Large and moderate deviations for a discrete-time marked Hawkes process

Haixu Wang
Department of Mathematics, Florida State University, Tallahassee, Florida, USA

ABSTRACT
Hawkes process is a continuous-time stochastic model that captures temporal stochastic self-exciting phenomena. In particular, the linear Hawkes process has been well studied and widely used in practice because of its mathematical tractability. However, in some contexts, a Hawkes model is not directly applicable because data is recorded in a discrete-time scheme or an aggregated way. Thus, a discrete-time Hawkes model is appealing for applications. In this paper, we study large and moderate deviations for a discrete-time marked Hawkes process first proposed in Xu, Zhu, and Wang (2020).

ARTICLE HISTORY
Received 28 July 2021
Accepted 25 December 2021

KEYWORDS
Discrete-time; marked Hawkes process; self-exciting; univariate; large deviations; moderate deviations

1. Introduction

Hawkes process is a self-exciting simple point process named after Hawkes (1971). Hawkes processes originate from statistical literature to model the occurrences of earthquakes and shocks after earthquakes (Vere-Jones 1975). In contrast to a standard Poisson process, the intensity of Hawkes process depends on its entire history, which can model the self-exciting or clustering effect. In finance, most applications of Hawkes processes are about high-frequency trading (Bauwens and Hautsch 2009; Chavez-Demoulin and McGill 2012). Furthermore, Hawkes processes have been used to model credit default and the arrival of company defaults in a bond portfolio (Egami, Kato, and Sawaki 2013; Giesecke, Goldberg, and Ding 2011). Recently, Hawkes models have been applied in social networks. For example, Fox et al. (2016) modeled the rate of sending email for each officer at the West Point Military Academy. The more applications of Hawkes process can be found in seismology, neuroscience, cosmology, ecology, and epidemiology. For a list of references for these applications, see Bordenave and Torrisi (2007), Zhu (2013c), and Liniger (2009).

Next, let us introduce the Hawkes process. Let \( N \) be a simple point process on \( \mathbb{R} \) and \( \mathcal{F}_t^\infty := \sigma(N(C), C \in \mathcal{B}(\mathbb{R}), C \in (-\infty, t]) \) be an increasing family of \( \sigma \)-algebras. Any non-negative \( \mathcal{F}_t^\infty \)-progressively measurable process \( \lambda_t \) is called the \( \mathcal{F}_t^\infty \)-intensity of \( N \) if

\[
\mathbb{E}[N(a, b) | \mathcal{F}_a^{-\infty}] = \mathbb{E}\left[ \int_a^b \lambda_s ds | \mathcal{F}_a^{-\infty} \right]
\]
a.s. for all interval \((a, b]\). \(N_t := N(0, t]\) denotes the number of points in the interval \((0, t]\). Hawkes processes is a simple point process with \(\mathcal{F}_t\)-intensity

\[
\lambda_t := \lambda \left( \int_{(-\infty, t]} h(t - s) N(ds) \right),
\]

where kernel function \(h(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+\) is integrable and \(|h|_{L^1} = \int_0^\infty h(t) dt < \infty\) and \(\lambda(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+\) is local integrable. A variant of Hawkes process is marked Hawkes process, which is a Hawkes process with random marks. The marked Hawkes process is a simple point process with \(\mathcal{F}_t\)-intensity

\[
\lambda_t := \lambda \left( \int_{(-\infty, t] \times \mathbb{X}} h(t - s, \ell) N(ds, d\ell) \right),
\]

where \(\lambda(\cdot) : \mathbb{R}^+ \to \mathbb{R}^+\) is locally integrable and left continuous, \(h(\cdot, \cdot) : \mathbb{R}^+ \times \mathbb{X} \to \mathbb{R}^+\) is integrable, \(\ell\) denotes the mark variable, and \(|h|_{L^1} = \int_0^\infty \int_{\mathbb{X}} h(t, \ell) q(d\ell) dt < \infty\). Here \(\mathbb{X}\) is measurable space with common law \(q(d\ell)\). \(h(\cdot, \cdot)\) and \(\lambda(\cdot, \cdot)\) are referred as exciting function and rate function, respectively. Local integrability assumption of \(\lambda(\cdot)\) ensures that the process is non-explosive and left continuity assumption ensures that \(\lambda_t\) is \(\mathcal{F}_t\)-predictable. The integral in Equation (1) stands for \(\int_{(-\infty, t] \times \mathbb{X}} h(t - s, \ell) N(ds, d\ell) = \sum_{\tau_i < t} \sum_{j \leq i} h(t - \tau_i, \ell_i)\), where \((\tau_i)_{i \geq 1}\) are the occurrences of the points before time \(t\), and the \((\ell_i)_{i \geq 1}\) are i.i.d. random marks, \(\ell_i\) being independent of previous arrival times \(\tau_j, j \leq i\).

When \(\lambda(\cdot)\) is linear, it is called a linear Hawkes process. There were extensive studies on the stability, law of large numbers, central limit theorems, large deviations, Bartlett spectrum, etc. In particular, Bacry et al. (2013) proved the functional law of large numbers and the functional central limit theorems for unmarked Hawkes process. Karabash and Zhu (2015) studied the limit theorems of linear marked Hawkes process. Horst and Xu (2021) studied the functional law of large number and central limit theorem of marked Hawkes process with homogeneous immigration. Bordenave and Torrisi (2007) derived large deviations of Hawkes process. For a survey on Hawkes processes and related self-exciting processes, Poisson cluster processes, affine point processes, etc., see Daley and Vere-Jones (2003).

When \(\lambda(\cdot)\) is nonlinear, it is known as a nonlinear Hawkes process. Because of the lack of immigration-birth representation and computational tractability, nonlinear Hawkes processes are much less studied. However, there were some efforts in this direction. A nonlinear Hawkes process was first introduced by Brémaud and Massoulié (1996). The central limit theorems, the large deviation principles for nonlinear Hawkes processes can be found in Zhu (2015, 2013a, 2013b, 2014b).

Hawkes process can also be extended to the multivariate setting. For a survey of multivariate processes and a short history of Hawkes process, we refer to Liniger (2009).

Before we proceed, we will briefly review the other related literature, the large deviation principle and the moderate deviation principle for Hawkes models.
1.1. Large deviation principles

Following Dembo and Zeitouni (1998), we introduce the definition of large deviation principle. A family of probability measures \( \{P_n\}_{n \in \mathbb{N}} \) on a topological space \((X,T)\) satisfies the large deviation principle with rate function \( I(\cdot) : X \to [0, \infty) \) and speed \( a_n \) if \( I \) is a lower semi-continuous function, \( a_n : [0, \infty) \to [0, \infty) \) is a measurable function which increases to infinity, and the following inequalities hold for every Borel set \( A \):

\[
- \inf_{x \in A^c} I(x) \leq \lim_{n \to \infty} \inf \frac{1}{a_n} \log P_n(A) \leq \lim_{n \to \infty} \sup \frac{1}{a_n} \log P_n(A) \leq - \inf_{x \in \bar{A}} I(x),
\]

where \( A^c \) is the interior of \( A \) and \( \bar{A} \) is the closure of \( A \). We say that the rate function \( I \) is good if for any \( m \geq 0 \), the level set \( \{x \in X : I(x) \leq m, \ m \geq 0\} \) is compact. In addition to Dembo and Zeitouni (1998), we also refer to Varadhan (1984) for a survey on large deviations.

1.2. Large deviations for Hawkes processes

We first review some large deviations results for Hawkes processes in the literature. We recall that the intensity of a unmarked linear Hawkes process with empty past history, i.e., \( N_{(-\infty,0]} = 0 \), is given by

\[
\lambda_t := \nu + \int_{(0,t)} h(t-s)N(ds)
\]

(2)

where \( \nu > 0 \). The integral in Equation (2) stands for \( \int_{(0,t)} h(t-s)N(ds) = \sum_{\tau_i < t} h(t-\tau_i) \), where \( (\tau_i)_{i \geq 1} \) are the occurrences of the points before time \( t \). If \( ||h||_{L^1} = \int_0^\infty h(t)dt < 1 \), the linear Hawkes process has an immigration-birth representation, and by ergodic theory, the law of large numbers for the linear Hawkes process (see, for instance, Daley and Vere-Jones [2003]) is derived as

\[
\lim_{t \to \infty} \frac{N_t}{t} = \frac{\nu}{1 - ||h||_{L^1}}.
\]

Bordenave and Torrisi (2007) showed that, if \( 0 < ||h||_{L^1} < 1 \) and \( \int_0^\infty th(t)dt < \infty \), then \( \mathbb{P}\left(\frac{N_t}{t} \in \cdot \right) \) satisfies the large deviation principle on \( \mathbb{R} \) with the good rate function:

\[
I(x) = \begin{cases} 
\bar{\theta}x + \nu - \frac{\nu x}{\nu + ||h||_{L^1}x}, & \text{if } x \in (0, \infty), \\
\nu, & \text{if } x = 0, \\
+\infty, & \text{otherwise},
\end{cases}
\]

(3)

where \( \bar{\theta} = \tilde{\theta}_x \) is the unique solution in \( (-\infty, ||h||_{L^1}) \) of \( \mathbb{E}[e^{\tilde{\theta}S}] = \frac{x}{\nu + x||h||_{L^1}}, \ x > 0 \), where \( S \) denotes the total number of descendants of an immigrant, including the immigrant itself.

The large deviation principle of a marked linear Hawkes process with empty history can be found in Karabash and Zhu (2015). Recall the notation of a general marked Hawkes process introduced in Section 1. The intensity of a marked univariate linear Hawkes process is given by
\[ \lambda_t := \nu + \int_{(0,t) \times \mathbb{X}} h(t - s, \ell) N(ds, d\ell). \]  

(4)

Let \( H(\ell) := \int_0^\infty h(t, \ell) dt \) for any \( \ell \in \mathbb{X} \). Assume that

\[ \int_\mathbb{X} H(\ell) q(d\ell) < 1. \]  

(5)

Under the above assumption, there exists an unique stationary version of the linear marked Hawkes process defined by Equation (4). And by ergodic theorem, a law of large numbers hold:

\[ \lim_{t \to \infty} \frac{N_t}{t} = \frac{\nu}{1 - \mathbb{E}[H(\ell)]}. \]

If there exists some \( \tilde{\theta} > 0 \), so that \( \int_\mathbb{X} e^{\tilde{\theta} H(\ell)} q(d\ell) < \infty \). Karabash and Zhu (2015) proved that \( P(N_t/t \to \cdot) \) satisfies a large deviation principle with rate function:

\[ I(x) = \begin{cases} \tilde{\theta}x - \nu(x - 1), & x \geq 0, \\ +\infty, & x < 0, \end{cases} \]  

(6)

where \( \tilde{\theta} \) and \( x \) satisfy the following equations

\[
\begin{align*}
\tilde{\theta}x &= \mathbb{E}[e^{\tilde{\theta}x - (x - 1)H(\ell)}], \\
\frac{x}{\nu} &= x + \frac{x}{\nu} \mathbb{E}[H(\ell)e^{\tilde{\theta}x - (x - 1)H(\ell)}].
\end{align*}
\]  

(7)

For nonlinear Hawkes processes, Zhu (2014b) established the level-3 large deviation principle first and then used the contraction principle to obtain the large deviation principle for \( P(N_t/t \to \cdot) \). Zhu (2015) proved the large deviations for Markovian Hawkes processes and generalized the proof to the case when \( h(\cdot) \) is a sum of exponentials starting with the case of exponential \( h(\cdot) \).

