J Theor Probab (2018) 31:41–67
https://doi.org/10.1007/s10959-016-0706-4

Large and Moderate Deviations for the Total Population Arising from a Sub-critical Galton–Watson Process with Immigration

Shihang Yu[1] · **Dehui Wang**[2] · **Xia Chen**[3]

Received: 24 February 2015 / Revised: 12 July 2016 / Published online: 26 July 2016

Abstract In this paper, we provide the exact forms of large and moderate deviations for the empirical mean of population and the centered total population of a sub-critical branching process with immigration. The rate functions in our large and moderate deviations are explicitly identified. Our theorems also apply to the models of the integer-valued autoregression. In computing the generating function requested by Gärtner-Ellis theorem, our treatment substantially relies on an algorithm specifically designed for the autoregressive structure of our models.

Keywords Branching process with immigration · Integer-valued AR model · Large deviations principle · Moderate deviation principle

Mathematics Subject Classification (2010) 60J80 · 60F10 · 62M10 · 62G20

1 Introduction

The goals of this paper are the large deviation principle (LDP) for the empirical mean

✉ Dehui Wang
wangdh@jlu.edu.cn

Shihang Yu
qqhrysh@163.com

Xia Chen
xchen@math.utk.edu

[1] School of Science, Qiqihar University, Qiqihar 161006, China

[2] Institute of Mathematics, Jilin University, Changchun 130012, China

[3] Department of Mathematics, University of Tennessee, Knoxville, TN 37996, USA

$$\frac{N_1 + \cdots + N_n}{n}$$

and the moderate deviation principle (MDP) for the centered total population

$$\sum_{k=1}^{n} (N_k - \lambda)$$

arising from the Galton–Watson process $\{N_k\}$ that is defined inductively by the equation

$$N_k = \sum_{j=1}^{N_{k-1}} \xi_{k-1,j} + \varepsilon_k \quad k = 1, 2, \ldots \tag{1.1}$$

where $\{\varepsilon_k, k \in \mathbb{N}^+\}$ and $\{\xi_{k-1,j}\}_{j,k}$ are two mutually independent i.i.d. sequence and array of nonnegative integer-valued random variables, $\lambda = \mathbb{E}\varepsilon/(1 - \mathbb{E}\xi)$ ($\mathbb{E}\xi < 1$ by assumption). For each k, ε_k and $\{\xi_{k-1,j}\}_j$ stand for, respectively, the number of the immigrants of the k-th generation and the numbers of the offsprings given birth by the individuals in the $(k-1)$-th generation. In this setup, the sequence N_k represents the size of the population in the k-th generation and the partial sum $N_1 + \cdots + N_n$ is the size of the total population ever lived on the earth up to the generation n.

In this paper, we use ε and ξ for the generic copies of $\{\varepsilon_k, k \in \mathbb{N}^+\}$ and $\{\xi_{k-1,j}\}_{j,k}$, respectively. Throughout, we focus on the sub-critical case given as $\mathbb{E}\xi < 1$. In other words, the average number of the children of each individual in the system is less than one. Since ξ takes nonnegative integers, the sub-criticity implies $\mathbb{P}\{\xi = 0\} > 0$. To avoid degeneracy, on the other hand, we always assume that $\mathbb{P}\{\xi = 0\} < 1$. Recall the well-known fact that without immigration, the sub-critical branching process extinct with probability 1. To prevent this from happening, we also assume that $\mathbb{P}\{\varepsilon = 0\} < 1$ throughout the paper.

Galton–Watson process is also called branching process in literature. We refer the book [4] by Athreya and Ney for the general information on branching processes and the paper [12] by Pakes for the central limit theorem for the centered total population of sub-critical branching process with immigration (Theorem 3, [12]). The branching process with immigration is also mathematically linked to the models in time series. In the special cases when ξ is Bernoulli ($\{0, 1\}$-valued) with $\mathbb{E}\xi = \alpha$ and when ξ is geometrically distributed with $\mathbb{E}\xi = \alpha/(1+\alpha)$, for example, $\{N_k\}$ becomes an integer-valued autoregressive process (known as INAR models) with the relation (1.1) being re-denoted as, respectively,

