end{aligned}\tag{3.60}$$

$$
\begin{aligned}
&= \frac{1}{t}\log\widehat{\mathbb{P}}\left(\frac{N_t}{t}\in B_\varepsilon(x)\right) \\
&\quad + \frac{1}{t}\log\left[\frac{1}{\widehat{\mathbb{P}}((N_t/t)\in B_\varepsilon(x))}\int_{(N_t/t)\in B_\varepsilon(x)}\frac{d\mathbb{P}}{d\widehat{\mathbb{P}}}\,d\widehat{\mathbb{P}}\right] \\
&\geq \frac{1}{t}\log\widehat{\mathbb{P}}\left(\frac{N_t}{t}\in B_\varepsilon(x)\right) \\
&\quad - \frac{1}{\widehat{\mathbb{P}}((N_t/t)\in B_\varepsilon(x))}\cdot\frac{1}{t}\widehat{\mathbb{E}}\left[1_{(N_t/t)\in B_\varepsilon(x)}\log\frac{d\widehat{\mathbb{P}}}{d\mathbb{P}}\right].
\end{aligned}
$$

By the ergodic theorem,

$$
\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t}\in B_\varepsilon(x)\right)\geq -\Lambda(x), \tag{3.61}
$$

where

$$
\Lambda(x) = \inf_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e^x}\left\{\int(\lambda-\hat{\lambda})\hat{\pi} + \int\log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi}\right\} \tag{3.62}
$$

and

$$
\mathcal{Q}_e^x = \left\{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e : \int\hat{\lambda}(z)\hat{\pi}(dz) = x\right\}. \tag{3.63}
$$

Notice that

$$
\begin{aligned}
\Gamma(\theta) &= \sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e}\left\{\int\theta\hat{\lambda}\hat{\pi} + \int(\hat{\lambda}-\lambda)\hat{\pi} - \int\log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi}\right\} \\
&= \sup_x\sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e^x}\left\{\int\theta\hat{\lambda}\hat{\pi} + \int(\hat{\lambda}-\lambda)\hat{\pi} - \int\log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi}\right\} \\
&= \sup_x\sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e^x}\left\{\int\frac{\theta b}{a}z\hat{\pi}(dz) + \int(\hat{\lambda}-\lambda)\hat{\pi} - \int\log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi}\right\} \\
&= \sup_x\{\theta x - \Lambda(x)\}.
\end{aligned} \tag{3.64}
$$

We prove in Lemma 5 that $\Lambda(x)$ is convex in x, identify it as the convex conjugate of $\Gamma(\theta)$ and thus complete the proof. □

LEMMA 5. *$\Lambda(x)$ in (3.62) is convex in x.*

PROOF. Define

$$
\widehat{H}(\hat{\lambda},\hat{\pi}) = \int(\lambda-\hat{\lambda})\hat{\pi} + \int\log(\hat{\lambda}/\lambda)\hat{\lambda}\hat{\pi}. \tag{3.65}
$$

Then

$$\Lambda(x) = \inf_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e^x} \widehat{H}(\hat{\lambda},\hat{\pi}). \tag{3.66}$$

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 $\varepsilon > 0$, we can choose $(\hat{\lambda}_k, \hat{\pi}_k) \in \mathcal{Q}_e^{x_k}$ such that $\widehat{H}(\hat{\lambda}_k, \hat{\pi}_k) \leq \Lambda(x_k) + \varepsilon/2$, for $k = 1, 2$. Set

$$\hat{\pi}_3 = \alpha\hat{\pi}_1 + \beta\hat{\pi}_2, \qquad \hat{\lambda}_3 = \frac{d(\alpha\hat{\pi}_1)}{d(\alpha\hat{\pi}_1 + \beta\hat{\pi}_2)}\hat{\lambda}_1 + \frac{d(\beta\hat{\pi}_2)}{d(\alpha\hat{\pi}_1 + \beta\hat{\pi}_2)}\hat{\lambda}_2. \tag{3.67}$$

Then for any test function f,

$$\int \widehat{\mathcal{A}}_3 f\hat{\pi}_3 = \alpha\int\widehat{\mathcal{A}}_1 f\hat{\pi}_1 + \beta\int\widehat{\mathcal{A}}_2 f\hat{\pi}_2 = 0, \tag{3.68}$$

which implies $(\hat{\lambda}_3, \hat{\pi}_3) \in \mathcal{Q}_e$. Furthermore,

$$\int\hat{\lambda}_3\hat{\pi}_3 = \alpha\int\hat{\lambda}_1\hat{\pi}_1 + \beta\int\hat{\lambda}_2\hat{\pi}_2 = \alpha x_1 + \beta x_2. \tag{3.69}$$

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

$$\begin{aligned}
\widehat{H}(\hat{\lambda}_3, \hat{\pi}_3) &= \int\big[(\lambda - \hat{\lambda}_3 - \hat{\lambda}_3\log\lambda) + \hat{\lambda}_3\log\hat{\lambda}_3\big]\hat{\pi}_3 \\
&\leq \int\bigg[(\lambda - \hat{\lambda}_3 - \hat{\lambda}_3\log\lambda) + \alpha\frac{d\hat{\pi}_1}{d\hat{\pi}_3}\hat{\lambda}_1\log\hat{\lambda}_1 + \beta\frac{d\hat{\pi}_2}{d\hat{\pi}_3}\hat{\lambda}_2\log\hat{\lambda}_2\bigg]\hat{\pi}_3 \\
&= \alpha\widehat{H}(\hat{\lambda}_1, \hat{\pi}_1) + \beta\widehat{H}(\hat{\lambda}_2, \hat{\pi}_2).
\end{aligned} \tag{3.70}$$

Therefore,

$$\begin{aligned}
\Lambda(\alpha x_1 + \beta x_2) &\leq \widehat{H}(\hat{\lambda}_3, \hat{\pi}_3) \\
&\leq \alpha\widehat{H}(\hat{\lambda}_1, \hat{\pi}_1) + \beta\widehat{H}(\hat{\lambda}_2, \hat{\pi}_2) \\
&\leq \alpha\Lambda(x_1) + \beta\Lambda(x_2) + \varepsilon.
\end{aligned} \tag{3.71}$$

□

LEMMA 6. *If* $\limsup_{z\to\infty}\frac{\lambda(z)}{bz} < \frac{1}{a}$, *then for any*

$$\theta < \log\left(\frac{b}{a\limsup_{z\to\infty}(\lambda(z)/z)}\right) - 1 + \frac{a}{b}\cdot\limsup_{z\to\infty}\frac{\lambda(z)}{z}, \tag{3.72}$$

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

$$\begin{aligned}
\Gamma(\theta) &\leq \inf_{g\in\mathcal{U}_\theta}\sup_{z\geq 0}\frac{\mathcal{A}g(z) + ((\theta b)/a)zg(z)}{g(z)} \leq \sup_{z\geq 0}\left\{\frac{\mathcal{A}e^{Kz}}{e^{Kz}} + \frac{\theta b}{a}z\right\} \\
&= \sup_{z\geq 0}\left\{-\left(bK - \frac{\theta b}{a}\right)z + \lambda(z)\big(e^{Ka} - 1\big)\right\}.
\end{aligned} \tag{3.73}$$

Define the function

$$F(K) = -K + \limsup_{z\to\infty} \frac{\lambda(z)}{bz} \cdot \big(e^{Ka} - 1\big). \tag{3.74}$$