1.3. Moderate deviation principles

For any \( \sqrt{n} \ll c_n \ll n \), a family of probability measures \( \{P_n\}_{n \in \mathbb{N}} \) on a topological space \((X, \mathcal{T})\) satisfies a moderate deviation principle with rate function \( J(\cdot) : X \to [0, \infty) \) if \( J \) is a lower semi-continuous function and for any Borel set \( A \)

\[
-\inf_{x \in A^o} J(x) \leq \liminf_{n \to \infty} \frac{1}{c_n^2} \log P_n(A) \leq \limsup_{n \to \infty} \frac{1}{c_n^2} \log P_n(A) \leq -\inf_{x \in \bar{A}} J(x). 
\]

That is, \( P_n \) satisfies a large deviation principle with speed \( c_n^2 \). For example, let \( X_1, \ldots, X_n \) be a sequence of i.i.d random variables commonly distributed as \( X \) and assume \( \mathbb{E}[e^{\tilde{\theta}X}] < \infty \) for \( \tilde{\theta} \) in some ball around the origin. Then, \( P := P\left( \frac{1}{c_n} \sum_{i=1}^n X_i \in \cdot \right) \) satisfies a large deviation principle with speed \( c_n^2 \). Moderate deviations fill the gap between ordinary deviations approximated by the central limit theorem and large deviations.
1.4. Moderate deviations for Hawkes processes

Zhu (2013b) proved the moderate deviation principle for a univariate linear Hawkes process, defined by Equation (2) in Section 1.2. With the assumption \( \sup_{t>0} t^{3/2} h(t) = C < \infty \), the moderate deviation principle holds with the rate function

\[
J(x) = \frac{x^2(1 - \|h\|_1)^3}{2\nu}.
\]

(8)

The moderate deviation principles for a marked linear Hawkes process was studied in Seol (2017). Recall the definition of a marked linear Hawkes process in Section 1.2. Seol (2017) showed the moderate deviation rate function is

\[
J(x) = \frac{x^2(1 - \mathbb{E}[H(\ell)])^3}{2\nu(1 + \text{Var}[H(\ell)])},
\]

(9)

with assumptions, \( \text{Var}[H(\ell)] < \infty \) and \( \sup_{t>0} t^{3/2} \int h(t, \ell) q(d\ell) \leq C < \infty \).

The other related literature. The large deviations of Cox-Ingersoll-Ross process with Hawkes jumps can be found in Zhu (2014a). Zhang et al. (2015) studied limit theorems of affine jump diffusion processes with Hawkes jumps. Gao and Zhu (2018) studied large deviations of the Hawkes process with large initial intensity and also discussed the applications of the model to insurance and queue systems. Yao (2018) studied the moderate deviation principle for multivariate unmarked linear Hawkes processes. And moderate deviation principles have been studied in mixing processes, Markov processes, martingales, etc. (see Gao [1996], Chen [2001], and Dembo [1996]).

1.5. Discrete-time Hawkes process

In contrast to the continuous setting, in reality, the arrivals of events are often recorded in a discrete-time scheme. For example, the data is collected on a fixed phase or the data only shows the aggregate results. Continuous-time Hawkes processes can model the unevenly spaced the arrival of events in time, while modeling the evenly spaced events in time requires a discrete-time type model. Therefore, discrete-time Hawkes processes are appealing for certain applications. However, there are few works on discrete-time Hawkes type models.

Xu, Zhu, and Wang (2020) proposed for the first time a discrete-time self-exciting and mutually-exciting model analogous to Hawkes process. More recently, the discrete-time self-exciting model was also applied to study the infection and death of COVID-19 in Browning et al. (2021). Wang (2020) extended the model of Xu, Zhu, and Wang (2020) in the univariate case and studied its limit theorems. Following the model in Wang (2020), for \( t \in \mathbb{N} \) let \( \alpha(t) : \mathbb{N} \to \mathbb{R}_+ \) be a positive function on \( \mathbb{N} \) and define \( X_0 = N_0 = 0 \). We define \( ||x||_1 := \sum_{t=1}^{\infty} x(t) \) as the \( \ell_1 \) norm of \( x \). Conditional on \( X_{t-1}, X_{t-2}, \ldots, X_1 \), we define \( Z_t \) as a Poisson random variable with mean

\[
\lambda_t = \nu + \sum_{s=1}^{t-1} \alpha(s) X_{t-s},
\]

(10)
and define
\[ X_t = \sum_{j=1}^{Z_t} \ell_{t,j}, \]
where \( \ell_{t,j} \) are positive random variables that are i.i.d. in both \( t \) and \( j \) with common law \( q(d\ell) \). And \( \ell_{t,j} \) are independent of \( Z_t \) and \( X_{t-1}, X_{t-2}, \ldots, X_1 \). Finally, we define \( N_t := \sum_{s=1}^{t} Z_s \) and \( L_t := \sum_{j=1}^{t} X_j \).

With the assumption that \( ||x||_1 E[\ell_{1,1}] := \sum_{i=1}^{\infty} x(t) \int_{\mathcal{X}} \ell_{1,1} q(d\ell) < 1 \), it can be derived that the law of large numbers hold (Wang 2020):
\[ \lim_{t \to \infty} \frac{N_t}{t} = \mu := \frac{\nu}{1 - ||x||_1 E[\ell_{1,1}]}, \quad \lim_{t \to \infty} \frac{L_t}{t} = \bar{\mu} := \frac{\nu E[\ell_{1,1}]}{1 - ||x||_1 E[\ell_{1,1}]}, \]
in probability as \( t \to \infty \), and the central limit theorem also holds, see Wang (2020):
\[ \frac{1}{\sqrt{t}} \left( N_t - \frac{\nu t}{1 - ||x||_1 E[\ell_{1,1}]} \right) \to \mathcal{N} \left( 0, \frac{\nu(1 + ||x||_1^2 \text{Var}(\ell_{1,1}))}{(1 - ||x||_1 E[\ell_{1,1}])^3} \right), \]
\[ \frac{1}{\sqrt{t}} \left( L_t - \frac{\nu E[\ell_{1,1}] t}{1 - ||x||_1 E[\ell_{1,1}]} \right) \to \mathcal{N} \left( 0, \frac{\nu E[\ell_{1,1}^2]}{(1 - ||x||_1 E[\ell_{1,1}])^3} \right), \]
in distribution as \( t \to \infty \) under the additional assumptions that

1. \( \lim_{t \to \infty} \frac{1}{\sqrt{t}} \sum_{u=1}^{t-1} \sum_{s=1+u}^{\infty} x(s) = 0 \),
2. The first four moments of \( \ell_{1,1} \) are finite.

**Other related literature.** A discrete-time Hawkes-type model with 0-1 arrivals was proposed by Seol (2015) and the limit theorems were studied. Let \( (X_n)_{n=1}^{\infty} \) be a sequence taking values on \( \{0, 1\} \) defined as follows. Let \( \hat{N} = \mathbb{N} \cup \{0\} \) and assume that for \( i \in \hat{N} \), \( x_i > 0 \) is a given sequence of positive numbers and \( \sum_{i=0}^{\infty} x_i < 1 \). (i) \( X_1 = 1 \) with probability \( x_0 \) and \( X_1 = 0 \) otherwise. (ii) Conditional on \( X_1, X_2, \ldots, X_{n-1} \), we have \( X_n = 1 \) with probability \( x_0 + \sum_{i=1}^{n-1} x_{n-i} X_i \), and \( X_n = 0 \) otherwise. Define \( S_n := \sum_{i=1}^{n} X_i \). Seol (2015) showed a law of large numbers theorem, i.e.,
\[ \frac{S_n}{n} \to \mu := \frac{x_0}{1 - \sum_{i=1}^{\infty} x_i}, \]
in probability as \( n \to \infty \). In addition, with assumption \( \sqrt{n} \sum_{i=1}^{n} x_i \to 0 \) as \( n \to \infty \) and \( \frac{1}{\sqrt{n}} \sum_{i=1}^{n} i x_i \to 0 \) as \( n \to \infty \), the central limit theorem follows:
\[ \frac{S_n - \mu n}{\sqrt{n}} \to \mathcal{N} \left( 0, \frac{\mu(1-\mu)}{(1 - \sum_{i=1}^{\infty} x_i)^2} \right), \]
in distribution as \( n \to \infty \).
Organization of this paper. The rest of the paper is organized as follows. In Section 2, we state our main results. The proof of the main results can be found in Section 3.

2. Main results

Recall the discrete-time Hawkes model introduced in Section 1, denote $t$ as time index, $t \in \mathbb{N}$ and $N_t := \sum_{s=1}^t Z_s$, $L_t := \sum_{s=1}^t X_s$ where $Z_t$ is a Poisson random variable conditional on $\mathcal{F}_{t-1}$ with intensity $\lambda_t$ defined by Equation (10) and $X_t$ is a compound Poisson random variable defined by Equation (11). We assume that

I. $||Z||_1 \mathbb{E}[\ell_{1,1}] < 1$.
II. There exists some $\theta > 0$ such that $\mathbb{E}[e^{\theta \ell_{1,1}}] < \infty$.

This section states the large deviations and moderate deviations of the discrete-time marked Hawkes process.

2.1. Large deviations

The formal definition of the large deviation principle has been introduced in Section 1.1. For the discrete-time Hawkes process, we prove the following large deviation principles.

Theorem 2.1. Assume the condition (I) and (II). $\mathbb{P}(N_t/t \in \cdot)$ satisfies a large deviation principle with the rate function

$$I(x) = \sup_{\theta \leq \theta_c} \{ \theta x - \Gamma(\theta) \},$$

where $\Gamma(\theta) := \nu(g(\theta) - 1)$, where $g(\theta)$ is the minimal solution to the equation $x = \mathbb{E}[e^{\theta ||x||_1 \ell_{1,1} + ||x||_1(x-1)}]$ for any $\theta \leq \theta_c$. Here $\theta_c = -\log \mathbb{E}[||x||_1 \ell_{1,1} e^{||x||_1(x-1)}]$, and $x_c > 1$ is the unique solution of the equation $x_c \mathbb{E}[||x||_1 \ell_{1,1} e^{||x||_1(x-1)}] = \mathbb{E}[e^{||x||_1(x-1)}]$.