$$N_k = \alpha \circ N_{k-1} + \varepsilon_k \tag{1.2}$$

and

$$N_k = \alpha * N_{k-1} + \varepsilon_k. \tag{1.3}$$

More specifically, the above autoregressive processes are classified as INAR(1) models in statistical literature, due to the fact that they are "single-operator" systems. The operators "$\circ$" (known as binomial thinning operator) and "$*$" (known as geometric thinning operator) are introduced in Steutel and Van Harn [14] and Ristic et al. [13], respectively. The INAR(1) models have been important research subjects in statistics and finance. In application, the random term N_k can be used, for example, to measure the claim numbers in the insurance industry or the number of the patients in a hospital at the k-th period. Thus, the sum $N_1+\cdots+N_n$ is the accumulated claims up to the n-th fiscal year. We refer the works by Mckenzie [11], Al-Osh and Alzaid [1,2], Alzaid and Al-Osh [3], Weiss [16], Jung and Tremayne [9,10] for the discussion on other aspects of the autoregressive models.

Theorem 1.1 *In the relation* (1.1), let $N_0 \geq 0$ be a deterministic integer. Under the assumptions

$$\mathbb{E}\xi < 1, \quad \mathbb{E}e^{\theta\xi} < \infty \quad \text{and} \quad \mathbb{E}e^{\theta\varepsilon} < \infty \quad \theta > 0,$$

we have

$$\limsup_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^{n} N_k \in F\Big) \leqslant -\inf_{x\in F} I_L(x), \tag{1.4}$$

for each close set $F \in \mathbb{R}^+$; and

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\Big(\frac{1}{n}\sum_{k=1}^{n} N_k \in G\Big) \geqslant -\inf_{x\in G} I_L(x) \tag{1.5}$$

for each open set $G \in \mathbb{R}^+$, where the rate function $I_L(x)$ is given as

$$I_L(x) = \sup_{\theta\in\mathbb{R}} \left\{ x\left(\theta - \log \mathbb{E}e^{\theta\xi}\right) - \log \mathbb{E}\exp\{\theta\varepsilon\} \right\} \quad x \geq 0.$$

Theorem 1.2 *Let b_n be a sequence of positive numbers satisfying $b_n \to \infty$ and $b_n/n \to 0$. Under the conditions same as the ones given in Theorem 1.1,*

$$\limsup_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left\{ \frac{1}{\sqrt{nb_n}} \sum_{k=1}^{n} \left(N_k - \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi} \right) \in F \right\} \leqslant -\inf_{x\in F} I_M(x) \tag{1.6}$$

for each close set $F \subset \mathbb{R}$; and

$$\liminf_{n\to\infty} \frac{1}{b_n} \log \mathbb{P}\left\{ \frac{1}{\sqrt{nb_n}} \sum_{k=1}^{n} \left(N_k - \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi} \right) \in G \right\} \geqslant -\inf_{x\in G} I_M(x) \tag{1.7}$$

for each open set $G \in \mathbb{R}$, where the rate function $I_M(\cdot)$ is given as

$$I_M(x) = \frac{x^2}{2\sigma^2} \quad x \in \mathbb{R}, \quad \textit{where} \quad \sigma^2 = \frac{\text{Var}\,(\varepsilon)}{(1-\mathbb{E}\xi)^2} + (\mathbb{E}\varepsilon)\frac{\text{Var}\,(\xi)}{(1-\mathbb{E}\xi)^3}.$$

In the large deviation theory, the result given in Theorem 1.2 is referred as moderate deviation principle (MDP) partially due to its connection to the following central limit theorem (Theorem 3, [12]):

$$\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(N_k - \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}\right) \xrightarrow{d} N(0,\sigma^2).$$

What Theorem 1.2 tells is the story that the large deviation principle of Gaussian tail passes through the central limit theorem.

It is worthy of mentioning that the rate functions in our theorems, especially in Theorem 1.1, are explicitly given. The explicitness in rate function is particularly important when it comes to application. A careful reader may have noticed our unorthodox way of setting the rate function $I_L(x)$. By putting $\xi = 0$ (so our system consists only of immigrants) and $N_k = \varepsilon_k$,

$$I_L(x) = \sup_{\theta\in\mathbb{R}}\left\{\theta x - \log \mathbb{E}e^{\theta\varepsilon}\right\} \quad x \geq 0$$

and Theorem 1.1 becomes the well-known Cramér's LDP (Theorem 2.2.3, p. 27, [8]).