Then $F(0) = 0$, F is convex and $F(K) \to \infty$ as $K \to \infty$ and its minimum is attained at

$$K^* = \frac{1}{a} \log\left(\frac{b}{a \limsup_{z\to\infty}(\lambda(z)/z)}\right) > 0, \tag{3.75}$$

and $F(K^*) < 0$. Therefore, $\Gamma(\theta) < \infty$ for any

$$\begin{aligned} \theta &< -a \min_{K>0}\left\{-K + \limsup_{z\to\infty} \frac{\lambda(z)}{bz} \cdot (e^{Ka} - 1)\right\} \\ &= \log\left(\frac{b}{a \limsup_{z\to\infty}(\lambda(z)/z)}\right) - 1 + \frac{a}{b} \cdot \limsup_{z\to\infty} \frac{\lambda(z)}{z} < K^* a. \end{aligned} \tag{3.76}$$

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. $\square$

4. 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, that is, $h(t) = \sum_{i=1}^d a_i e^{-b_i t}$. Let

$$Z_i(t) = \sum_{\tau_j < t} a_i e^{-b_i(t-\tau_j)}, \qquad 1 \le i \le d \tag{4.1}$$

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 2 with $\lambda(Z_1(t), Z_2(t), \dots, Z_d(t))$ taking the form $\lambda(\sum_{i=1}^d Z_i(t))$. It is easy to see that $(Z_1, \dots, Z_d)$ is Markovian with generator

$$\begin{aligned} \mathcal{A}f &= -\sum_{i=1}^d b_i z_i \frac{\partial f}{\partial z_i} \\ &\quad + \lambda\left(\sum_{i=1}^d z_i\right) \cdot \big[f(z_1 + a_1, \dots, z_d + a_d) - f(z_1, \dots, z_d)\big]. \end{aligned} \tag{4.2}$$

Here $b_i > 0$ for any $1 \le i \le d$ 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 $t \ge 0$ for the rest of this paper. In particular, $h(0) = \sum_{i=1}^d a_i > 0$. If $a_i > 0$, then $Z_i(t) \ge 0$ almost surely; if $a_i < 0$, then $Z_i(t) \le 0$ almost surely.

THEOREM 7. *Assume* $\lim_{z\to\infty}\frac{\lambda(z)}{z}=0$, $\lambda(\cdot)$ *is continuous and bounded below by a positive constant. Then*

$$\lim_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]=\inf_{u\in\mathcal{U}_\theta}\sup_{(z_1,\dots,z_d)\in\mathcal{Z}}\left\{\frac{\mathcal{A}u}{u}+\frac{\theta}{\sum_{i=1}^d a_i}\sum_{i=1}^d b_i z_i\right\}, \tag{4.3}$$

where $\mathcal{Z}=\{(z_1,\dots,z_d): a_i z_i\geq 0, 1\leq i\leq d\}$ *and*

$$\mathcal{U}_\theta=\{u\in C_1(\mathbb{R}^d,\mathbb{R}^+), u=e^f, f\in\mathcal{F}\}, \tag{4.4}$$

where

$$\mathcal{F}=\left\{f=g+\frac{\theta\sum_{i=1}^d z_i}{\sum_{i=1}^d a_i}+L, L\in\mathbb{R}, g\in\mathcal{G}\right\}, \tag{4.5}$$

where

$$\mathcal{G}=\left\{\sum_{i=1}^d K\varepsilon_i z_i+g, K>0, g \text{ is } C_1 \text{ with compact support}\right\}. \tag{4.6}$$

PROOF. Notice that

$$dZ_i(t)=-b_i Z_i(t)\,dt+a_i\,dN_t, \qquad 1\leq i\leq d. \tag{4.7}$$

Hence $a_i N_t=Z_i(t)-Z_i(0)+\int_0^t b_i Z_i(s)\,ds$ and

$$\mathbb{E}[e^{\theta N_t}]=\mathbb{E}\left[\exp\left\{\frac{\theta\sum_{i=1}^d Z_i(t)-Z_i(0)}{\sum_{i=1}^d a_i}+\frac{\theta}{\sum_{i=1}^d a_i}\int_0^t\sum_{i=1}^d b_i Z_i(s)\,ds\right\}\right]. \tag{4.8}$$

Following the same arguments in the proof of Theorem 2, we obtain the upper bound

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]\leq\inf_{u\in\mathcal{U}_\theta}\sup_{(z_1,\dots,z_d)\in\mathcal{Z}}\left\{\frac{\mathcal{A}u}{u}+\frac{\theta}{\sum_{i=1}^d a_i}\sum_{i=1}^d b_i z_i\right\}. \tag{4.9}$$

As before, we can obtain the lower bound

$$\begin{aligned}
&\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]\\
&\quad\geq\sup_{(\hat\lambda,\hat\pi)\in\mathcal{Q}_e}\int[\theta\hat\lambda-\lambda+\hat\lambda-\hat\lambda\log(\hat\lambda/\lambda)]\hat\pi(dz_1,\dots,dz_d)\\
&\quad\geq\sup_{(\hat\lambda,\hat\pi)\in\mathcal{Q}}\inf_{g\in\mathcal{G}}\int[\theta\hat\lambda-\lambda+\hat\lambda-\hat\lambda\log(\hat\lambda/\lambda)+\hat{\mathcal{A}}g]\hat\pi\\
&\quad=\sup_{(\hat\lambda,\hat\pi)\in\mathcal{Q}}\inf_{f\in\mathcal{F}}\int\left[\frac{\theta\sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i}-\lambda+\hat\lambda-\hat\lambda\log(\hat\lambda/\lambda)+\hat{\mathcal{A}}f\right]\hat\pi.
\end{aligned} \tag{4.10}$$

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

$$\mathcal{G} = \left\{ \sum_{i=1}^d K\varepsilon_i z_i + g, K > 0, g \text{ is } C_1 \text{ with compact support} \right\}. \tag{4.11}$$

Here, $\varepsilon_i = a_i/|a_i|$, $1 \leq i \leq d$. Define

$$F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) = \int \left[\frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} + \widehat{\mathcal{A}} f \right] \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}). \tag{4.12}$$

F is linear in f and hence convex in f. Also $\widehat{H}$ 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 $f = \frac{\theta \sum_{i=1}^d z_i}{\sum_{i=1}^d a_i} + \sum_{i=1}^d K\varepsilon_i z_i + g + L \in \mathcal{F}$,

$$\begin{aligned} F(\hat{\lambda}\hat{\pi}, \hat{\pi}, f) &= \int \left[\theta + \sum_{i=1}^d K\varepsilon_i a_i \right] \hat{\lambda}\hat{\pi} \\ &\quad - \int \sum_{i=1}^d K\varepsilon_i b_i z_i \hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) + \int \widehat{\mathcal{A}} g \hat{\pi}. \end{aligned} \tag{4.13}$$

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

$$\begin{aligned} &\int \left[\theta + \sum_{i=1}^d K\varepsilon_i a_i \right] \lambda_n \pi_n + \int \widehat{\mathcal{A}} g \pi_n \\ &\quad \to \int \left[\theta + \sum_{i=1}^d K\varepsilon_i a_i \right] \gamma_\infty + \int \widehat{\mathcal{A}} g \pi_\infty. \end{aligned} \tag{4.14}$$

Since $-\sum_{i=1}^d K\varepsilon_i b_i z_i$ is continuous and nonpositive on $\mathcal{Z}$, we have

$$\limsup_{n\to\infty} \int \left[-\sum_{i=1}^d K\varepsilon_i b_i z_i \right] \pi_n \leq \int \left[-\sum_{i=1}^d K\varepsilon_i b_i z_i \right] \pi_\infty. \tag{4.15}$$

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

$$\left\{ (\hat{\lambda}\hat{\pi}, \hat{\pi}) : \int \left[-\frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} - \widehat{\mathcal{A}} f \right] \hat{\pi} + \widehat{H}(\hat{\lambda}, \hat{\pi}) \leq C \right\}. \tag{4.16}$$