Theorem 2.2. Assume the condition (I) and (II). $\mathbb{P}(L_t/t \in \cdot)$ satisfies a large deviation principle with the rate function

$$I_L(x) = \sup_{\tilde{\theta} \leq \tilde{\theta}_c} \{ \tilde{\theta} x - \Gamma_L(\tilde{\theta}) \},$$

where $\Gamma_L(\tilde{\theta}) := \nu(g_L(\tilde{\theta}) - 1)$, and $g_L(\tilde{\theta})$ is the minimal solution to the equation $x = \mathbb{E}[e^{\tilde{\theta} \ell_{1,1} + ||x||_1(x-1)}]$ for any $\tilde{\theta} \leq \tilde{\theta}_c$. Here $\tilde{\theta}_c > 0$ and $\tilde{x}_c > 1$ satisfy the following equations

$$\left\{ \begin{array}{l}
\mathbb{E}[||x||_1 \ell_{1,1} e^{\tilde{\theta} \ell_{1,1} + ||x||_1(x-1)}] = 1, \\
\mathbb{E}[e^{\tilde{\theta} \ell_{1,1} + ||x||_1(x-1)}] = \tilde{x}_c.
\end{array} \right.$$
2.2. Moderate deviations

In terms of the moderate deviations of the discrete-time Hawkes process, we assume sup_{t>0} t^{3/2} z(t) = C < \infty. Recall the Equation (12), where \( \mu \) and \( \bar{\mu} \) denote the limits in the law of large numbers for \( N_t \) and \( L_t \) respectively. We obtain the following moderate deviation principles for the discrete-time Hawkes process.

**Theorem 2.3.** Assume the condition (I) and (II). Furthermore, we assume sup_{t>0} t^{3/2} z(t) = C < \infty. For any Borel set \( A \) and time sequence \( c(t) \) such that \( \sqrt{t} \ll c(t) \ll t \), we have the following moderate deviation principle.

\[
- \inf_{x \in A} J(x) \leq \liminf_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P} \left( \frac{N_t - \mu t}{c(t)} \in A \right) \leq \limsup_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P} \left( \frac{N_t - \mu t}{c(t)} \in A \right) \leq - \inf_{x \in A} J(x),
\]

where \( \mu \) is defined by Equation (12) and

\[
J(x) = \frac{x^2 \left( 1 - \mathbb{E} [\ell_{1,1}||x||_1] \right)^3}{2\nu \left( 1 + \text{Var}(\ell_{1,1})||x||_1^2 \right)},
\]

(17)

**Theorem 2.4.** Assume the condition (I) and (II). Furthermore, we assume sup_{t>0} t^{3/2} z(t) = C < \infty. For any Borel set \( A \) and time sequence \( c(t) \) such that \( \sqrt{t} \ll c(t) \ll t \), we have the following moderate deviation principle.

\[
- \inf_{x \in A^c} J(x) \leq \liminf_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P} \left( \frac{L_t - \bar{\mu} t}{c(t)} \in A \right) \leq \limsup_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P} \left( \frac{L_t - \bar{\mu} t}{c(t)} \in A \right) \leq - \inf_{x \in A^c} J(x),
\]

where \( \bar{\mu} \) is defined by Equation (12) and

\[
J(x) = \frac{x^2 \left( 1 - \mathbb{E} [\ell_{1,1}||x||_1] \right)^3}{2\nu \mathbb{E} \left[ \ell_{1,1}^2 \right]}.
\]

(18)

3. Proof of main results

This section states the proof of our main results. Before we proceed, let us recall a version of Gärtner-Ellis theorem which will be used in our proof.

**Theorem 3.1** (Gärtner-Ellis theorem (Theorem 2.3.6 (Dembo and Zeitouni [1998]))). Let \( (\mu_n)_{n \in \mathbb{N}} \) be a sequence of probability measures on \( (\mathbb{R}, \mathcal{B}(\mathbb{R})) \). Define the logarithmic moment generating function

\[
\Lambda_n(\theta) := \int_{\mathbb{R}} e^{\theta x} d\mu_n,
\]

(19)

and assume that for all \( \theta \in \mathbb{R} \) a possibly infinite limit \( \Lambda(\theta) \) in (20)

\[
\Lambda(\theta) := \lim_{n \to \infty} \frac{1}{n} \log \left[ \Lambda_n(n\theta) \right],
\]

(20)
exists and \(0 \in \mathcal{D}_\Lambda^0\), where \(\mathcal{D}_\Lambda^0\) is the interior of \(\mathcal{D}_\Lambda\) and \(\mathcal{D}_\Lambda := \{\theta \in \mathbb{R} : \Lambda(\theta) < \infty\}\). Suppose in addition that \(\Lambda\) is lower semi-continuous on \(\mathbb{R}\), differentiable on \(\mathcal{D}_\Lambda^0\), and \(\Lambda\) is steep, i.e.,

\[
\lim_{n \to \infty} |\Lambda'(\theta_n)| = \infty
\]

whenever \(\theta_n \in \mathcal{D}_\Lambda^0\), \(\theta_n \to \theta \in \partial \mathcal{D}_\Lambda^0\) as \(n \to \infty\). Then \((\mu_n)_{n \in \mathbb{N}}\) satisfies the LDP with rate function \(I\), which is Fenchel-Legendre transform of \(\Lambda\),

\[
I(x) = \sup_{\theta \in \mathbb{R}} \{\theta x - \Lambda(\theta)\}. \tag{21}
\]

### 3.1. Proof of large deviations

#### Proof of Theorem 2.1

For any \(\theta \in \mathbb{R}\), we can compute that

\[
\mathbb{E}[e^{\theta N_t}] = \mathbb{E}[e^{\theta N_{t-1} + \theta Z_t}]
\]

\[
= \mathbb{E}[\mathbb{E}[e^{\theta N_{t-1} + \theta Z_t} | \mathcal{F}_{t-1}]]
\]

\[
= \mathbb{E}[e^{\theta N_{t-1}} \mathbb{E}[e^{\theta Z_t} | \mathcal{F}_{t-1}]]
\]

\[
= \mathbb{E}[e^{\theta N_{t-1} + (\theta - 1)\lambda_t}],
\]

where we used the fact that \(Z_t\) is Poisson with parameter \(\lambda_t\), conditional on \(\mathcal{F}_{t-1}\), the natural filtration up to time \(t - 1\). By the definition of \(\lambda_t\), we have

\[
\mathbb{E}[e^{\theta N_t}] = \mathbb{E}[e^{(\theta - 1)\lambda_t + (\theta - 1)^{t-1} \sum_{s=1}^{t-1} \lambda(s)X_{t-s}}]
\]

\[
= e^{(\theta - 1)\lambda_t} \mathbb{E}[e^{(\theta - 1)^{t-1} \sum_{s=1}^{t-1} \lambda(s)X_{t-s}}]
\]

\[
= e^{(\theta - 1)\lambda_t} \mathbb{E}[e^{(\theta - 1)\lambda(1)X_{t-1} + (\theta - 1)^{t-1} \sum_{s=2}^{t-1} \lambda(s)X_{t-s}}].
\]

Let \(f_0(\theta) = \theta\) and \(f_1(\theta) = \theta + \log \mathbb{E}[e^{(\theta - 1)\lambda(1)X_{t-1}}].\) Then,

\[
\mathbb{E}[e^{\theta N_t}] = e^{(\theta - 1)\lambda_t} \mathbb{E}[e^{\theta N_{t-2} + f_1(\theta)X_{t-1} + (\theta - 1)^{t-1} \sum_{s=2}^{t-1} \lambda(s)X_{t-s}}]
\]

\[
= e^{(\theta - 1)\lambda_t} \mathbb{E}[e^{\theta N_{t-2} + (\theta - 1)\lambda_{t-1} + (\theta - 1)^{t-1} \sum_{s=2}^{t-1} \lambda(s)X_{t-s}}].
\]
By the definition of $\lambda_{t-1}$, we get
\[
\mathbb{E}[e^{\theta N_t}]
= e^{(\theta-1)\nu} \mathbb{E}
\left[
\theta N_{t-2} + (e^{\theta(1)} - 1)(\nu + \sum_{s=1}^{t-2} \lambda(s) X_{t-1-s}) + (e^{\theta-1} - 1) \sum_{s=1}^{t-1} \lambda(s) X_{t-s}
\right]
= e^{(\theta-1)\nu + (e^{\theta(1)} - 1)\nu} \mathbb{E}
\left[
\theta N_{t-2} + (e^{\theta(1)} - 1)\nu + \sum_{s=1}^{t-2} \lambda(s) X_{t-1-s} + (e^{\theta-1} - 1) \sum_{s=1}^{t-1} \lambda(s) X_{t-s}
\right]
= e^{(\theta-1)\nu + (e^{\theta(1)} - 1)\nu} \mathbb{E}
\left[
\theta N_{t-2} + \log \mathbb{E}[e^{(e^{\theta(1)} - 1)\nu + (e^{\theta-1} - 1)\nu} \lambda_{t-1} Z_{t-2}]
\right]
= e^{(\theta-1)\nu + (e^{\theta(1)} - 1)\nu} \mathbb{E}
\left[
\theta N_{t-2} + \log \mathbb{E}[e^{(e^{\theta(1)} - 1)\nu + (e^{\theta-1} - 1)\nu} \lambda_{t-1} Z_{t-2}]
\right].
\]

By induction on $t$, we get
\[
\mathbb{E}[e^{\theta N_t}] = e^{(\theta-1)\nu + (e^{\theta(1)} - 1)\nu + \ldots + (e^{\theta-t+1} - 1)\nu} \mathbb{E}[e^{(\theta-1)\nu}]
= e^{\theta\nu + (e^{\theta(1)} - 1)\nu + \ldots + (e^{\theta-t+1} - 1)\nu} e^{(\theta-1)\nu - 1},
\]
where $f_0(\theta) = \theta$, $f_1(\theta) = \theta + \log \mathbb{E}[e^{(e^{\theta(1)} - 1)\nu} \lambda_{t,1}]$, and
\[
f_2(\theta) = \theta + \log \mathbb{E}[e^{(e^{\theta(1)} - 1)\nu + (e^{\theta-1} - 1)\nu} \lambda_{t,1}],
\]
and more generally, for every $s \geq 1$,
\[
f_s(\theta) = \theta + \log \mathbb{E}[e^{(e^{\theta(1)} - 1)\nu + \ldots + (e^{\theta-s+1} - 1)\nu} \lambda_{t,1}],
\]
(22)

This implies that
\[
\lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \nu(e^{\theta\nu} - 1),
\]
(24)

where
\[
f_\infty(\theta) = \theta + \log \mathbb{E}[e^{(e^{\theta(1)} - 1)\nu} \lambda_{t,1}],
\]
(25)

Let $x = e^{\theta\nu}$. Thus, Equation (25) can be rewritten as
\[
x = \mathbb{E}[e^{\theta + (x-1)\nu} \lambda_{t,1}],
\]
(26)

When $\theta \leq 0$, $e^{\theta(1)}(x)$ is decreasing in $t$ and $0 < e^{\theta(1)} \leq 1$. Thus, $e^{\theta(1)}(x)$ converges to a finite limit $x$ as $t \to \infty$, which satisfies Equation (26).