In the Sect. 2 below, we shall give the "close forms" of the rate function $I_L(x)$ in some settings of INAR(1) models. Here we would like to list some properties of $I_L(x)$ in the following lemma.

Lemma 1.3 *(1) $I_L(x) \geq 0$ for each $x \geq 0$. Further, $I_L(x) = 0$ if and only if*

$$x = \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}.$$

(2) Let $k_0 \geq 0$ be the least integer such that $\mathbb{P}\{\varepsilon = k_0\} > 0$. Then $I_L(x) = \infty$ on $[0, k_0)$ as $k_0 \geq 1$ and $I_L(0) = \log\left(\mathbb{P}\{\varepsilon = 0\}\right)^{-1}$ when $k_0 = 0$.
(3) $I_L(x)$ is a good rate function in the sense that $I_L(x)$ is lower semi-continuous on $\mathbb{R}^+$ and for each $l > 0$, the level set $\{x \in \mathbb{R}^+;\quad I_L(x) \leq l\}$ is compact.

The proof of this lemma will be given in Sect. 3 right after the proof of Theorem 1.1. Here we give a probabilistic interpretation of (2) instead. By the relation (1.1),

$$N_k \geq \varepsilon_k \geq k_0 \quad a.s. \quad k = 1, 2, \cdots.$$

Consequently, the probability that $(N_1 + \cdots + N_n)/n$ takes a value less than k_0 is zero.

Using (1) and (3) in Lemma 1.3, one can show that

$$\inf_{|x-\frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}|\geq\delta} I_L(x) > 0$$

for every $\delta > 0$. Notice that $\frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}$ is the expectation of N under the invariant distribution. As a consequence of Theorem 1.1, we are able to claim the genuine exponential decay for the probability that the sample average deviates away from its mathematical equilibrium value.

Corollary 1.4 *Under the assumption of Theorem 1.1, for every $\delta > 0$ there is a constant $c_\delta > 0$ such that*

$$\mathbb{P}\Big(\Big|\frac{1}{n}\sum_{k=1}^{n} N_k - \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}\Big| \geq \delta\Big) \leq \exp\{-c_\delta n\} \tag{1.8}$$

as n is sufficiently large.

In the language of population, Corollary 1.4 shows that a society with sub-critical birth rate and sizable immigration is "exponentially" stable as far as its population is concerned.

Since $\{N_k\}$ can be viewed as a Markov chain taking values in $\mathbb{Z}^+ = \{0, 1, 2, \ldots\}$, we intend to make a comparison between our theorems and the existing large and moderate deviations for additive functionals of an ergodic Markov chain $\{X_k\}$. In the general setting, the underline random sequences are the sample average (LDP) of the form

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \quad n = 1, 2, \ldots$$

and the centered quantity (MDP)

$$\frac{1}{\sqrt{nb_n}}\sum_{k=1}^{n}\Big\{f(X_k) - \int f(x)\pi(dx)\Big\} \quad \text{(where } \pi(dx) \text{ is the invariant distribution)}$$

and the results are formulated in a way comparable to Theorem 1.1 (see, e.g., Theorem 1, de Acosta and Ney [7] for the latest development) and Theorem 1.2 (see, e.g., Chapter 3, Chen [6]), respectively. Less general as they are, the theorems established in this work have the following advantages. First, Theorem 1.1 provides explicit and complete information on the rate function I_L that is crucial in applications but largely ignored in literature where the rate function is essentially incomputable in any non-trivial setting. Without knowing that $I_L(x) > 0$ for all $x \neq \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}$, for example, one could not claim the exponential decay given in (1.8). To the best knowledge of the authors, the only work in literature with the claim comparable to (1.8) is [7]. Second, the general results usually require strong ergodicity of the underline Markov chains coupling with some technical integrability condition on the function f that demands f to be bounded, to say the least (see, e.g., Theorem 1, de Acosta and Ney [7] for details). These assumptions are either too technical to be examined or fail in our setting. In particular, Theorem 1 in de Acosta and Ney [7] does not apply to the function $f(x) = x$ (in connection to Theorem 1.1). In addition, the super-exponential ergodicity condition (applied to our setting),