For any $f = \frac{\theta \sum_{i=1}^d z_i}{\sum_{i=1}^d a_i} + \sum_{i=1}^d K\varepsilon_i z_i + g + L \in \mathcal{F}$, where g is C_1 with compact support, etc., there exist some $C_1, C_2 > 0$ such that

$$
\begin{aligned}
C_1 &\geq \widehat{H} + \sum_{i=1}^d K b_i \varepsilon_i \int z_i \hat{\pi} - C_2 \int \hat{\lambda}\hat{\pi} \\
&\geq \int_{\hat{\lambda} \geq \sum_{i=1}^d c_i z_i + \ell} [\lambda - \hat{\lambda} + \hat{\lambda}\log(\hat{\lambda}/\lambda)]\hat{\pi} \\
&\quad + \sum_{i=1}^d K b_i \varepsilon_i \int z_i \hat{\pi} \\
&\quad - C_2 \int_{\hat{\lambda} \geq \sum_{i=1}^d c_i z_i + \ell} \hat{\lambda}\hat{\pi} - C_2 \int_{\hat{\lambda} < \sum_{i=1}^d c_i z_i + \ell} \hat{\lambda}\hat{\pi} \\
&\geq \left[\min_{(z_1,\dots,z_d)\in\mathcal{Z}} \log \frac{c_1 z_1 + \cdots + c_d z_d + \ell}{\lambda(z_1 + \cdots + z_d)} - 1 - C_2 \right] \int_{\hat{\lambda} \geq \sum_{i=1}^d c_i z_i + \ell} \hat{\lambda}\hat{\pi} \\
&\quad + \sum_{i=1}^d [-c_i \cdot C_2 + K b_i \varepsilon_i] \int z_i \hat{\pi} - \ell C_2.
\end{aligned} \tag{4.17}
$$

If $a_i > 0$, then $\varepsilon_i > 0$, pick up $c_i > 0$ such that $-c_i \cdot C_2 + K b_i \varepsilon_i > 0$. If $a_i < 0$, then $\varepsilon_i < 0$, pick up c_i such that $-c_i \cdot C_2 + K b_i \varepsilon_i < 0$. Finally, choose ℓ big enough such that the big bracket above is positive. Then

$$
\int |z_i|\hat{\pi} \leq C_3, \qquad \int_{\hat{\lambda} \geq \sum_{i=1}^d c_i z_i + \ell} \hat{\lambda}\hat{\pi} \leq C_4. \tag{4.18}
$$

Hence, $\int \hat{\lambda}\hat{\pi} \leq C_5$ and $\widehat{H} \leq C_6$. We can use a method similar to the proof of Theorem 2 to show that

$$
\lim_{\ell\to\infty} \sup_n \int_{|z_i| > \ell} \lambda_n \pi_n = 0, \qquad 1 \leq i \leq d. \tag{4.19}
$$

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,

$$
\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \\
&\quad \geq \sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}} \inf_{f\in\mathcal{F}} \int \left[\frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} - \lambda + \hat{\lambda} - \hat{\lambda}\log(\hat{\lambda}/\lambda) + \widehat{\mathcal{A}} f \right] \hat{\pi} \\
&\quad = \inf_{f\in\mathcal{F}} \sup_{\hat{\pi}} \sup_{\hat{\lambda}} \int \left[\frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} - \lambda + \hat{\lambda} - \hat{\lambda}\log(\hat{\lambda}/\lambda) + \widehat{\mathcal{A}} f \right] \hat{\pi} \\
&\quad = \inf_{f\in\mathcal{F}} \sup_{(z_1,\dots,z_d)\in\mathcal{Z}} \frac{\theta \sum_{i=1}^d b_i z_i}{\sum_{i=1}^d a_i} + \lambda\left(e^{f(z_1+a_1,\dots,z_d+a_d) - f(z_1,\dots,z_d)} - 1\right)
\end{aligned} \tag{4.20}
$$

$$
\begin{aligned}
&-\sum_{i=1}^{d} b_i z_i \frac{\partial f}{\partial z_i} \\
&\geq \inf_{u\in\mathcal{U}_\theta} \sup_{(z_1,\ldots,z_d)\in\mathcal{Z}} \left\{ \frac{\mathcal{A}u}{u} + \frac{\theta}{\sum_{i=1}^{d} a_i} \sum_{i=1}^{d} b_i z_i \right\}.
\end{aligned}
$$

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 $f \in \mathcal{F}$, $u = e^f \in \mathcal{U}_\theta$. $\square$

THEOREM 8. *Assume* $\lim_{z\to\infty} \frac{\lambda(z)}{z} = 0$, $\lambda(\cdot)$ *is positive 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* $\Gamma(\cdot)$,

$$
I(x) = \sup_{\theta\in\mathbb{R}} \{\theta x - \Gamma(\theta)\}, \tag{4.21}
$$

where

$$
\Gamma(\theta) = \sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e} \int [\theta\hat{\lambda} - \lambda + \hat{\lambda} - \hat{\lambda}\log(\hat{\lambda}/\lambda)]\hat{\pi}. \tag{4.22}
$$

PROOF. The proof is the same as in the case of exponential $h(\cdot)$. $\square$

5. Large deviations for linear Hawkes processes: An alternative proof. In this section, we use our method to recover the result proved in Bordenave and Torrisi [2]. We prove the existence of the limit of logarithmic moment generating function first. The strategy is to use the tilting method to prove the lower bound. This requires an ergodic lemma, which we state as Lemma 9. For the upper bound, we can opitimize over a special class of testing functions for the linear rate with the sum of exponential exciting function h_n. Any continuous and integrable h can be approximated by a sequence h_n. By a coupling argument, we can use that to approximate the upper bound for the logarithmic moment generating function when the exciting function is h. Finally, by a tilting argument for the lower bound and the Gärtner–Ellis theorem for the upper bound, we can prove the large deviations for the linear Hawkes processes.

LEMMA 9. *Assume* $\lambda(z) = \alpha + \beta z$ *and* $\mu = \int_0^\infty h(t)\,dt < \infty$. *If* $\beta\mu < 1$, *then there exists a stationary and ergodic probability measure* π *for* Z_t *and* $\int z\pi = \frac{\alpha\mu}{1-\beta\mu}$.

PROOF. The ergodicity is a well-known result for linear Hawkes process; see Hawkes and Oakes [13]. Let π be the invariant probability measure for Z_t, then

$$
\lim_{t\to\infty} \frac{N_t}{t} = \int \lambda(z)\pi(dz) = \alpha + \beta \int z\pi(dz). \tag{5.1}
$$

If Z_t is invariant in t, taking expectations to $Z_t = \int_{-\infty}^t h(t-s)\,dN_s$,

$$(5.2)\qquad \begin{aligned}\mathbb{E}[Z_t] &= \int z\pi(dz) = \int \lambda(z)\pi(dz)\int_{-\infty}^t h(t-s)\,ds \\ &= \mu \int \lambda(z)\pi(dz),\end{aligned}$$

which implies that $\int z\pi = \frac{\alpha\mu}{1-\beta\mu}$. $\square$

REMARK 10. In Lemma 9, we assumed that $\lambda(z) = \alpha + \beta z$ and $\beta\|h\|_{L^1} < 1$. However, when do the LDP for linear Hawkes process and when we prove Theorem 12, we assume that $\lambda(z) = \nu + z$ since $\lambda(z) = \nu + \beta z$ is equivalent to the case $\lambda(z) = \nu + z$ if we change $h(\cdot)$ to $\beta h(\cdot)$. The reason we used $\lambda(z) = \alpha + \beta z$ in Lemma 9 is because we need to use it when we tilt $\lambda(z) = \nu + z$ to $K\lambda(z) = K\nu + Kz$ in the proof of lower bound in Theorem 12.