When $\theta > 0$, $f_s(\theta)$ is increasing in $t$. Either $f_s(\theta)$ converges to infinity as $t \to \infty$, in which case $\lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \infty$, or $e^{\theta(1)}(x)$ converges to a finite limit $x$ as $t \to \infty$. 

which satisfies Equation (26). Next, we need to determine for what values of $\theta$ the solution of Equation (26) exists. Let

$$G(x) = \mathbb{E}[e^{\theta(x-1)}\|x\|_1^\ell_{1,1}] - x. \quad (27)$$

It is easy to see that $G(x)$ is increasing in $\theta$ and $G''(x) > 0$. If $\theta = 0$, then $G(x) = \mathbb{E}[e^{(x-1)}\|x\|_1^\ell_{1,1}] - x$ satisfies $G(1) = 0$. Moreover, $G'(1) = \mathbb{E}[\|x\|_1^\ell_{1,1}] - 1$. By the assumption (I), we have $G'(1) < 0$. It implies $\min_{x>1} G(x) < 0$. Hence, there exists some critical $\theta_c > 0$ such that $\min_{x>1} G(x) = 0$. In other words, with $\theta_c$, we can find critical value $x_c$ such that $G(x_c) = G'(x_c) = 0$. By $G'(x_c) = 0$, we can find

$$\theta_c = -\log \mathbb{E}[\|x\|_1^\ell_{1,1} e^{(x-1)}\|x\|_1^\ell_{1,1}]$$

and substitute $\theta_c$ into $G(x_c) = 0$, we can find $x_c > 1$ satisfies

$$x_c \mathbb{E}[\|x\|_1^\ell_{1,1} e^{(x-1)}\|x\|_1^\ell_{1,1}] = \mathbb{E}[e^{(x-1)}\|x\|_1^\ell_{1,1}] . \quad (29)$$

Now let

$$H(x_c) = x_c \mathbb{E}[\|x\|_1^\ell_{1,1} e^{(x-1)}\|x\|_1^\ell_{1,1}] - \mathbb{E}[e^{(x-1)}\|x\|_1^\ell_{1,1}] . \quad (30)$$

We can compute that $H'(x_c) = \mathbb{E}[\|x\|_1^\ell_{1,1} e^{(x-1)}\|x\|_1^\ell_{1,1}] > 0$. In other words, $H(x_c)$ is increasing function of $x_c$ and has a unique solution such that $H(x_c) = 0$. Note that $\theta_c$ is also unique. Therefore, Equation (26) has finite solutions if and only if $\theta \leq \theta_c$.

$G(x)$ is strictly convex in $x$. Hence, there are at most two solutions for Equation (26). As discussed above, when $\theta = \theta_c$, $e^{\ell_{1,0}(0)}$ converges to a finite limit $x_c$, which is the unique solution of Equation (26). Indeed, Equation (26) has two repeated solutions in this case. When $0 < \theta < \theta_c$, Equation (26) has two different solutions. It is not hard to check $G(1) = e^0 - 1 > 0$ and $G'(1) = \mathbb{E}[\|x\|_1^\ell_{1,1} e^0] - 1 < 0$. $f_t(\theta)$ is increasing in $t$ and for $t = 0$, $e^{\ell_{1,0}(0)} = e^0 > 1$ and

$$G(e^0) = e^0 \left[ \mathbb{E}[e^{(e^0-1)}\|x\|_1^\ell_{1,1}] - 1 \right] > 0 . \quad (31)$$

Thus, as $t \to \infty$, $e^{\ell_{1,0}(0)}$ converges to a finite limit $x$. Overall, when $\theta \leq \theta_c$, $e^{\ell_{1,0}(0)}$ converges to the smaller solution of Equation (26) as $t \to \infty$.

When $\theta > \theta_c$, Equation (26) does not have solution, i.e., $f_t(\theta)$ does not converge to a finite value as $t \to \infty$. And $f_t(\theta)$ is monotonic in $t$. Hence, $f_t(\theta)$ must converge to infinity as $t \to \infty$, i.e., $\lim_{t \to \infty} \frac{1}{t} \log \mathbb{E}[e^{\theta N_t}] = \infty$.

When $\theta < 0$, we can check $G(1) < 0$ and $G'(1) < 0$. $f_t(\theta)$ is decreasing in $t$ and at $t = 0$, $e^{\ell_{1,0}(0)} = e^0 < 1$ with $G(e^0) < 0$. Thus, as $t \to \infty$, $e^{\ell_{1,0}(0)}$ converges to a finite limit. It must converge to the finite limit $x$ which is also the smaller solution of Equation (26).
Finally, we need to check the essential smoothness condition of $\nu(e^{\tilde{\ell}_s(\theta)} - 1)$.

$$f'_\infty(\theta) = \frac{\mathbb{E}[e^{(\tilde{\ell}_s(\theta) - 1) ||x||_1^2 \ell_{1,1}}]}{\mathbb{E}[e^{(\tilde{\ell}_s(\theta) - 1) ||x||_1^2 \ell_{1,1}}] - \nu e^{\tilde{\ell}_s(\theta)}} \mathbb{E}[||x||_1^2 \ell_{1,1} e^{(\tilde{\ell}_s(\theta) - 1) ||x||_1^2 \ell_{1,1}}]$$  \hspace{1cm} (32)

By Equations (28) and (29), it is not hard to find $|f'_\infty(\theta)| \to \infty$ as $\theta \to \theta_c$, the conclusion then follows Gärtner-Ellis theorem. \hfill \Box

**Proof of Theorem 2.2.** For any $\tilde{\theta} \in \mathbb{R}$, we can compute that

$$\mathbb{E}[e^{\tilde{\theta} L_t}] = \mathbb{E}[e^{\tilde{\theta} L_{t-1} + \tilde{\theta} X_t}] = \mathbb{E}\left[\mathbb{E}[e^{\tilde{\theta} L_{t-1} + \tilde{\theta} X_t} | F_{t-1}]\right]$$

$$= \mathbb{E}[e^{\tilde{\theta} L_{t-1}} \mathbb{E}[e^{\tilde{\theta} X_t} | F_{t-1}]] = \mathbb{E}[\mathbb{E}[e^{\tilde{\theta} L_{t-1}} \mathbb{E}[e^{\tilde{\theta} X_t} | F_{t-1}]]$$

where we used the fact that $X_t$ is compound Poisson with intensity $\lambda_t$ conditional on $F_{t-1}$, the natural filtration up to time $t - 1$. By the definition of $\lambda_t$, we have

$$\mathbb{E}[e^{\tilde{\theta} L_t}] = \mathbb{E}\left[e^{\tilde{\theta} L_{t-1} + (E[e^{\tilde{\theta} X_1}] - 1) \nu + (E[e^{\tilde{\theta} X_1}] - 1) \sum_{i=1}^{\lambda_{t-1}}} x(s) X_{t-i} \right]$$

$$= e^{(E[e^{\tilde{\theta} X_1}] - 1) \nu} \mathbb{E}\left[e^{\tilde{\theta} L_{t-2} + (E[e^{\tilde{\theta} X_1}] - 1) \sum_{i=1}^{\lambda_{t-2}} x(s) X_{t-i}} \right]$$

By the definition of $\lambda_{t-1}$, we get

$$\mathbb{E}[e^{\tilde{\theta} L_t}] = e^{(E[e^{\tilde{\theta} X_1}] - 1) \nu} \mathbb{E}\left[e^{\tilde{\theta} L_{t-2} + (E[e^{\tilde{\theta} X_1}] - 1) \sum_{i=1}^{\lambda_{t-2}} x(s) X_{t-i}} \right]$$

By induction on $t$, we get

$$\mathbb{E}[e^{\tilde{\theta} L_t}] = e^{(g_0(\tilde{\theta}) - 1) + (g_1(\tilde{\theta}) - 1) + \ldots + (g_{t-1}(\tilde{\theta}) - 1) \nu} = \mathbb{E}[e^{\nu(\sum_{i=0}^{t-1} x_i(\tilde{\theta}) - 1)}]$$  \hspace{1cm} (33)

where $g_0(\tilde{\theta}) = \mathbb{E}[e^{\tilde{\theta} L_1}]$, $g_1(\tilde{\theta}) = \mathbb{E}[e^{(\tilde{\theta} + (g_0(\tilde{\theta}) - 1)x(1)) L_1}]$, and more generally, for every $s \geq 1$, 

where
\[
E \left[ e^{(\bar{\theta} + (g_{\infty}(\bar{\theta}) - 1) x(1)) + (g_{\infty}(\bar{\theta}) - 1) x(2) + \cdots + (g_{\infty}(\bar{\theta}) - 1) x(s)} \right]_{\ell_{1,1}}
\]
\[= E \left[ e^{(\sum_{i=1}^{s} (g_{\infty}(\bar{\theta}) - 1) x(i))} \right]_{\ell_{1,1}}.
\] (34)

This implies that
\[
\lim_{t \to \infty} \frac{1}{t} \log E \left[ e^{\bar{\theta} L_t} \right] = \nu (g_{\infty}(\bar{\theta}) - 1),
\] (35)

where
\[
g_{\infty}(\bar{\theta}) = E \left[ e^{(\bar{\theta} + (g_{\infty}(\bar{\theta}) - 1) x(1))} \right]_{\ell_{1,1}}.
\] (36)

Let \( \tilde{G}(x) = E \left[ e^{(\bar{\theta} + (x-1) |x|)} \right]_{\ell_{1,1}} - x \). Similar as before, \( \tilde{G}(x) \) is increasing in \( \bar{\theta} \) and there exists finite \( \bar{\theta}_c > 0 \) such that we have \( \lim_{t \to \infty} \frac{1}{t} \log E \left[ e^{\bar{\theta} L_t} \right] = \nu (g_{\tilde{\ell}}(\bar{\theta}) - 1) \), where \( g_{\tilde{\ell}}(\bar{\theta}) \) is the minimal solution to the equation \( x = E \left[ e^{\tilde{\ell} \tilde{\ell}_{1,1} + |x|} \tilde{\ell}_{1,1} \right] \) for any \( \bar{\theta} \leq \bar{\theta}_c \). Note that \( \bar{\theta}_c > 0 \) satisfies the equation \( E \left[ ||x|| \tilde{\ell}_{1,1} + |x| \tilde{\ell}_{1,1} x \right] = 1 \), where \( \tilde{x} > 1 \) satisfies the equation \( \tilde{x}_c = E \left[ e^{\tilde{\ell}_{1,1} + |x|} \tilde{\ell}_{1,1} (\tilde{x}_c - 1) \right] \).