$$\mathbb{E}_0 \exp\{\theta\tau_0\} < \infty \quad \forall\theta > 0$$

assumed in Theorem 1, de Acosta and Ney [7] with $\tau_0 = \inf\{k \geq 1; \quad N_k = 0\}$, is not satisfied in our model. Unlike the previous work, the setup of our theorems does not require the information on the invariant distribution which is not explicitly given in some of our settings. Finally, our method is much more elementary and does not use the any special tools developed in the area of Markov processes.

Suggested by the referee, we remark on Gaussian autoregressive process(AR(1))[5]

$$N_k = \alpha N_{k-1} + \varepsilon_k$$

for the purpose of comparison (Notice that the relation (1.1) only works for integral-valued processes). Here we assume that $0 < \alpha < 1$, N_0 is a constant and $\{\varepsilon_k\}$ is a sequence of i.i.d. $N(\mu, \sigma_0^2)$-random variables known as white noise. By the relation

$$\begin{aligned}\sum_{k=1}^{n} N_k &= \frac{\alpha}{1-\alpha}(N_0 - N_n) + \frac{1}{1-\alpha}\sum_{k=1}^{n}(N_k - \alpha N_{k-1}) \\ &= \frac{\alpha}{1-\alpha}(N_0 - N_n) + \frac{1}{1-\alpha}\sum_{k=1}^{n}\varepsilon_k\end{aligned}$$

the summation on the right hand side is Gaussian for each $n \geq 1$ and

$$\mathbb{E}\left(\sum_{k=1}^{n} N_k\right) = \frac{n\mu}{1-\alpha} + O(1) \quad \text{and} \quad \mathrm{Var}\left(\sum_{k=1}^{n} N_k\right) = \frac{n\sigma_0^2}{(1-\alpha)^2} + O(1).$$

By the large deviation for Gaussian random variables (by Cramérs large deviation principle (Theorem 2.2.3, [8]) with the i.i.d. $N\Big(\mu(1-\alpha)^{-1}, \sigma_0^2(1-\alpha)^{-2}\Big)$-random variables, for example), Theorem 1.1 holds with the rate function $I_L(x) = I(x - \lambda)$ and Theorem 1.2 holds with the rate function $I_M(x) = I(x)$ and $\lambda = \mu(1-\alpha)^{-1}$, where

$$I(x) = \frac{(1-\alpha)^2}{2\sigma_0^2}x^2 \quad x \in \mathbb{R}.$$

It shows that in the Gaussian setting, the LDP and MDP are essentially the same thing. It is not difficult to extend this observation to some other types of Gaussian autoregressive processes.

We now comment on the approaches used in this work. As many other works in the large deviations, our treatment involves Gärtner-Ellis theorem (see, e.g., Theorem 2.3.6, p. 44, [8]) which reduces the problem to the computation of the logarithmic generating functions. What sets this work aside is our algorithm designed for the autoregressive structure of our model. The computation of the logarithmic generating functions will be carried out in Sect. 3 (for the LDP) and Sect. 4 (for the MDP).

2 Examples

In this section we consider some special settings that are linked to the integer-valued autoregressive models. These cases are included by Theorem 1.1 and Theorem 1.2. All we need is to specify the rate functions.

Example 1 Correspondent to the INAR(1) model given in (1.2) is the setting when ξ is Bernoulli $\mathbb{P}\{\xi = 1\} = 1 - \mathbb{P}\{\xi = 0\} = \alpha$ and ε has the Poisson distribution with $\mathbb{E}\varepsilon = (1-\alpha)\lambda$. In this case,

$$I_M(x) = \Big(\frac{1-\alpha}{1+\alpha}\Big)\frac{x^2}{2} \quad x \in \mathbb{R}.$$

To find $I_L(x)$, we need to maximize the function

$$f(\theta) = x\Big(\theta - \log\Big(1 - \alpha + \alpha e^\theta\Big)\Big) - (1-\alpha)\lambda\Big(e^\theta - 1\Big)$$

for $x > 0$. By a simple but tedious calculus, the maximizer is

$$\theta = \log\left(\frac{\sqrt{\lambda^2(1-\alpha)^2 + 4\lambda\alpha x} - \lambda(1-\alpha)}{2\lambda\alpha}\right).$$