LEMMA 11. *If $h(t) > 0$, $\int_0^\infty h(t)\,dt < \infty$, $\lim_{t\to\infty} h(t) = 0$, 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 $X = \mathbb{R}_{\geq 0} \cup \{\infty\} = [0,\infty]$ that is compactified and $C[0,\infty]$ consists of continuous functions vanishing at ∞ and the constant function 1.

By the 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

$$(5.3)\qquad \sup_{t\geq 0}\left| h(t) - \sum_{j=0}^n a_j e^{-jt}\right| \leq \varepsilon.$$

In fact, since $h(\infty) = 0$, we get $|a_0| \leq \varepsilon$. Thus

$$(5.4)\qquad \sup_{t\geq 0}\left| h(t) - \sum_{j=1}^n a_j e^{-jt}\right| \leq 2\varepsilon.$$

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.

$$\|h - h_n e^{-\delta_n t}\|_{L^\infty} \le \|h - h_n\|_{L^\infty} + \|h_n - h_n e^{-\delta_n t}\|_{L^\infty}. \tag{5.5}$$

Notice that $(1 - e^{-\delta_n t}) h_n \le (1 - e^{-\delta_n t})(h(t) + \varepsilon)$. Since $h(\infty) = 0$, there exists some $M > 0$, such that for $t > M$, $h(t) \le \varepsilon$ so that $(1 - e^{-\delta_n t}) h_n \le 2\varepsilon$ for $t > M$. For $t \le M$, $(1 - e^{-\delta_n t}) h_n \le (1 - e^{-\delta_n M})(\|h\|_{L^\infty} + \varepsilon)$ which is small if δ_n is small. □

THEOREM 12. *Assume $\lambda(z) = \nu + z$, $\nu > 0$. $h(\cdot)$ satisfies the assumptions in Lemma 11 and $\int_0^\infty h(t)\, dt < 1$. We have*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \nu(x - 1), \tag{5.6}$$

where x is the minimal solution to $x = e^{\theta + \mu(x-1)}$, where $\mu = \int_0^\infty h(t)\, dt$.

PROOF. By Lemma 9, we have

$$\begin{aligned}
&\liminf_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \\
&\quad\ge \sup_{(\hat\lambda, \hat\pi) \in \mathcal{Q}_e} \int [\theta \hat\lambda + \hat\lambda - \lambda - \hat\lambda \log(\hat\lambda/\lambda)] \hat\pi \\
&\quad\ge \sup_{(K\lambda, \hat\pi) \in \mathcal{Q}_e, K \in \mathbb{R}^+} \int [\theta \hat\lambda + \hat\lambda - \lambda - \hat\lambda \log(\hat\lambda/\lambda)] \hat\pi \\
&\quad\ge \sup_{0 < K < 1/\mu, (K\lambda, \hat\pi) \in \mathcal{Q}_e} \int \left[\theta + 1 - \frac{1}{K} - \log K\right] \hat\lambda \hat\pi \\
&\quad\ge \sup_{0 < K < 1/\mu} \left[\theta + 1 - \frac{1}{K} - \log K\right] \cdot \frac{K\nu}{1 - K\mu} \\
&\quad= \begin{cases} \nu(x - 1), & \text{if } \theta \in (-\infty, \mu - 1 - \log \mu], \\ +\infty, & \text{otherwise}, \end{cases}
\end{aligned} \tag{5.7}$$

where x is the minimal solution to $x = e^{\theta + \mu(x-1)}$.

By Lemma 11, we can find a sequence of h_n, where $h_n(t) = \sum_{i=1}^n a_i e^{-b_i t}$ such that $h_n \to h$ as $n \to \infty$ in both L^1 and L^∞ norms. Let $h_\varepsilon(t) = |h(t) - h_n(t)|$. Then $0 \le h_n - h_\varepsilon \le h \le h_n + h_\varepsilon$.

Let D_1 be the set of points generated by the Hawkes process with intensity $\lambda(\sum_{\tau \in D_1, \tau < t} h_n(t - \tau))$ and then conditional on D_1, let D_2 be the set of points generated by the point process with intensity $\lambda(\sum_{\tau \in D_1, \tau < t} (h_n + h_\varepsilon)(t -$

$\tau)) - \lambda(\sum_{\tau\in D_1,\tau<t} h_n(t-\tau))$ and then iteratively, conditional on $D_1,\dots,D_{j-1}$, let D_j be the set of points generated by the point process with intensity $\lambda(\sum_{\tau\in\bigcup_{i=1}^{j-1} D_i,\tau<t}(h_n+h_\varepsilon)(t-\tau)) - \lambda(\sum_{\tau\in\bigcup_{i=1}^{j-2} D_i,\tau<t}(h_n+h_\varepsilon)(t-\tau))$, for any $j\geq 3$. Let $D_j(t)$ correspond to the number of points in D_j by time t. Therefore, $\sum_{j=1}^{\infty} D_j(t)$ equals the number of points generated by Hawkes process with intensity $\lambda(\sum_{\tau<t}(h_n+h_\varepsilon)(t-\tau))$. Our coupling argument is essentially the same as the one used in Brémaud and Massoulié [3]. For a more formal treatment, one can use Poisson canonical space and Poisson embeddings; we refer to Brémaud and Massoulié [3] for the details.

Assume that $\theta > 0$, and we therefore have

$$\mathbb{E}[e^{\theta N_t}] \leq \mathbb{E}[e^{\theta\sum_{j=1}^{\infty} D_j(t)}]. \tag{5.8}$$

Now, for any $N\in\mathbb{N}$,

$$\begin{aligned}
&\mathbb{E}\left[\exp\left\{\theta\sum_{j=1}^{N} D_j(t)\right\}\right]\\
&\quad=\mathbb{E}\Bigg[\exp\left\{\theta\sum_{j=1}^{N-1} D_j(t)\right\}\\
&\qquad\times\exp\Bigg\{(e^\theta-1)\int_0^t \lambda\Bigg(\sum_{\tau\in\bigcup_{i=1}^{N-1} D_i,\tau<s}(h_n+h_\varepsilon)(s-\tau)\Bigg)\\
&\qquad\qquad-\lambda\Bigg(\sum_{\tau\in\bigcup_{i=1}^{N-2} D_i,\tau<s}(h_n+h_\varepsilon)(s-\tau)\Bigg)ds\Bigg\}\Bigg]\\
&\quad\leq\mathbb{E}\left[\exp\left\{\theta\sum_{j=1}^{N-2} D_j(t)\right\}\exp\{((e^\theta-1)\|h_n+h_\varepsilon\|_{L^1}+\theta)D_{N-1}(t)\}\right]\\
&\quad\leq\cdots\\
&\quad\leq\mathbb{E}[\exp\{\theta D_1(t)+f_{N-1}(\theta)D_2(t)\}]\\
&\quad\leq\mathbb{E}[\exp\{\theta D_1(t)+(\exp\{f_{N-1}(\theta)\}-1)\|h_\varepsilon\|_{L^1}D_1(t)\}],
\end{aligned}\tag{5.9}$$

where $f_j(\theta)=(e^{f_{j-1}(\theta)}-1)\|h_n+h_\varepsilon\|_{L^1}+\theta$, for $j\geq 2$ and $f_1(\theta)=\theta$. Thus, for any $\theta\leq\|h_n+h_\varepsilon\|_{L^1}-1-\log(\|h_n+h_\varepsilon\|_{L^1})$, $e^{f_{N-1}(\theta)}$ converges to y_n as $N\to\infty$, where y_n is the minimal solution to $y_n=e^{\theta+\|h_n+h_\varepsilon\|_{L^1}(y_n-1)}$. Since $D_1(t)$ is the Hawkes process with exciting function h_n,