Finally, we can check the essential smoothness condition similar as before.
\[
g'_{\infty}(\bar{\theta}) = \frac{E \left[ e^{(\bar{\theta} + (g_{\infty}(\bar{\theta}) - 1)) |x|} \tilde{\ell}_{1,1} |x| \tilde{\ell}_{1,1} \right]}{1 - E \left[ ||x|| \tilde{\ell}_{1,1} e^{\tilde{\ell}_{1,1} + |x|} \tilde{\ell}_{1,1} (\tilde{x}_c - 1) \right]},
\] (37)

it is not hard to find \( |g'_{\infty}(\bar{\theta})| \to \infty \) as \( \bar{\theta} \to \bar{\theta}_c \), the conclusion then follows Gärnter-Ellis theorem.

\[\square\]

3.2. Proof of moderate deviations

Proof of Theorem 2.3. First, for any \( \theta \in \mathbb{R} \), we prove that
\[
\lim_{t \to \infty} \frac{t}{c^2(t)} \log E \left[ e^{\frac{(\theta - \mu) N_t}{\sigma^2(t)}} \right] = \frac{\nu \theta^2 \left( 1 + \text{Var}(\ell_{1,1}) ||x||^2 \right)}{2 \left( 1 - E [\ell_{1,1} ||x||^2] \right)},
\]
where \( \mu \) is defined by Equation (12).

By the proof of Theorem 2.1, we get
\[
E \left[ e^{\frac{(\theta - \mu) N_t}{\sigma^2(t)}} \right] = e^{\nu \left( (f_0(\theta_t) - 1) + (f_1(\theta_t) - 1) + \cdots + (f_{s-1}(\theta_t) - 1) \right) E [f_{s-1}(\theta_t) N_t]}
\]
\[= e^{\nu \left( (f_0(\theta_t) - 1) + (f_1(\theta_t) - 1) + \cdots + (f_{s-1}(\theta_t) - 1) \right) N_t},
\]
where \( f_0(\theta_t) = \theta_t - \frac{c(t)}{T} \), \( f_1(\theta_t) = \theta_t + \log E \left[ e^{(\theta_t - 1) x(1)} \ell_{1,1} \right] \), and \( f_2(\theta_t) = \theta_t + \log E \left[ e^{(\theta_t - 1) x(1) + (\theta_t - 1) x(2)} \ell_{1,1} \right] \). More generally, for every \( s \geq 1 \),
\[
f_s(\theta_t) = \theta_t + \log E \left[ e^{(f_{s-1}(\theta_t) - 1) x(1) + (f_{s-1}(\theta_t) - 1) x(2) + \cdots + (f_{s-1}(\theta_t) - 1) x(s)} \ell_{1,1} \right].
\]
Then we can rewrite the above equation such that

\[ e^{c_t(t_0)} = e^{\beta_0} \mathbb{E} \left[ e^{\left( \alpha (d_{t-1}^{(t_0)})^{-1} \alpha (1) + \alpha (d_{t-2}^{(t_0)})^{-1} \alpha (2) + \cdots + \alpha (d_{t}^{(t_0)})^{-1} \alpha (s) \right) \ell_{t-1}} \right]. \]

Let us define \( G_t(s) = e^{c_t(t_0)} - 1 \) so that \( G_t(s) = \mathbb{E} \left[ e^{\theta_t + \ell_{t-1} \sum_{i=1}^{s} \alpha(i) G_t(s-i)} \right] - 1. \) Then we have

\[ \mathbb{E} \left[ e^{c_t(t_0) N_t} \right] = e^{\sum_{i=0}^{t-1} G_t(s)}. \] (38)

We write \( G_t(s) \) instead of \( G(s) \) to indicate its dependence on \( t \) because of the term, \( \frac{c(t)}{t} \). As the proof of Theorem 2.1 shows and \( \frac{c(t)}{t} \) is sufficient small so that we have \( \frac{c(t)}{t} \theta \leq \theta_c \) where \( \theta_c = -\log \mathbb{E} \left[ ||x||_1 \ell_1 e^{||x||_1(x_s,-1)} \right] \) and as \( s \to \infty \), we get that \( G_t(\infty) \) is the minimal solution to the equation \( x_t = \mathbb{E} \left[ e^{c(t) \theta + \ell_{i-1} \sum_{i=1}^{\infty} \alpha(i) x_i} \right] - 1. \) Because of assumption (I), it is easy to see that \( x_t = O((c(t)/t)) \). Because \( x_t = O((c(t)/t)) \), we have \( G_t(s) = O((c(t)/t)) \) uniformly in \( s \). By Taylor’s expansion,

\[ G_t(s) = \frac{c(t) \theta}{t} + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_t(s-i) \right] \\
+ \frac{1}{2} \left( \frac{c(t) \theta}{t} \right)^2 + \frac{1}{2} \mathbb{E} \left[ \left( \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_t(s-i) \right)^2 \right] \\
+ \frac{c(t) \theta}{t} \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_t(s-i) \right] + O \left( \left( \frac{c(t)}{t} \right)^3 \right). \] (39)

Now, let

\[ G_t(s) = \frac{c(t) \theta}{t} G_1(s) + \left( \frac{c(t) \theta}{t} \right)^2 G_2(s) + \epsilon_t(s), \] (40)

where \( G_1(s) \) satisfies

\[ G_1(s) = 1 + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_1(s-i) \right] \], \quad (41)

and \( G_1(0) = 1 \), and \( G_2(s) \) satisfies

\[ G_2(s) = \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_2(s-i) \right] + \frac{1}{2} + (G_1(s) - 1) + \frac{1}{2} \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_1(s-i) \right]^2 \], \quad (42)

and \( G_2(0) = 1/2 \). Then we can substitute \( G_t(s) \) in terms of Equations (41) and (42) into the left and right side of Equation (39) so that we get \( \epsilon_t(s) = O \left( \left( \frac{c(t)}{t} \right)^3 \right) \).
By Equation (38),
\[
\frac{t}{c^2(t)} \log \mathbb{E} \left[ e^{c \theta(N_t - \mu)} \right] = \frac{t}{c^2(t)} \left( \nu \sum_{s=0}^{t-1} G_t(s) - \frac{\mu \theta t}{c(t)} \right).
\]

Then we can rewrite the above equation in terms of Equation (40),
\[
\frac{t}{c^2(t)} \log \mathbb{E} \left[ e^{c \theta(N_t - \mu)} \right] = \frac{t}{c^2(t)} \left( \nu \sum_{s=0}^{t-1} \left( \frac{c(t) \theta}{t} G_t(s) + \left( \frac{c(t) \theta}{t} \right)^2 G_2(s) + \epsilon_t(s) \right) \right) - \frac{\mu \theta t}{c(t)}
\]
\[
= \frac{\nu \theta}{c(t)} \sum_{s=0}^{t-1} G_t(s) - \frac{\mu \theta t}{c(t)} + \frac{\nu \theta^2}{t} \sum_{s=0}^{t-1} G_2(s) + O \left( \left( \frac{c(t)}{t^2} \right) \right).
\]

Now let us compute $\sum_{s=0}^{t-1} G_1(s)$. By Equation (41),
\[
\sum_{s=0}^{t-1} G_1(s) = \sum_{s=1}^{t-1} G_1(s) + 1 = 1 + \sum_{s=1}^{t-1} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_1(s - i) \right]
\]
\[
= t + \sum_{s=1}^{t-1} \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) G_1(s - i) \right]
\]
\[
= t + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \sum_{s=i}^{t-1} \alpha(i) G_1(s - i) \right]
\]
\[
= t + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \sum_{j=0}^{t-1-i} \alpha(i) G_1(j) \right]
\]
\[
= t + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \left( \sum_{j=0}^{t-1} G_1(j) - \sum_{j=t-i}^{t-1} G_1(j) \right) \right].
\]

After rewriting the above equation, we get
\[
\sum_{s=0}^{t-1} G_1(s) - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \sum_{j=0}^{t-1} \alpha(i) G_1(j) \right] = t - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \sum_{j=t-i}^{t-1} \alpha(i) G_1(j) \right]
\]
\[
\sum_{s=0}^{t-1} G_1(s) = \frac{t - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \sum_{j=t-i}^{t-1} \alpha(i) G_1(j) \right]}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \right]}.
\]
\[ \frac{\nu t}{c(t)} \sum_{s=0}^{t-1} G_1(s) - \frac{\mu t}{c(t)} = \nu t \left( \frac{t - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \sum_{j=t-i}^{t-1} G_1(j) \right]}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]} - \frac{t}{1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}]} \right) \]

\[ \frac{\nu t}{c(t)} \left( \frac{t}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]} - \frac{t}{1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}]} \right) \]

\[ = \frac{\mu t}{c(t)} \left( -\|\varphi\|_1 \mathbb{E}[\ell_{1,1}] + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right] \right) \left( \frac{t}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]} \right) \left( 1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}] \right) \]

\[ \leq \frac{\mu t}{c(t)} \left( \frac{\mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]}{1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}]} \right)^2 \]

(43)

For the first term in Equation (43), we can compute

\[ \left| \frac{\nu t}{c(t)} \left( \frac{t}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]} - \frac{t}{1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}]} \right) \right| \]

\[ = \frac{\mu t}{c(t)} \left( -\|\varphi\|_1 \mathbb{E}[\ell_{1,1}] + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right] \right) \left( \frac{t}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]} \right) \left( 1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}] \right) \]

\[ \leq \frac{\mu t}{c(t)} \left( \frac{\mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \varphi(i) \right]}{1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}]} \right)^2 \]

According to the assumption \( \sup_{t>0} t^{3/2} \|\varphi\| = C < \infty \), we can find \( \sum_{i=t}^{\infty} \varphi(i) \leq \sum_{i=t}^{\infty}Ci^{-3/2} < 2C(t-1)^{-1/2} \). Thus,

\[ \frac{\nu t}{c(t)} \left( \frac{\theta \mathbb{E}[\ell_{1,1}]^{\frac{2}{t^{1/2}}} \sqrt{t-1}}{1 - \|\varphi\|_1 \mathbb{E}[\ell_{1,1}]} \right) \]

By Lemma 3.1, \( G_1(t) \) is uniformly bounded. Then for the second term in Equation (43),
\[
\limsup_{t \to \infty} \nu \theta \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \chi(i) \sum_{j=t-i}^{t-1} G_1(j) \right] \leq G_1(\infty) \limsup_{t \to \infty} \nu \theta \frac{\mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} (i-1) \chi(i) \right]}{1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} \]
\[
= G_1(\infty) \limsup_{t \to \infty} \nu \theta \frac{\sum_{i=1}^{t-1} \mathbb{E}[\ell_{1,1} (i-1) \chi(i)]}{1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} .
\]