That gives

$$I_L(x) = x\log\frac{4\lambda x}{\Big((1-\alpha)\lambda + \sqrt{\lambda^2(1-\alpha)^2 + 4\lambda\alpha x}\Big)^2} + \left(1 - \frac{2x}{\lambda(1-\alpha) + \sqrt{\lambda^2(1-\alpha)^2 + 4\lambda\alpha x}}\right) \quad x \ge 0.$$

One can directly exam that

$$I_L(\lambda) = 0 \quad \text{for} \quad \lambda = \frac{\mathbb{E}\varepsilon}{1 - \mathbb{E}\xi}.$$

Example 2 We now consider the INAR(1) model defined by (1.3) where ξ and ε are geometric distributed random variables with probability mass function given by

$$\mathbb{P}(\xi = k) = \mathbb{P}(\varepsilon = k) = \frac{\alpha^k}{(1+\alpha)^{k+1}} \quad k = 0, 1, 2, \ldots.$$

We have that $\mathbb{E}\xi = \mathbb{E}\varepsilon = \alpha$, $\text{Var}\,(\xi) = \text{Var}\,(\varepsilon) = \alpha$ and

$$\lambda \equiv \frac{\mathbb{E}\varepsilon}{1 - \mathbb{E}\xi} = \frac{\alpha}{1-\alpha}.$$

In particular,

$$I_M(x) = \frac{(1-\alpha)^3}{2\alpha(1+\alpha)} x^2 \quad x \in \mathbb{R}.$$

Further,

$$\mathbb{E}\exp\{\theta\xi\} = \mathbb{E}\exp\{\theta\varepsilon\} = \frac{1}{1+\alpha(1-e^\theta)}.$$

To compute $I_L(x)$, we maximize the function

$$f(\theta) = x\Big(\theta + \log\left(1+\alpha-\alpha e^\theta\right)\Big) + \log\left(1+\alpha-\alpha e^\theta\right)$$

for $x > 0$. Indeed, the maximizer is

$$\theta = \log\frac{(1+\alpha)x}{\alpha(2x+1)}$$

and

$$I_L(x) = x\log\frac{(1+\alpha)^2 x(x+1)}{\alpha(2x+1)^2} + \log\frac{(1+\alpha)(x+1)}{2x+1} \quad x \geq 0.$$

One can check that $I_L\left(\frac{\alpha}{1-\alpha}\right) = 0$.

3 Proof of Theorem 1.1

According to Gärtner-Ellis theorem, our proof relies on computing the limit

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\kappa\sum_{j=1}^n N_j\right\} \quad \kappa \in \mathbb{R}.$$

To this end, the function

$$\phi(\theta) \triangleq \theta - \log \mathbb{E} \exp\left\{\theta\xi\right\} \quad \theta \in \mathbb{R}$$

plays a crucial role. In the following, we shall list some analytic properties of $\phi(\theta)$ that are relevant to our work. First notice that $\phi(\theta)$ is a concave continuous function and $\phi(0) = 0$. By Jensen's inequality, $\log \mathbb{E}e^{\theta\xi} \geq \mathbb{E} \log e^{\theta\xi} = \theta\mathbb{E}\xi$. Consequently, $\phi(\theta) \leq \theta(1 - \mathbb{E}\xi)$. By the assumption that $\mathbb{E}\xi < 1$ we have that $\phi(\theta) < 0$ for all $\theta < 0$. In addition, by the fact that

$$\mathbb{E}e^{\theta\xi} = \mathbb{P}\{\xi = 0\} + e^{\theta} \sum_{k=1}^{\infty} \mathbb{P}\{\xi = k\}e^{(k-1)\theta} \sim \mathbb{P}\{\xi = 0\} \quad (\theta \to -\infty)$$

$\phi(\theta)$ is asymptotically linear in the negative direction

$$\phi(\theta) \sim \theta \quad (\theta \to -\infty) \quad \text{and} \quad \lim_{\theta \to -\infty} \phi'(\theta) = 1. \tag{3.1}$$