$$\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]\leq\Gamma_n(p_n\theta), \tag{5.10}$$

where $p_n = 1 + y_n \|h - h_n\|_{L^1}$. For $\Gamma_n(p_n\theta)$, we have

$$\begin{aligned}\Gamma_n(p_n\theta) &= \inf_{u\in\mathcal{U}_{p_n\theta}} \sup_{(z_1,\dots,z_n)\in\mathcal{Z}} \left\{\frac{\mathcal{A}u}{u} + \frac{p_n\theta}{\sum_{i=1}^n a_i}\sum_{i=1}^n b_i z_i\right\} \\ &\le \inf_{u=e^{\sum_{i=1}^n c_i z_i}\in\mathcal{U}_{p_n\theta}} \sup_{(z_1,\dots,z_n)\in\mathcal{Z}} \left\{\frac{\mathcal{A}u}{u} + \frac{p_n\theta}{\sum_{i=1}^n a_i}\sum_{i=1}^n b_i z_i\right\} \\ &= \inf_{c_1,\dots,c_n} \sup_{(z_1,\dots,z_n)\in\mathcal{Z}} \left\{-\sum_{i=1}^n b_i c_i z_i + (\nu + z_1 + \cdots + z_n)\big(e^{\sum_{i=1}^n c_i a_i} - 1\big)\right. \\ &\qquad\qquad \left. + \frac{p_n\theta}{\sum_{i=1}^n a_i}\sum_{i=1}^n b_i z_i\right\} \\ &= \nu\big(e^{\sum_{i=1}^n c_i^* a_i} - 1\big) = \nu(x_n - 1),\end{aligned}$$

where c_i^* satisfies $-b_i c_i^* + e^{\sum_{i=1}^n c_i^* a_i} - 1 + \frac{p_n\theta}{\sum_{i=1}^n a_i} b_i = 0$, for each $1 \le i \le n$. By some computation, it is not hard to see that $x_n = e^{\sum_{i=1}^n c_i^* a_i}$ satisfies

$$\begin{aligned}x_n &= \exp\left\{p_n\theta + \sum_{i=1}^n \frac{a_i}{b_i}(x_n - 1)\right\} \\ &= \exp\left\{(1 + y_n\|h - h_n\|_{L^1})\theta + (x_n - 1)\int_0^\infty h_n(t)\,dt\right\}.\end{aligned} \tag{5.11}$$

Since $h_n \to h$ in L^1 norm, it is not hard to see that x_n converges to the minimal solution of $x = e^{\theta + \|h\|_{L^1}(x-1)}$ as $n \to \infty$. If $\theta < 0$, consider $h \ge h_n - h_\varepsilon \ge 0$ and the argument is similar. □

THEOREM 13. *Assume $\lambda(z) = \nu + z$, $h:[0,\infty) \to \mathbb{R}^+$, $\mu := \int_0^\infty h(t)\,dt < 1$ and h is continuous. Then $(N_t/t \in \cdot)$ satisfies a large deviation principle with the rate function $I(x)$ given by*

$$I(x) = \begin{cases} x\log\left(\dfrac{x}{\nu + x\mu}\right) - x + \mu x + \nu, & \text{if } x \in [0,\infty), \\ +\infty, & \text{otherwise.}\end{cases} \tag{5.12}$$

PROOF. For the upper bound, apply the Gärtner–Ellis theorem. For the lower bound, use the tilting method and identify $I(x)$ as the Fenchel–Legendre transform of $\Gamma(\theta)$. □

REMARK 14. In Bordenave and Torrisi [2], their $I(x)$ has the form

$$I(x) = \begin{cases} x\theta_x + \nu - \dfrac{\nu x}{\nu + \mu x}, & \text{if } x \in [0,\infty), \\ +\infty, & \text{otherwise,}\end{cases} \tag{5.13}$$

where $\theta = \theta_x$ is the unique solution in $(-\infty, \mu - 1 - \log \mu]$ of $\mathbb{E}[e^{\theta S}] = \frac{x}{\nu + x\mu}$, $x > 0$. Here, $\mathbb{E}[e^{\theta S}]$ satisfies the equation

$$\mathbb{E}[e^{\theta S}] = e^{\theta} \exp\{\mu(\mathbb{E}[e^{\theta S}] - 1)\}, \tag{5.14}$$

which implies that $\theta_x = \log(\frac{x}{\nu + x\mu}) - \mu(\frac{x}{\nu + x\mu} - 1)$. Substituting into the formula, their rate function is the same as what we got.

REMARK 15. In Bordenave and Torrisi [2], the assumption in proving the large deviations for linear Hawkes processes is slightly different from ours. They did not require $h(\cdot)$ to be continuous, but they further assumed that $\int_0^\infty t h(t)\, dt < \infty$.

6. Large deviations for a special class of nonlinear Hawkes processes: An approximation approach. In this section, we prove the large deviation results for $(N_t/t \in \cdot)$ for a very special class of nonlinear $\lambda(\cdot)$ and $h(\cdot)$ that satisfies the assumptions in Lemma 11.

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. Let us define

$$\Gamma_n(\theta) = \lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}^{P_n}[e^{\theta N_t}]. \tag{6.1}$$

We have the following results.

LEMMA 16. *For any $K > 0$ and $\theta_1, \theta_2 \in [-K, K]$, there exists some constant $C(K)$ such that for any n,*

$$|\Gamma_n(\theta_1) - \Gamma_n(\theta_2)| \leq C(K)|\theta_1 - \theta_2|. \tag{6.2}$$

PROOF. Without loss of generality, take $\theta_2 > \theta_1$. Then

$$\begin{aligned}\Gamma_n(\theta_1) &\leq \Gamma_n(\theta_2) \\ &= \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e^*} \int (\theta_2 - \theta_1)\hat{\lambda}\hat{\pi} + \theta_1 \hat{\lambda}\hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) \\ &\leq \sup_{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e^*} \int (\theta_2 - \theta_1)\hat{\lambda}\hat{\pi} + \Gamma_n(\theta_1),\end{aligned} \tag{6.3}$$

where

$$\mathcal{Q}_e^* = \left\{(\hat{\lambda}, \hat{\pi}) \in \mathcal{Q}_e : \int \theta_1 \hat{\lambda}\hat{\pi} - \widehat{H}(\hat{\lambda}, \hat{\pi}) \geq \Gamma_n(\theta_1) - 1\right\}. \tag{6.4}$$

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 constant $C(K)>0$ depending only on K. Define $u=u(z_1,\dots,z_n)=e^{\sum_{i=1}^n c_i z_i}$ where

$$c_i=\frac{3K}{\sum_{i=1}^n(a_i/b_i)}\cdot\frac{1}{b_i},\qquad 1\leq i\leq n. \tag{6.5}$$

Define $V=-\frac{\mathcal{A}u}{u}$ such that

$$V(z_1,\dots,z_n)=\frac{3K}{\sum_{i=1}^n(a_i/b_i)}\sum_{i=1}^n z_i-\lambda(z_1+\cdots+z_n)\big(e^{3K}-1\big). \tag{6.6}$$

Notice that $\int\widehat{\mathcal{A}}f\hat{\pi}=0$ for any test function f with certain regularities. If we try $f=\frac{z_i}{b_i}$, $1\leq i\leq n$, we get

$$-\int z_i\hat{\pi}+\frac{a_i}{b_i}\int\hat{\lambda}\hat{\pi}=0,\qquad 1\leq i\leq n. \tag{6.7}$$

Summing over $1\leq i\leq n$, we get

$$\int\hat{\lambda}\hat{\pi}=\frac{1}{\sum_{i=1}^n(a_i/b_i)}\int\sum_{i=1}^n z_i\hat{\pi}. \tag{6.8}$$

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

$$\begin{aligned}\theta_1\int\hat{\lambda}\hat{\pi}&\leq K\int\hat{\lambda}\hat{\pi}\\&=\frac{K}{\sum_{i=1}^n(a_i/b_i)}\int\sum_{i=1}^n z_i\hat{\pi}\\&\leq\frac{1}{2}\int V\hat{\pi}+C_{1/2}(K),\end{aligned} \tag{6.9}$$

where $C_{1/2}(K)$ is some positive constant depending only on K.