And we can compute
\[
\limsup_{t \to \infty} \nu \theta \frac{\sum_{i=1}^{t-1} \mathbb{E}[\ell_{1,1} (i-1) \chi(i)]}{1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} = \limsup_{t \to \infty} \nu \theta \frac{\sum_{i=1}^{t-1} \mathbb{E}[\ell_{1,1} (i-1) \chi(i)]}{1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} \frac{\nu |\theta|}{c(t) 1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} \frac{\sum_{i=1}^{t-1} \mathbb{E} [\ell_{1,1} C \sqrt{i}]}{c(t) 1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} \frac{2C \sqrt{t} \nu |\theta|}{c(t) \mathbb{E}[\ell_{1,1}]} \frac{1}{1 - ||\chi||_1 \mathbb{E}[\ell_{1,1}]} = 0
\]
Hence,
\[
\lim_{t \to \infty} \left[ \nu \theta \frac{t}{c(t)} \sum_{s=0}^{t-1} G_1(s) - \frac{\mu \theta t}{c(t)} \right] = 0.
\]
Furthermore, we also get
\[
\lim_{t \to \infty} \sum_{s=0}^{t-1} \frac{1}{t} G_1(s) = \frac{1}{1 - \mathbb{E}[\ell_{1,1}] ||\chi||_1}.
\]

According to Lemma 3.2, \(G_2(t)\) is uniformly bounded in \(t\). Then we can compute
\[
\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} G_2(s) = \frac{1}{2} \left( 1 + 2 \left( \frac{1}{1 - \mathbb{E}[\ell_{1,1}] ||\chi||_1} - 1 \right) + \frac{\mathbb{E} \left[ \left( \ell_{1,1} \sum_{i=1}^{t} \chi(i) \right)^2 \right]}{(1 - \mathbb{E}[\ell_{1,1}] ||\chi||_1)^2} \right) \]
\[
= \frac{1 + \text{Var}(\ell_{1,1}) ||\chi||_1^2}{2 (1 - \mathbb{E}[\ell_{1,1}] ||\chi||_1)^3}.
\]

Now we can have
\[
\lim_{t \to \infty} \frac{\nu \theta^2}{t} \sum_{s=0}^{t-1} G_2(s) = \frac{\nu \theta^2 \left( 1 + \text{Var}(\ell_{1,1}) ||\chi||_1^2 \right)}{2 (1 - \mathbb{E}[\ell_{1,1}] ||\chi||_1)^3}.
\]
Thus, we can prove
\[
\lim_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{E} \left[ e^{\frac{\nu t^2}{c^2(t)}} \right] = \frac{\nu \theta^2 \left( 1 + \text{Var}(\ell_{1,1})||x||_1^2 \right)}{2 \left( 1 - \mathbb{E}[\ell_{1,1}]||x||_1 \right)^3}.
\]

Applying the Gärtner-Ellis theorem, we conclude that, for any Borel set \( A \),
\[
- \inf_{x \in A} J(x) \leq \liminf_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P} \left( \frac{N_t - \mu t}{c(t)} \in A \right) \leq \limsup_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P} \left( \frac{N_t - \mu t}{c(t)} \in A \right) \leq - \inf_{x \in A} J(x),
\]
where
\[
J(x) = \sup_{\theta \in \mathbb{R}} \left\{ \theta x - \frac{\nu \theta^2 \left( 1 + \text{Var}(\ell_{1,1})||x||_1^2 \right)}{2 \left( 1 - \mathbb{E}[\ell_{1,1}]||x||_1 \right)^3} \right\} = \frac{x^2 \left( 1 - \mathbb{E}[\ell_{1,1}]||x||_1 \right)^3}{2 \nu \left( 1 + \text{Var}(\ell_{1,1})||x||_1^2 \right)}.
\]

\[ \blacksquare \]

\textbf{Proof of Theorem 2.4.} First, let us prove for any \( \tilde{\theta} \in \mathbb{R} \)
\[
\lim_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{E} \left[ e^{\frac{\nu t^2}{c^2(t)}(L_t - \tilde{\mu} t)} \right] = \frac{\nu \tilde{\theta}^2 \mathbb{E} \left[ \ell_{1,1}^2 \right]}{2 \left( 1 - \mathbb{E}[\ell_{1,1}]||x||_1 \right)^3},
\]
where \( \tilde{\mu} \) is defined in Equation (12).

As the proof of Theorem 2.2 shows, we can find
\[
\mathbb{E} \left[ e^{\frac{\nu t^2}{c^2(t)} \tilde{\theta}^2 t} \right] = e^{\nu \tilde{\theta}^2 \left( g_0(\tilde{\theta}) + \gamma(\tilde{\theta}) \right)} = e^{\nu \tilde{\theta}^2 \sum_{i=0}^{\tilde{\theta}^2} (g_i(\tilde{\theta})) - 1},
\]
where \( \tilde{\theta} = \frac{\nu t^2}{c^2(t)} \tilde{\theta}, g_0(\tilde{\theta}) = \mathbb{E} \left[ e^{\tilde{\theta} \ell_{1,1}} \right], g_1(\tilde{\theta}) = \mathbb{E} \left[ e^{\tilde{\theta} \ell_{i,1}} \right], \) and in general for every \( s \geq 1, \)
\[
g_s(\tilde{\theta}) = \mathbb{E} \left[ e^{\tilde{\theta}(s(\tilde{\theta}) - 1)\ell_{1,1}} \right] = \mathbb{E} \left[ e^{\tilde{\theta}(s(\tilde{\theta}) - 1)\ell_{1,1}} \right].
\]

Thus, \( \tilde{G}_t(0) = \mathbb{E} \left[ e^{\tilde{\theta} \ell_{1,1}} \right] - 1, \tilde{G}_t(s) = \mathbb{E} \left[ e^{\tilde{\theta} \sum_{i=1}^{s} \tilde{\theta}(i) \ell_{1,1}} \right] - 1 \) for \( s \geq 1. \)

Because of \( \frac{\nu t^2}{c^2(t)} \tilde{\theta} \), we write \( \tilde{G}_t(s) \) instead of \( G_t(s) \) to indicate its dependence on \( t. \)

According to the proof of Theorem 2.2, \( \frac{\nu t^2}{c^2(t)} \tilde{\theta} \) is sufficient small so that we have \( \frac{\nu t^2}{c^2(t)} \tilde{\theta} \leq \tilde{\theta}_c \), where \( \tilde{\theta}_c \) satisfies \( \mathbb{E} \left[ ||x||_1 \ell_{1,1} e^{\tilde{\theta}_c \ell_{1,1}} ||x||_1^{s(\tilde{\theta}_c - 1)} \right] = 1 \) and as \( s \to \infty \), we get that
\( \tilde{G}_t(\infty) \) is the minimal solution to the equation \( \tilde{x}_t = \mathbb{E} \left[ e^{\frac{\nu t^2}{c^2(t)} \ell_{1,1} + \frac{\nu t^2}{c^2(t)} \sum_{i=1}^{s(\tilde{\theta}_c - 1)} x(i) \tilde{x}_i} \right] \) - 1.

Because of the assumption (I), it is easy to see that \( \tilde{x}_t = O((c(t)/t)) \). Because \( \tilde{x}_t = O((c(t)/t)) \), we have \( \tilde{G}_t(s) = O((c(t)/t)) \) uniformly in \( s. \) By Taylor’s expansion,
\[ \tilde{G}_t(s) = \frac{c(t)\tilde{\theta}}{t} \mathbb{E}[\ell_{1,1}] + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_t(s - i) \right] \\
+ \frac{1}{2} \left( \frac{c(t)\tilde{\theta}}{t} \right)^2 \mathbb{E} \left[ \ell_{1,1}^2 \right] + \frac{1}{2} \mathbb{E} \left[ \left( \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_t(s - i) \right)^2 \right] \\
+ \frac{c(t)\tilde{\theta}}{t} \mathbb{E} \left[ \ell_{1,1}^2 \sum_{i=1}^{s} \alpha(i) \tilde{G}_t(s - i) \right] + O \left( \left( \frac{c(t)}{t} \right)^3 \right). \] (51)

We can let

\[ \tilde{G}_t(s) = \frac{c(t)\tilde{\theta}}{t} \tilde{G}_1(s) + \left( \frac{c(t)\tilde{\theta}}{t} \right)^2 \tilde{G}_2(s) + \epsilon_t(s), \] (52)

where \( \tilde{G}_1(s) \) satisfies

\[ \tilde{G}_1(s) = \mathbb{E}[\ell_{1,1}] + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_1(s - i) \right], \] (53)

\( \tilde{G}_1(0) = \mathbb{E}[\ell_{1,1}] \) and \( \tilde{G}_2(s) \) satisfies

\[ \tilde{G}_2(s) = \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_2(s - i) \right] + \frac{\mathbb{E} \left[ \ell_{1,1}^2 \right]}{2}, \] (54)

\[ \tilde{G}_2(0) = \mathbb{E}[\ell_{1,1}^2] / 2. \] Then we can substitute \( \tilde{G}_t(s) \) in terms of Equations (53) and (54) into the left and right side of Equation (51) and we find \( \epsilon_t(s) = O \left( \left( \frac{c(t)}{t} \right)^3 \right) \).