The fact that $\phi(0) = 0$ and $\phi'(0) = 1 - \mathbb{E}\xi > 0$ implies that $\phi(\theta)$ maintains to be positive and increasing at least in a right neighborhood of $\theta = 0$. The behavior of $\phi(\theta)$ on the far right of zero divides our model into the following two different cases.
Case 1 The only setting contained in this case is when ξ is Bernoulli: $\mathbb{P}\{\xi = 1\} = 1 - \mathbb{P}\{\xi = 0\} = \alpha$. In this case, $\phi(\theta) = \theta - \log\left((1-\alpha) + \alpha e^{\theta}\right)$ is strictly increasing on the real line and

$$\lim_{\theta \to \infty} \phi(\theta) = \log \frac{1}{\alpha}. \tag{3.2}$$

Case 2 The remaining setting, where $\mathbb{P}\{\xi = k\} > 0$ for some integer $k \geq 2$. Consequently, $\phi(\theta) \leq \theta - k\theta - \log \mathbb{P}\{\xi = k\}$ and the right hand side tends to $-\infty$ as $\theta \to \infty$. By the continuity and the concavity of $\phi(\theta)$, there are $0 < \theta_0 < \theta_1$ such that $\phi(\theta)$ is strictly increasing on $(0, \theta_0)$ (therefore on $(-\infty, \theta_0)$) and strictly decreasing on (θ_0, ∞); and that $\phi(\theta) > 0$ on $(0, \theta_1)$, $\phi(\theta) < 0$ on (θ_1, ∞) (therefore on $(-\infty, 0) \cup (\theta_1, \infty)$). Obviously, $\phi'(\theta_0) = 0$ and $\phi(\theta_1) = 0$.

Example 1 and Example 2 listed in Sect. 2 belong to Case 1 and Case 2, respectively. The graphs of $\phi(\theta)$ in these two examples are given in Fig. 1 below with $\alpha = 0.05$.

To have a uniform treatment for the both cases, we take $\theta_0 = \infty$, $\theta_1 = \infty$ and $\phi(\theta_0) = \log 1/\alpha$ in Case 1 in the following discussion.

The crucial step of our proof is to show

Lemma 3.1 *For any $\theta < \theta_0$,*

$$\lim_{n \to \infty} \frac{1}{n} \log \mathbb{E} \exp\left\{\phi(\theta) \sum_{j=1}^{n} N_j\right\} = \log \mathbb{E} \exp\{\theta\varepsilon\}. \tag{3.3}$$

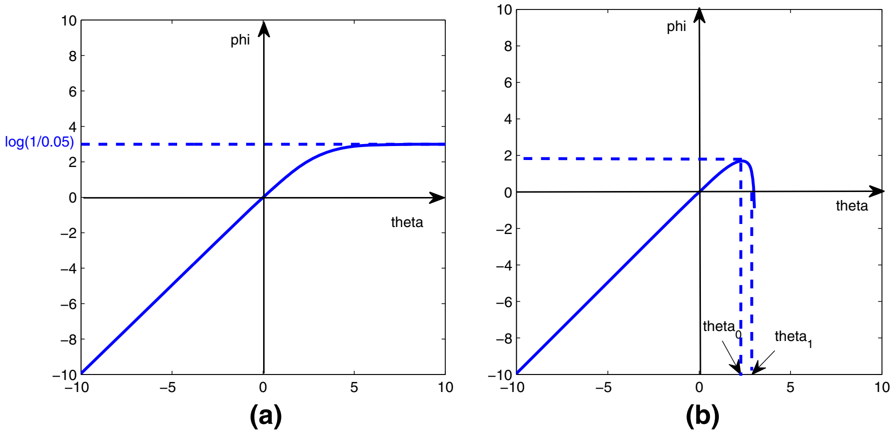

Fig. 1 Graphs of $\phi(\theta)$ in Case 1 (**a**) and in Case 2 (**b**)