We claim that $\int V(z)\hat{\pi}\leq\widehat{H}(\hat{\pi})$ for any $\hat{\pi}\in\mathcal{Q}_e^*$. Let us prove it. By the ergodic theorem and Jensen's inequality,

$$\begin{aligned}\int V(z)\hat{\pi}&=\lim_{t\to\infty}\mathbb{E}^{\hat{\pi}}\left[\frac{1}{t}\int_0^t V(Z_s)\,ds\right]\\&\leq\limsup_{t\to\infty}\frac{1}{t}\log\mathbb{E}^{\pi}\big[e^{\int_0^t V(Z_s)\,ds}\big]+\widehat{H}(\hat{\pi}).\end{aligned} \tag{6.10}$$

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

$$Y_t=\sum_{i=1}^n\frac{Z_i(t)}{b_i}=\sum_{\tau_j<t}\sum_{i=1}^n\frac{a_i}{b_i}e^{-b_i(t-\tau_j)}=\sum_{\tau_j<t}g(t-\tau_j), \tag{6.11}$$

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 $\frac{\mathcal{A}u}{u} + V = 0$ and $u \geq 1$, by the Feynman–Kac formula and Dynkin's formula,

$$
\begin{aligned}
\mathbb{E}^\pi\bigl[e^{\int_0^t V(Z_s)\,ds}\bigr] &\leq \mathbb{E}^\pi\bigl[u(Z_t)e^{\int_0^t V(Z_s)\,ds}\bigr] \\
&= u(Z_0) + \int_0^t \mathbb{E}^\pi\bigl[\bigl(\mathcal{A}u(Z_s) + V(Z_s)u(Z_s)\bigr)e^{\int_0^s V(Z_u)\,du}\bigr]\,ds \\
&= u(Z_0),
\end{aligned}
\tag{6.12}
$$

and therefore $\int V(z)\hat{\pi} \leq \widehat{H}(\hat{\pi})$ for any $\hat{\pi} \in \mathcal{Q}_e^*$. Hence

$$
\theta_1 \int \hat{\lambda}\hat{\pi} \leq \frac{1}{2}\int V(z) + C_{1/2}(K) \leq \frac{1}{2}\widehat{H} + C_{1/2}(K). \tag{6.13}
$$

Notice that

$$
-\infty < \Gamma_n(\theta_1) - 1 \leq \theta_1 \int \hat{\lambda}\hat{\pi} - \widehat{H} \leq \Gamma_n(\theta_1) < \infty. \tag{6.14}
$$

Hence

$$
\Gamma_n(\theta_1) - 1 + \frac{1}{2}\widehat{H} \leq \theta_1 \int \hat{\lambda}\hat{\pi} - \frac{1}{2}\widehat{H} \leq C_{1/2}(K), \tag{6.15}
$$

which implies $\widehat{H} \leq 2(C_{1/2}(K) - \Gamma_n(\theta_1) + 1)$ and so also,

$$
\begin{aligned}
\int \hat{\lambda}\hat{\pi} &\leq \frac{1}{2K}\int V\hat{\pi} + \frac{1}{K}C_{1/2}(K) \\
&\leq \frac{1}{K}\bigl(C_{1/2}(K) - \Gamma_n(\theta_1) + 1\bigr) + \frac{1}{K}C_{1/2}(K).
\end{aligned}
\tag{6.16}
$$

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

$$
\Gamma_n(\theta_1) \geq \Gamma_n(-K) \geq \Gamma_g(-K), \tag{6.17}
$$

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 $\lambda(z) = \nu_\varepsilon + \varepsilon z$ that dominates the original $\lambda(\cdot)$ for $\nu_\varepsilon > 0$ big enough and $\varepsilon > 0$ small enough so that $\varepsilon\|g\|_{L^1} < 1$. Now, we have $\Gamma_g(-K) \geq \Gamma_{\varepsilon g}^{\nu_\varepsilon}(-K)$ which is finite (see Theorem 12), where $\Gamma_{\varepsilon g}^{\nu_\varepsilon}(-K)$ corresponds to the case when $\lambda(z) = \nu_\varepsilon + \varepsilon z$. Hence

$$
\sup_{(\hat{\lambda},\hat{\pi})\in\mathcal{Q}_e^*} \int \hat{\lambda}\hat{\pi} \leq C(K), \tag{6.18}
$$

for some $C(K) > 0$ depending only on K. $\square$

LEMMA 17. *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,

$$\exp\left\{q \int_0^t \log\left(\frac{\lambda(H_m(s))}{\lambda(H_n(s))}\right) dN_s - \int_0^t \left(\frac{\lambda(H_m(s))^q}{\lambda(H_n(s))^{q-1}} - \lambda\bigl(H_n(s)\bigr)\right) ds\right\} \tag{6.19}$$

is a martingale under P_n. By Hölder's inequality, for any $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$,

$$\begin{aligned}
\mathbb{E}^{P_m}\bigl[e^{\theta N_t}\bigr] &= \mathbb{E}^{P_n}\left[e^{\theta N_t} \frac{dP_m}{dP_n}\right] \\
&= \mathbb{E}^{P_n}\bigl[e^{\theta N_t - \int_0^t (\lambda(H_m(s)) - \lambda(H_n(s)))\, ds - \int_0^t \log(\lambda(H_n(s))/\lambda(H_m(s)))\, dN_s}\bigr] \\
&\leq \mathbb{E}^{P_n}\bigl[e^{p\theta N_t - p\int_0^t (\lambda(H_m(s)) - \lambda(H_n(s)))\, ds}\bigr]^{1/p} \\
&\quad \times \mathbb{E}^{P_n}\bigl[e^{q\int_0^t \log(\lambda(H_m(s))/\lambda(H_n(s)))\, dN_s}\bigr]^{1/q}.
\end{aligned} \tag{6.20}$$

By the Cauchy–Schwarz inequality,

$$\begin{aligned}
&\mathbb{E}^{P_n}\bigl[e^{q\int_0^t \log(\lambda(H_m(s))/\lambda(H_n(s)))\, dN_s}\bigr]^{1/q} \\
&\quad \leq \mathbb{E}^{P_n}\bigl[e^{\int_0^t (\lambda(H_m(s))^{2q}/\lambda(H_n(s))^{2q-1} - \lambda(H_n(s)))\, ds}\bigr]^{1/(2q)} \\
&\quad \leq \mathbb{E}^{P_n}\bigl[e^{(1/c^{2q-1}) L_{2q} \int_0^t \sum_{\tau < s} |h_m(s-\tau) - h_n(s-\tau)|\, ds}\bigr]^{1/(2q)} \\
&\quad \leq \mathbb{E}^{P_n}\bigl[e^{(1/c^{2q-1}) L_{2q} \|h_m - h_n\|_{L^1} N_t}\bigr]^{1/(2q)}.
\end{aligned} \tag{6.21}$$