By Equation (49),

\[ \frac{t}{c^2(t)} \log \mathbb{E} \left[ e^{\frac{c(t)\tilde{\theta}}{c(t)}(L_t - \tilde{\mu}t)} \right] = \frac{t}{c^2(t)} \left( \nu \sum_{s=0}^{t-1} \tilde{G}_t(s) \right) - \frac{\tilde{\mu}t}{c(t)}, \]

Then by Equation (52), we can rewrite the above equation,

\[ \frac{t}{c^2(t)} \log \mathbb{E} \left[ e^{\frac{c(t)\tilde{\theta}}{c(t)}(L_t - \tilde{\mu}t)} \right] = \frac{t}{c^2(t)} \left( \nu \sum_{s=0}^{t-1} \frac{c(t)\tilde{\theta}}{t} \tilde{G}_1(s) + \left( \frac{c(t)\tilde{\theta}}{t} \right)^2 \tilde{G}_2(s) + \epsilon_t(s) \right) - \frac{\tilde{\mu}t}{c(t)} \]

\[ = \nu \frac{\tilde{\theta}}{c(t)} \sum_{s=0}^{t-1} \tilde{G}_1(s) - \frac{\tilde{\mu}t}{c(t)} + \nu \frac{\tilde{\theta}^2}{t} \sum_{s=0}^{t-1} \tilde{G}_2(s) + O \left( \left( \frac{c(t)}{t^2} \right)^3 \right). \]
Now let us compute $\sum_{s=0}^{t-1} \tilde{G}_1(s)$. By Equation (53),

$$
\sum_{s=0}^{t-1} \tilde{G}_1(s) = \sum_{s=1}^{t-1} \tilde{G}_1(s) + \mathbb{E}[\ell_{1,1}] = \mathbb{E}[\ell_{1,1}] + \sum_{s=1}^{t-1} \mathbb{E}[\ell_{1,1}] + \sum_{s=1}^{t-1} \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_1(s - i) \right]
$$

$$
= t \mathbb{E}[\ell_{1,1}] + \sum_{s=1}^{t-1} \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_1(s - i) \right]
$$

$$
= t \mathbb{E}[\ell_{1,1}] + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \sum_{s=i}^{t-1} \alpha(i) \tilde{G}_1(j) \right]
$$

$$
= t \mathbb{E}[\ell_{1,1}] + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \left( \sum_{j=0}^{t-1} \tilde{G}_1(j) - \sum_{j=t-i}^{t-1} \tilde{G}_1(j) \right) \right].
$$

After rewriting the above equation, we get

$$
\sum_{s=0}^{t-1} \tilde{G}_1(s) - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \sum_{j=0}^{t-1} \tilde{G}_1(j) \right] = t \mathbb{E}[\ell_{1,1}] - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \sum_{j=t-i}^{t-1} \tilde{G}_1(j) \right]
$$

$$
\sum_{s=0}^{t-1} \tilde{G}_1(s) = \frac{t \mathbb{E}[\ell_{1,1}] - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \sum_{j=t-i}^{t-1} \tilde{G}_1(j) \right]}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \right]}. 
$$

And we can compute

$$
\frac{\nu \bar{\theta}}{c(t)} \sum_{s=0}^{t-1} \tilde{G}_1(s) - \frac{\tilde{\mu} \bar{t}}{c(t)} = \nu \bar{\theta} \left( \frac{t \mathbb{E}[\ell_{1,1}] - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \sum_{j=t-i}^{t-1} \tilde{G}_1(j) \right]}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \right]} - \frac{t \mathbb{E}[\ell_{1,1}]}{1 - ||\alpha|| \mathbb{E}[\ell_{1,1}]} \right)
$$

$$
= \nu \bar{\theta} \left( \frac{t \mathbb{E}[\ell_{1,1}]}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \right]} - \frac{t \mathbb{E}[\ell_{1,1}]}{1 - ||\alpha|| \mathbb{E}[\ell_{1,1}]} \right)
$$

$$
- \nu \bar{\theta} \left( \frac{t \mathbb{E}[\ell_{1,1}]}{1 - \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \right]} - \frac{t \mathbb{E}[\ell_{1,1}]}{1 - ||\alpha|| \mathbb{E}[\ell_{1,1}]} \right).
$$

(56)
For the first term on the right hand side of the Equation (56), we can compute

\[
\frac{\hat{\nu} \cdot \theta}{c(t)} \left( \frac{t \mathbb{E}[\ell_{1,1}]}{1 - \mathbb{E}[\ell_{1,1} \sum_{i=1}^{t-1} \alpha(i)]} - \frac{t \mathbb{E}[\ell_{1,1}]}{1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]} \right) = \frac{\hat{\nu} \cdot t \mathbb{E}[\ell_{1,1}]}{c(t)} \left( \frac{-||\alpha||_1 \mathbb{E}[\ell_{1,1}] + \mathbb{E}[\ell_{1,1} \sum_{i=1}^{t-1} \alpha(i)]}{\left(1 - \mathbb{E}[\ell_{1,1} \sum_{i=1}^{t-1} \alpha(i)]\right) \left(1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]\right)} \right) \leq \frac{\hat{\nu} \cdot t \mathbb{E}[\ell_{1,1}]}{c(t)} \left( \frac{\mathbb{E}[\ell_{1,1} \sum_{i=1}^{\infty} \alpha(i)]}{\left(1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]\right)^2} \right) = \frac{\nu t}{c(t)} \left( \frac{\mathbb{E}[\ell_{1,1} \sum_{i=1}^{\infty} \alpha(i)]}{\left(1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]\right)^2} \right).
\]

According to the assumption \(\sup_{t > 0} t^{3/2} \alpha(t) = C < \infty\), we get \(\sum_{i=t}^{\infty} \alpha(i) \leq \sum_{i=t}^{\infty} Ci^{-3/2} < 2C(t - 1)^{-1/2}\). Therefore,

\[
\frac{\nu t \mathbb{E}[\ell_{1,1}]}{c(t)} \left( \frac{\mathbb{E}[\ell_{1,1} \sum_{i=1}^{\infty} \alpha(i)]}{\left(1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]\right)^2} \right) \leq \frac{\nu t}{c(t)} \left( \frac{\mathbb{E}[\ell_{1,1} \sum_{i=1}^{\infty} \alpha(i)]}{\left(1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]\right)^2} \right) \rightarrow 0, \text{ ast } \rightarrow \infty.
\]

Next, by Lemma 3.1, \(\tilde{G}_2(t)\) is uniformly bounded in \(t\), then for the second term on the right hand side of the Equation (56), we have

\[
\lim_{t \to \infty} \sup \left| \frac{\hat{\nu} \cdot \theta}{c(t)} \mathbb{E}[\ell_{1,1} \sum_{i=1}^{t-1} \alpha(i) \sum_{j=t-i}^{t-1} \tilde{G}_1(j)] \right| = \tilde{G}_1(\infty) \lim_{t \to \infty} \sup \left| \frac{\hat{\nu} \cdot \theta}{c(t)} \mathbb{E}[\ell_{1,1} \sum_{i=1}^{t-1} (i-1) \alpha(i)] \right| = \tilde{G}_1(\infty) \lim_{t \to \infty} \left( \frac{\nu \cdot \tilde{\theta} \cdot \mathbb{E}[\ell_{1,1} \sum_{i=1}^{t-1} (i-1) \alpha(i)]}{1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]} \right).
\]

Then we can get

\[
\lim_{t \to \infty} \sup \left| \frac{\nu \cdot \tilde{\theta}}{c(t)} \sum_{i=t}^{t-1} \mathbb{E}[\ell_{1,1}(i-1) \alpha(i)] \right| = \lim_{t \to \infty} \sup \left| \frac{\nu \cdot \tilde{\theta}}{c(t)} \sum_{i=t}^{t-1} \mathbb{E}[\ell_{1,1}(i-1) \alpha(i)] \right| \leq \lim_{t \to \infty} \left( \frac{\nu \cdot \tilde{\theta} \cdot \mathbb{E}[\ell_{1,1}(i-1) \alpha(i)]}{1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]} \right) \leq 2C\sqrt{\nu} \cdot \frac{\tilde{\theta}}{c(t)} \frac{\mathbb{E}[\ell_{1,1}]}{1 - ||\alpha||_1 \mathbb{E}[\ell_{1,1}]} = 0.
\]
Therefore, we can compute
\[
\lim_{t \to \infty} \left[ \frac{\nu \tilde{\theta}}{c(t)} \sum_{s=0}^{t-1} G_1(s) - \frac{\tilde{\mu} t}{c(t)} \right] = 0.
\]

Furthermore, we can show
\[
\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} G_1(t) = \frac{\mathbb{E}[\ell_{1,1}]}{1 - \mathbb{E}[\ell_{1,1}||x||_1]}.
\] (57)

For \( \frac{1}{t} \sum_{s=0}^{t-1} \tilde{G}_2(s) \), by Lemma 3.2, \( \tilde{G}_2(t) \) is uniformly bounded in \( t \). Thus, we can compute
\[
\lim_{t \to \infty} \frac{1}{t} \sum_{s=0}^{t-1} \tilde{G}_2(s) = \frac{\mathbb{E}\left[\ell_{1,1}^2\right]}{2\left(1 - \mathbb{E}[\ell_{1,1}||x||_1]\right)^3}.
\] (58)

Noe we have
\[
\lim_{t \to \infty} \frac{\nu \tilde{\theta}^2}{c(t)} \sum_{s=0}^{t-1} \tilde{G}_2(s) = \frac{\nu \tilde{\theta}^2 \mathbb{E}\left[\ell_{1,1}^2\right]}{2\left(1 - \mathbb{E}[\ell_{1,1}||x||_1]\right)^3}.
\] (59)

Thus, we can derive
\[
\lim_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{E}\left[e^{\nu \tilde{\theta}^2 \tilde{L}_t - \tilde{\mu} t}\right] = \frac{\nu \tilde{\theta}^2 \mathbb{E}\left[\ell_{1,1}^2\right]}{2\left(1 - \mathbb{E}[\ell_{1,1}||x||_1]\right)^3}.
\]

Applying the Gärtner-Ellis theorem, see Dembo and Zeitouni (1998), we conclude that, for any Borel set \( A \),
\[
- \inf_{x \in A} J(x) \leq \lim_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P}\left( \frac{L_t - \tilde{\mu} t}{c(t)} \in A \right) \\
\leq \limsup_{t \to \infty} \frac{t}{c^2(t)} \log \mathbb{P}\left( \frac{L_t - \tilde{\mu} t}{c(t)} \in A \right) \leq - \inf_{x \in A} J(x),
\] (60)

where
\[
J(x) = \sup_{\tilde{\theta} \in \mathbb{R}} \left\{ \tilde{\theta} x - \frac{\nu \tilde{\theta}^2 \mathbb{E}\left[\ell_{1,1}^2\right]}{2\left(1 - \mathbb{E}[\ell_{1,1}||x||_1]\right)^3} \right\} = \frac{x^2 \left(1 - \mathbb{E}[\ell_{1,1}||x||_1]\right)^3}{2\nu \mathbb{E}\left[\ell_{1,1}^2\right]}.
\] (61)
Lemma 3.1. For any $s \in \mathbb{N}$,

$$G_1(s) \leq \frac{1}{1 - \mathbb{E}[\ell_{1,1}||z||_1]} ,$$

where $G_1(s) = 1 + \mathbb{E}[\ell_{1,1} \sum_{i=1}^{s} z(i)G_1(s-i)]$ for $s \geq 1$ and $G_1(0) = 1$. and

$$\tilde{G}_1(s) \leq \frac{\mathbb{E}[\ell_{1,1}]}{1 - \mathbb{E}[\ell_{1,1}||z||_1]} ,$$

where $\tilde{G}_1(0) = \mathbb{E}[\ell_{1,1}]$ and for $s \geq 1$, $\tilde{G}_1(s) = \mathbb{E}[\ell_{1,1}] + \mathbb{E}[\ell_{1,1} \sum_{i=1}^{s} z(i)\tilde{G}_1(s-i)]$.