Proof Write

$$L(\theta) = \log \mathbb{E} \exp \Big\{\theta\xi\Big\}$$

and let $\mathcal{F}_k = \sigma(N_1, \cdots, N_k)$ be the σ-field generated by $\{N_1, \cdots, N_k\}$. Notice that for any k,

$$\begin{aligned}
&\mathbb{E}\Big\{\exp\Big\{(\theta N_k - L(\theta)N_{k-1}\Big\}\Big|\mathcal{F}_{k-1}\Big\} \qquad (3.4)\\
&= \exp\Big\{-L(\theta)N_{k-1}\Big\}\mathbb{E}\Big\{\exp\Big\{\theta N_k\Big\}\Big|\mathcal{F}_{k-1}\Big\}\\
&= \exp\Big\{-L(\theta)N_{k-1}\Big\}\mathbb{E}\exp\Big\{\theta\varepsilon\Big\}\mathbb{E}\Big\{\exp\Big\{\theta\sum_{j=1}^{N_{k-1}}\xi_{k-1,j}\Big\}\Big|\mathcal{F}_{k-1}\Big\}\\
&= \exp\Big\{-L(\theta)N_{k-1}\Big\}\mathbb{E}\exp\Big\{\theta\varepsilon\Big\}\Big(\mathbb{E}\exp\Big\{\theta\xi\Big\}\Big)^{N_{k-1}}\\
&= \mathbb{E}\exp\Big\{\theta\varepsilon\Big\} \quad (\theta \in \mathbb{R}).
\end{aligned}$$

Consequently,

$$\mathbb{E}\exp\Big\{\sum_{k=1}^{n+1}(\theta N_k - L(\theta)N_{k-1})\Big\} = \Big(\mathbb{E}\exp\Big\{\theta\varepsilon\Big\}\Big)^{n+1}.$$

By the facts that

$$\sum_{k=1}^{n+1}(\theta N_k - L(\theta)N_{k-1}) = \theta N_{n+1} - L(\theta)N_0 + \phi(\theta)\sum_{j=1}^{n} N_j$$

and that N_0 is deterministic

$$\mathbb{E}\exp\left\{\theta N_{n+1}+\phi(\theta)\sum_{j=1}^{n}N_j\right\}=\exp\left\{L(\theta)N_0\right\}\Big(\mathbb{E}\exp\left\{\theta\varepsilon\right\}\Big)^{n+1}\quad \theta\in\mathbb{R}. \tag{3.5}$$

Our goal in the remaining argument is to remove the term θN_{n+1} from (3.5) without drastic dynamic change. Notice that $\theta_1<\infty$ in Case 2 and $\phi(\theta)\le 0$ as $\theta\ge\theta_1$. From (3.5)

$$\mathbb{E}\exp\left\{\theta N_{n+1}\right\}\ge\exp\left\{L(\theta)N_0\right\}\Big(\mathbb{E}\exp\left\{\theta\varepsilon\right\}\Big)^{n+1} \tag{3.6}$$

for all $\theta\ge\theta_1$. This shows that the routine exponential approximation by Hölder inequality is no longer working here as the term θN_{n+1} can be exponentially significant for large $\theta>0$.

First notice that (3.3) holds automatically for $\theta=0$ as $\phi(0)=0$. In the following, we treat the cases $0<\theta<\theta_0$ and $\theta<0$ separately.

Let $0<\theta<\theta_0$ be fixed. By (3.5)

$$\mathbb{E}\exp\left\{\big(\theta-L(\theta)\big)\sum_{j=1}^{n}N_j\right\}\le\exp\left\{L(\theta)N_0\right\}\Big(\mathbb{E}\exp\left\{\theta\varepsilon\right\}\Big)^{n+1}. \tag{3.7}$$

Consequently,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\big(\theta-L(\theta)\big)\sum_{j=1}^{n}N_j\right\}\le\log\mathbb{E}\exp\{\theta\varepsilon\}. \tag{3.8}$$

We now work on the lower bound. Let $\delta>0$ be small but fixed and let $m\ge 1$ be a large but fixed integer. Define the stopping time

$$\tau=\min\{k\ge n-m;\quad N_k\le\delta n\}.$$

Consider the decomposition

$$\begin{aligned}
&\mathbb{E}\exp\left\{\theta N_{n+1}+\big(\theta-L(\theta)\big)\sum_{j=1}^{n}N_j\right\}\\
&=\mathbb{E}\exp\left\{\theta N_{n+1}+\big(\theta-L(\theta)\big)\sum_{j=1}^{n}N_j\right\}1_{\{\tau\le n\}}\\
&\quad+\mathbb{E}\exp\left\{\theta N_{n+1}+\big(\theta-L(\theta)\big)\sum_{j=1}^{n}N_j\right\}1_{\{\tau>n\}}\\
&=A(n,m)+B(n,m)\quad\text{(say)}.
\end{aligned}$$