We also have

$$\begin{aligned}
&\mathbb{E}^{P_n}\bigl[e^{p\theta N_t - p\int_0^t (\lambda(H_m(s)) - \lambda(H_n(s)))\, ds}\bigr]^{1/p} \\
&\qquad \leq \mathbb{E}^{P_n}\bigl[e^{p\theta N_t + pL_1 \|h_m - h_n\|_{L^1} N_t}\bigr]^{1/p}.
\end{aligned} \tag{6.22}$$

Therefore, by Lemma 16 and the fact $\Gamma_n(0) = 0$ for any n, we have

$$\begin{aligned}
&\Gamma_m(\theta) - \Gamma_n(\theta) \\
&\quad \leq \frac{1}{p}\Gamma_n(p\theta + pL_1\varepsilon_{m,n}) + \frac{1}{2q}\Gamma_n\left(\frac{L_{2q}\varepsilon_{m,n}}{c^{2q-1}}\right) - \Gamma_n(\theta) \\
&\quad \leq C(K)L_1\varepsilon_{m,n} + \frac{C(K)}{2q} \cdot \frac{L_{2q}\varepsilon_{m,n}}{c^{2q-1}} \\
&\qquad + \frac{1}{p}\Gamma_n(p\theta) - \frac{1}{p}\Gamma_n(\theta) + \left(1 - \frac{1}{p}\right)\bigl|\Gamma_n(\theta)\bigr|, \\
&\quad \leq C(K)L_1\varepsilon_{m,n} + \frac{C(K)}{2q} \cdot \frac{L_{2q}\varepsilon_{m,n}}{c^{2q-1}} \\
&\qquad + \frac{C(K)(p-1)K}{p} + \left(1 - \frac{1}{p}\right)C(K)K,
\end{aligned} \tag{6.23}$$

where $\varepsilon_{m,n} = \|h_m - h_n\|_{L^1}$. Hence,

$$\limsup_{m,n\to\infty}\{\Gamma_m(\theta) - \Gamma_n(\theta)\} \le 2\left(1 - \frac{1}{p}\right)C(K)K, \tag{6.24}$$

which is true for any $p > 1$. Letting $p \downarrow 1$, we get the desired result. □

REMARK 18. If $\lambda(\cdot) \ge c > 0$ and $\lim_{z\to\infty} \frac{\lambda(z)}{z^\alpha} = 0$ for any $\alpha > 0$, then, $\lambda(\cdot)^\sigma$ is Lipschitz for any $\sigma \ge 1$. For instance, $\lambda(z) = [\log(z + c)]^\beta$ satisfies the conditions if $\beta > 0$ and $c > 1$.

THEOREM 19. *Assume that $\lambda(\cdot) \ge c$ 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 $\alpha \ge 1$. Then, for any $\theta \in \mathbb{R}$,*

$$\lim_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \Gamma(\theta) = \lim_{n\to\infty} \Gamma_n(\theta). \tag{6.25}$$

PROOF. By Lemma 17, $\Gamma_n(\theta)$ tends to $\Gamma(\theta)$ uniformly on any compact set $[-K, K]$. Since $\Gamma_n(\theta)$ is Lipschitz by Lemma 16, it is continuous and the limit Γ is also continuous. Let $\varepsilon_n = \|h_n - h\|_{L^1} \le \varepsilon$. As in the proof of Lemma 17, for any $\theta \in [-K, K]$, $p, q > 1$, $\frac{1}{p} + \frac{1}{q} = 1$, we get

$$\begin{aligned} &\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \\ &\quad\le \Gamma_n(\theta) + C(K)L_1\varepsilon_n + \frac{C(K)}{2q} \cdot \frac{L_{2q}\varepsilon_n}{c^{2q-1}} + 2\left(1 - \frac{1}{p}\right)C(K)K. \end{aligned} \tag{6.26}$$

Letting $n \to \infty$ first and then $p \downarrow 1$, we get $\limsup_{t\to\infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] \le \Gamma(\theta)$. Similarly, for any $p', q' > 1$ with $\frac{1}{p'} + \frac{1}{q'} = 1$,

$$\begin{aligned} \Gamma_n(\theta) &\le \liminf_{t\to\infty} \frac{1}{pt} \log \mathbb{E}[e^{(p\theta + pL_1\varepsilon_n)N_t}] \\ &\quad + \liminf_{t\to\infty} \frac{1}{2qt} \log \mathbb{E}[e^{((L_{2q}\varepsilon_n)/c^{2q-1})N_t}] \\ &\le \liminf_{t\to\infty} \frac{1}{pp't} \log \mathbb{E}[e^{pp'\theta N_t}] + \liminf_{t\to\infty} \frac{1}{pq't} \log \mathbb{E}[e^{q'pL_1\varepsilon_n N_t}] \\ &\quad + \liminf_{t\to\infty} \frac{1}{2qt} \log \mathbb{E}[e^{((L_{2q}\varepsilon_n)/c^{2q-1})N_t}]. \end{aligned} \tag{6.27}$$

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

$$\Gamma(\theta) \le \liminf_{t\to\infty} \frac{1}{pp't} \log \mathbb{E}[e^{pp'\theta N_t}]. \tag{6.28}$$

This holds for any θ and thus

$$\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{E}[e^{\theta N_t}]\geq pp'\Gamma\left(\frac{\theta}{pp'}\right). \tag{6.29}$$

Letting $p, p'\downarrow 1$ and using the continuity of $\Gamma(\cdot)$, we get the desired result. □

THEOREM 20. *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 that $(N_t/t\in\cdot)$ satisfies the large deviation principle with the rate function*

$$I(x)=\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma(\theta)\}. \tag{6.30}$$

PROOF. For the upper bound, apply the Gärtner–Ellis theorem. Let us prove the lower bound. Let $B_{\varepsilon}(x)$ denote the open ball centered at x with radius $\varepsilon>0$. By Hölder's inequality, for any $p,q>1$ with $\frac{1}{p}+\frac{1}{q}=1$,

$$P_n\left(\frac{N_t}{t}\in B_{\varepsilon}(x)\right)\leq\left\|\frac{dP_n}{d\mathbb{P}}\right\|_{L^p(\mathbb{P})}\mathbb{P}\left(\frac{N_t}{t}\in B_{\varepsilon}(x)\right)^{1/q}. \tag{6.31}$$

Therefore, letting $t\to\infty$, we have

$$\begin{aligned}\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma_n(\theta)\}&=\lim_{t\to\infty}\frac{1}{t}\log P_n\left(\frac{N_t}{t}\in B_{\varepsilon}(x)\right)\\&\leq\frac{1}{pp'}\Gamma(pp'L_1\varepsilon_n)+\frac{1}{2pq'}\Gamma\left(\frac{L_{2pq'}\varepsilon_n}{c^{2pq'-1}}\right)\\&\quad+\frac{1}{q}\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t}\in B_{\varepsilon}(x)\right),\end{aligned} \tag{6.32}$$

where $\varepsilon_n=\|h_n-h\|_{L^1}$. Hence, letting $n\to\infty$, see that

$$\frac{1}{q}\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t}\in B_{\varepsilon}(x)\right)\geq\limsup_{n\to\infty}\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma_n(\theta)\}. \tag{6.33}$$

Since $\Gamma_n(\theta)\to\Gamma(\theta)$ uniformly on any compact set K,

$$\sup_{\theta\in K}\{\theta x-\Gamma_n(\theta)\}\to\sup_{\theta\in K}\{\theta x-\Gamma(\theta)\}, \tag{6.34}$$

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

$$\liminf_{\theta\to+\infty}\frac{\Gamma_n(\theta)}{\theta}\geq\liminf_{\theta\to+\infty}\frac{(e^{\theta}-1)c}{\theta}=+\infty, \tag{6.35}$$

which implies that $\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma_n(\theta)\}\to\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma(\theta)\}$. Therefore,

$$\frac{1}{q}\liminf_{t\to\infty}\frac{1}{t}\log\mathbb{P}\left(\frac{N_t}{t}\in B_\varepsilon(x)\right)\geq\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma(\theta)\}. \tag{6.36}$$