Proof of Lemma 3.1. We prove Lemma 3.1 by induction on $s$. By assumption (I), $G_1(0) \leq \frac{1}{1 - \mathbb{E}[\ell_{1,1}||z||_1]}$. Now, let us assume that $G_1(s) \leq \frac{1}{1 - \mathbb{E}[\ell_{1,1}||z||_1]}$. Then we can compute,

$$G_1(s + 1) = 1 + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s+1} z(i)G_1(s+1-i) \right]$$

$$\leq 1 + \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} z(i) \frac{1}{1 - \mathbb{E}[\ell_{1,1}||z||_1]} \right]$$

$$= 1 - \mathbb{E}[\ell_{1,1}||z||_1] + \mathbb{E}[\ell_{1,1} \sum_{i=1}^{s} z(i)]$$

$$\leq \frac{1}{1 - \mathbb{E}[\ell_{1,1}||z||_1]}$$

Hence, we proved that for every $s \in \mathbb{N}$, $G_1(s) \leq \frac{1}{1 - \mathbb{E}[\ell_{1,1}||z||_1]}$. Similarly, we can show $\tilde{G}_1(s) \leq \frac{\mathbb{E}[\ell_{1,1}]}{1 - \mathbb{E}[\ell_{1,1}||z||_1]}$. □

Lemma 3.2. For any $s \in \mathbb{N}$,

$$G_2(s) \leq \frac{1 + \text{Var}(\ell_{1,1})||z||_1^2}{2(1 - \mathbb{E}[\ell_{1,1}||z||_1])^3} ,$$

where $G_2(s) = \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} z(i)G_2(s-i) \right] + \frac{1}{2} + (G_1(s) - 1) + \frac{1}{2} \mathbb{E} \left[ (\ell_{1,1} \sum_{i=1}^{s} z(i)G_1(s-i))^2 \right]$ for $s \geq 1$ and $G_2(0) = 1/2$. And

$$\tilde{G}_2(s) \leq \frac{\mathbb{E}[\ell_{1,1}^2]}{2(1 - \mathbb{E}[\ell_{1,1}||z||_1])^3}$$

where $\tilde{G}_2(0) = \frac{\mathbb{E}[\ell_{1,1}]}{2}$ and for $s \geq 1$, 


\[
\tilde{G}_2(s) = \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_2(s-i) \right] + \frac{\mathbb{E} \left[ \ell_{1,1}^2 \right]}{2} \\
+ \mathbb{E} \left[ \ell_{1,1}^2 \sum_{i=1}^{s} \alpha(i) \tilde{G}_1(s-i) \right] + \frac{1}{2} \mathbb{E} \left[ \left( \ell_{1,1} \sum_{i=1}^{s} \alpha(i) \tilde{G}_1(s-i) \right)^2 \right].
\]

**Proof of Lemma 3.2.** We prove Lemma 3.2 by induction on s. By assumption (I), it is not hard to see \( G_2(0) \leq \frac{1 + \text{Var}(\ell_{1,1}) ||x||_1^2}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3} \).

Now, let us assume that \( G_2(s) \leq \frac{1 + \text{Var}(\ell_{1,1}) ||x||_1^2}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3} \). Then we can compute,

\[
G_2(s + 1) = \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s+1} \alpha(i) G_2(s+1-i) \right] + \frac{1}{2} (G_1(s) - 1) \\
+ \frac{1}{2} \mathbb{E} \left[ \left( \ell_{1,1} \sum_{i=1}^{s+1} \alpha(i) G_1(s+1-i) \right)^2 \right] \\
\leq \mathbb{E} \left[ \ell_{1,1} \sum_{i=1}^{s+1} \alpha(i) \frac{1 + \text{Var}(\ell_{1,1}) ||x||_1^2}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3} \right] + \frac{1}{2} + \frac{\mathbb{E}[\ell_{1,1} ||x||_1]}{1 - \mathbb{E}[\ell_{1,1} ||x||_1]} \\
+ \frac{1}{2} \mathbb{E} \left[ \left( \ell_{1,1} \sum_{i=1}^{s+1} \alpha(i) \frac{1}{1 - \mathbb{E}[\ell_{1,1} ||x||_1]} \right)^2 \right] \\
\leq \frac{1 + \text{Var}(\ell_{1,1}) ||x||_1^2}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3} \mathbb{E}[\ell_{1,1} ||x||_1] + \frac{1}{2} + \frac{\mathbb{E}[\ell_{1,1} ||x||_1]}{1 - \mathbb{E}[\ell_{1,1} ||x||_1]} \\
+ \frac{\mathbb{E}[\ell_{1,1}^2 ||x||_1^2]}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^2} \\
\leq \frac{1 + \text{Var}(\ell_{1,1}) ||x||_1^2}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3}.
\]

Hence, we proved that for every \( s \in \mathbb{N} \), \( G_2(s) \leq \frac{1 + \text{Var}(\ell_{1,1}) ||x||_1^2}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3} \). And similarly, we can also show that \( \tilde{G}_2(s) \leq \frac{\mathbb{E}[\ell_{1,1}^2]}{2(1 - \mathbb{E}[\ell_{1,1} ||x||_1])^3} \). \( \square \)

**References**

Bacry, E., S. Delattre, M. Hoffmann, and J. F. Muzy. 2013. Scaling limits for hawkes processes and application to financial statistics. *Stochastic Processes and Their Applications* 123 (7): 2475–99. doi:10.1016/j.spa.2013.04.007.

Bauwens, L., and N. Hautsch. 2009. Modelling financial high frequency data using point processes. *Handbook of Financial Time Series* :953–79.
Bordenave, C., and G. L. Torrisi. 2007. Large deviations of poisson cluster processes. *Stochastic Models* 23 (4):593–625. doi:10.1080/15326340701645959.

Brémaud, P., and L. Massoulié. 1996. Stability of nonlinear hawkes processes. *The Annals of Probability* 24 (3):1563–88. doi:10.1214/aop/1065725193.

Browning, R., D. Sulem, K. Mengersen, V. Rivoirard, and J. Rousseau. 2021. Simple discrete-time self-exciting models can describe complex dynamic processes: a case study of COVID-19. *Plos ONE* 16 (4):e0250015. doi:10.1371/journal.pone.0250015.

Chavez-Demoulin, V., and J. A. McGill. 2012. High-frequency financial data modeling using hawkes processes. *Journal of Banking & Finance* 36 (12):3415–26. doi:10.1016/j.jbankfin.2012.08.011.

Chen, X. 2001. Moderate deviations for markovian occupation times. *Stochastic Processes and Their Applications* 94 (1):51–70. doi:10.1016/S0304-4149(01)00079-5.

Daley, D. J., and D. Vere-Jones. 2003. *An introduction to the theory of point processes, Vols. I and II*. New York: Springer.

Dembo, A. 1996. Moderate Deviations for Martingales with Bounded Jumps. *Electronic Communications in Probability* 1:11–7. doi:10.1214/ECP.v1-973.

Dembo, A., and O. Zeitouni. 1998. *Large deviations techniques and applications*. 2nd ed. New York: Springer.

Egami, M., Y. Kato, and T. Sawaki. 2013. An analysis of CDS market liquidity by the Hawkes Process. SSRN eLibrary.

Fox, E. W., M. B. Short, F. P. Schoenberg, K. D. Coronges, and A. L. Bertozzi. 2016. Modeling E-mail networks and inferring leadership using self-exciting point processes. *Journal of the American Statistical Association* 111 (514):564–84. doi:10.1080/01621459.2015.1135802.

Gao, F. 1996. Moderate deviations for martingales and mixing random processes. *Stochastic Processes and Their Applications* 61 (2):263–75. doi:10.1016/0304-4149(95)00078-X.

Gao, X., and L. Zhu. 2018. Large deviations and applications for Markovian Hawkes processes with a large initial intensity. *Bernoulli* 24 (4A):2875–905. doi:10.3150/17-BEJ948.

Giesecke, K., L. L. Goldberg, and X. Ding. 2011. A top-down approach to multi-name credit. *Operations Research* 59 (2):283–300. doi:10.1287/opre.1100.0855.

Hawkes, A. G. 1971. Spectra of some self-exciting and mutually exciting point processes. *Biometrika* 58 (1):83–90. doi:10.1093/biomet/58.1.83.

Horst, U., and W. Xu. 2021. Functional limit theorems for marked Hawkes point measures. *Stochastic Processes and Their Applications* 134 (C):94–131. doi:10.1016/j.spa.2020.12.002.

Karabash, D., and L. Zhu. 2015. Limit theorems for Marked Hawkes processes with application to a risk model. *Stochastic Models* 31 (3):433–51. doi:10.1080/15326349.2015.1024868.

Liniger, T. J. 2009. Multivariate Hawkes processes,” ETH Zrich.

Seol, Y. 2015. Limit theorems for discrete Hawkes processes. *Statistics & Probability Letters* 99:223–9. doi:10.1016/j.spl.2015.01.023.

Seol, Y. 2017. Moderate deviations for Marked Hawkes processes. *Acta Mathematica Sinica, English Series* 33 (10):1297–304. doi:10.1007/s10114-017-6433-7.

Varadhan, S. R. S. 1984. *Large deviations and applications*. Philadelphia: SIAM.

Vere-Jones, D. 1975. Stochastic models for earthquake sequences. *Geo-Physical Journal International* 42 (2):811–26. doi:10.1111/j.1365-246X.1975.tb05893.x.

Wang, H. 2020. *Limit theorems for a discrete-time Marked Hawkes process*. arXiv: 2003.02810 [math.PR].

Xu, Y., L. Zhu, and H. Wang. 2020. A self- and mutual-exciting model for discrete-time data: case study on online money market fund. SSRN. doi:10.2139/ssrn.3665436.

Yao, N. 2018. Moderate deviations for multivariate Hawkes processes. *Statistics & Probability Letters* 140:71–6. doi:10.1016/j.spl.2018.04.023.

Zhang, X., J. Blanchet, K. Giesecke, and P. W. Glynn. 2015. Affine point processes: approximation and efficient simulation. *Mathematics of Operations Research* 40 (4):797–1088. doi:10.1287/moor.2014.0696.

Zhu, L. 2013a. Central limit theorem for nonlinear Hawkes processes. *Journal of Applied Probability* 50 (3):760–71. doi:10.1239/jap/1378401234.
Zhu, L. 2013b. Moderate deviations for Hawkes processes. *Statistics & Probability Letters* 83 (3): 885–90. doi:10.1016/j.spl.2012.12.011.

Zhu, L. 2013c. Nonlinear Hawkes processes. New York University.

Zhu, L. 2014a. Limit theorems for a Cox-Ingersoll-Ross process with Hawkes jumps. *Journal of Applied Probability* 51 (3):699–712. doi:10.1239/jap/1409932668.

Zhu, L. 2014b. Process-level large deviations for nonlinear Hawkes point processes. *Annales de L'Institut Henri Poincaré-Probabilités et Statistiques* 50 (3):845–71. doi:10.1214/12-AIHP532.

Zhu, L. 2015. Large deviations for Markovian nonlinear Hawkes processes. *Annals of Applied Probability* 25 (2):548–81. doi:10.1214/14-AAP1003.