Letting $q\downarrow 1$, we get the desired result. □

REMARK 21. The class of nonlinear Hawkes processes 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. In fact, by the coupling argument we used in the proof of the case of linear $\lambda(\cdot)$ in Theorem 12, we can prove that in the case when $\lim_{z\to\infty}\frac{\lambda(z)}{z}=0$ and $\lambda(\cdot)$ is α-Lipshcitz and $\lambda(\cdot)\geq c>0$, $\Gamma(\theta)=\lim_{n\to\infty}\Gamma_n(\theta)$ for $\theta\leq\mu-1-\log\mu$, where $\mu=\int_0^\infty h(t)\,dt$ and Γ and Γ_n are the limit of logarithmic moment generating functions when the exciting functions are h and h_n, respectively, and $h_n\to h$ in L^1. For the linear case, since $\Gamma(\theta)=\infty$ for $\theta>\mu-1-\log\mu$, the coupling argument is good enough. However, for the sublinear $\lambda(\cdot)$, $\Gamma(\theta)<\infty$ for any θ and the coupling argument is not enough. In fact, it will appear in Zhu [25] that under the condition that $\lim_{z\to\infty}\frac{\lambda(z)}{z}=0$, $\lambda(\cdot)$ is positive, increasing, α-Lipshcitz and $\lambda(\cdot)\geq c>0$ and $h(\cdot)$ is positive, decreasing and $\int_0^\infty h(t)\,dt<\infty$, there is a level-3 large deviation principle from which we can use the contraction principle to get the level-1 large deviation principle for $(N_t/t\in\cdot)$. Therefore, we conjecture that in the sublinear case, $\Gamma(\theta)=\lim_{n\to\infty}\Gamma_n(\theta)$ for any θ and $(N_t/t\in\cdot)$ satisfies the large deviation principle with rate function $I(x)=\sup_{\theta\in\mathbb{R}}\{\theta x-\Gamma(\theta)\}$. The advantage of approximating the general case by the case when h is a sum of exponentials is that $\Gamma_n(\theta)$ can be evaluated by an optimization problem, which should be computable by some numerical scheme.

Acknowledgments. The author is enormously grateful to his advisor Professor S. R. S. Varadhan for suggesting this topic and for his superb guidance, understanding, patience and generosity. He would also like to thank his colleague Dmytro Karabash for valuable discussions on this project. The author would also thank the anonymous referees who provided very helpful suggestions for the improvement of this paper, and to whom the author is much indebted. The author also thanks an associate editor for helpful remarks. The author also thanks Professor Henry McKean for pointing out some typos and minor mistakes in the manuscript.

REFERENCES

[1] BACRY, E., DELATTRE, S., HOFFMANN, M. and MUZY, J. F. (2013). Some limit theorems for Hawkes processes and application to financial statistics. *Stochastic Process. Appl.* **123** 2475–2499. MR3054533

[2] BORDENAVE, C. and TORRISI, G. L. (2007). Large deviations of Poisson cluster processes. *Stoch. Models* **23** 593–625. MR2362700

[3] BRÉMAUD, P. and MASSOULIÉ, L. (1996). Stability of nonlinear Hawkes processes. *Ann. Probab.* **24** 1563–1588. MR1411506
[4] COX, D. R. and ISHAM, V. (1980). *Point Processes*. Chapman & Hall, London. MR0598033
[5] DALEY, D. J. and VERE-JONES, D. (1988). *An Introduction to the Theory of Point Processes*, 2nd ed. Springer, New York. MR0950166
[6] DAVIS, M. H. A. (1993). *Markov Models and Optimization*. Chapman & Hall, London. MR1283589
[7] DEMBO, A. and ZEITOUNI, O. (1998). *Large Deviations Techniques and Applications*, 2nd ed. Springer, New York. MR1619036
[8] ECHEVERRÍA, P. (1982). A criterion for invariant measures of Markov processes. *Probab. Theory Related Fields* **61** 1–16.
[9] FAN, K. (1953). Minimax theorems. *Proc. Natl. Acad. Sci. USA* **39** 42–47. MR0055678
[10] FRENK, J. B. G. and KASSAY, G. (2003). The level set method of Joó and its use in minimax theory. Technical Report E.I 2003-03, Econometric Institute, Erasmus Univ., Rotterdam.
[11] HAIRER, M. (2010). Convergence of Markov processes. *Lecture Notes*. Univ. Warwick. Available at http://www.hairer.org/notes/Convergence.pdf.
[12] HAWKES, A. G. (1971). Spectra of some self-exciting and mutually exciting point processes. *Biometrika* **58** 83–90. MR0278410
[13] HAWKES, A. G. and OAKES, D. (1974). A cluster process representation of a self-exciting process. *J. Appl. Probab.* **11** 493–503. MR0378093
[14] JAGERS, P. (1975). *Branching Processes with Biological Applications*. Wiley, London. MR0488341
[15] JOÓ, I. (1984). Note on my paper: "A simple proof for von Neumann's minimax theorem" [Acta Sci. Math. (Szeged) **42** (1980), no. 1-2, 91–94; MR0576940 (81i:49008)]. *Acta Math. Hungar.* **44** 363–365. MR0764631
[16] KARABASH, D. and ZHU, L. (2012). Limit theorems for marked Hawkes processes with application to a risk model. Preprint. Available at arXiv:1211.4039.
[17] KORALOV, L. B. and SINAI, Y. G. (2007). *Theory of Probability and Random Processes*, 2nd ed. Springer, Berlin. MR2343262
[18] LINIGER, T. (2009). Multivariate Hawkes processes. Ph.D. thesis, ETH, Zürich.
[19] LIPTSER, R. S. and SHIRYAEV, A. N. (2001). *Statistics of Random Processes. II*, 2nd ed. Springer, Berlin. MR1800858
[20] OAKES, D. (1975). The Markovian self-exciting process. *J. Appl. Probab.* **12** 69–77. MR0362522
[21] STABILE, G. and TORRISI, G. L. (2010). Risk processes with non-stationary Hawkes claims arrivals. *Methodol. Comput. Appl. Probab.* **12** 415–429. MR2665268
[22] VARADHAN, S. R. S. (2001). *Probability Theory*. Amer. Math. Soc., Providence, RI. MR1852999
[23] VARADHAN, S. R. S. (2008). Large deviations. *Ann. Probab.* **36** 397–419. MR2393987
[24] ZHU, L. (2014). Limit theorems for a Cox–Ingersoll–Ross process with Hawkes jumps. *J. Appl. Probab.* To appear.
[25] ZHU, L. (2014). Process-level large deviations for nonlinear Hawkes point processes. *Ann. Inst. Henri Poincaré Probab. Stat.* **50** 845–871. MR3224291
[26] ZHU, L. (2013). Nonlinear Hawkes processes. Ph.D. thesis, New York Univ.
[27] ZHU, L. (2013). Ruin probabilities for risk processes with non-stationary arrivals and subexponential claims. *Insurance Math. Econom.* **53** 544–550. MR3130449
[28] ZHU, L. (2013). Central limit theorem for nonlinear Hawkes processes. *J. Appl. Probab.* **50** 760–771. MR3102513

[29] ZHU, L. (2013). Moderate deviations for Hawkes processes. *Statist. Probab. Lett.* **83** 885–890. MR3040318

COURANT INSTITUTE OF MATHEMATICAL SCIENCES
NEW YORK UNIVERSITY
251 MERCER STREET
NEW YORK, NEW YORK 10012
USA
E-MAIL: ling@cims.nyu.edu