Applying (3.5) on the left hand side,

$$\log \mathbb{E} \exp \big\{\theta\varepsilon\big\} \le \max \left\{ \liminf_{n\to\infty} \frac{1}{n} \log A(n,m), \quad \limsup_{n\to\infty} \frac{1}{n} \log B(n,m) \right\}. \tag{3.9}$$

For each $n - m \le k \le n$, write

$$\theta N_{n+1} + \big(\theta - L(\theta)\big) \sum_{j=1}^{n} N_j = \theta N_k + \big(\theta - L(\theta)\big) \sum_{j=1}^{k-1} N_j + \sum_{j=k+1}^{n+1} \big(\theta N_j - L(\theta) N_{j-1}\big).$$

Applying (3.4) repeatedly,

$$\mathbb{E}\Bigg[\exp \Bigg\{ \sum_{j=k+1}^{n+1} \big(\theta N_j - L(\theta) N_{j-1}\big) \Bigg\} \Bigg| \mathcal{F}_k \Bigg] = \Big(\mathbb{E} \exp\{\theta\varepsilon\} \Big)^{n-k+1}.$$

Thus

$$\begin{aligned} &\mathbb{E} \exp \left\{ \theta N_{n+1} + \big(\theta - L(\theta)\big) \sum_{j=1}^{n} N_j \right\} 1_{\{\tau = k\}} \\ &= \Big(\mathbb{E} \exp\{\theta\varepsilon\} \Big)^{n-k+1} \mathbb{E} \exp \left\{ \theta N_k + \big(\theta - L(\theta)\big) \sum_{j=1}^{k-1} N_j \right\} 1_{\{\tau = k\}} \\ &\le \Big(\mathbb{E} \exp\{\theta\varepsilon\} \Big)^{n-k+1} e^{\theta\delta n} \mathbb{E} \exp \left\{ \big(\theta - L(\theta)\big) \sum_{j=1}^{k-1} N_j \right\} 1_{\{\tau = k\}}. \end{aligned}$$

Hence

$$\begin{aligned} A(n,m) &\le e^{\delta\theta n} \sum_{k=n-m}^{n} \Big(\mathbb{E} \exp\{\theta\varepsilon\} \Big)^{n-k+1} \mathbb{E} \exp \left\{ \big(\theta - L(\theta)\big) \sum_{j=1}^{k-1} N_j \right\} 1_{\{\tau = k\}} \\ &\le e^{\delta\theta n} \Big(\mathbb{E} \exp\{\theta\varepsilon\} \Big)^{m+1} \sum_{k=n-m}^{n} \mathbb{E} \exp \left\{ \big(\theta - L(\theta)\big) \sum_{j=1}^{n} N_j \right\} 1_{\{\tau = k\}} \\ &\le e^{\delta\theta n} \Big(\mathbb{E} \exp\{\theta\varepsilon\} \Big)^{m+1} \mathbb{E} \exp \left\{ \big(\theta - L(\theta)\big) \sum_{j=1}^{n} N_j \right\}. \end{aligned}$$

Consequently,

$$\liminf_{n\to\infty} \frac{1}{n} \log A(n,m) \le \delta\theta + \liminf_{n\to\infty} \frac{1}{n} \log \mathbb{E} \exp \left\{ \big(\theta - L(\theta)\big) \sum_{j=1}^{n} N_j \right\} \tag{3.10}$$

for any $m \ge 1$.

We now show that for large m, the quantity $B(n, m)$ is negligible. By (3.7) and the facts that $N_{n-m} > \delta n, \ldots, N_n > \delta n$ on $\{\tau > n\}$, and that $\theta - L(\theta) > 0$ on $\theta < \theta_0$,

$$\exp\Big\{L(\theta)N_0\Big\}\Big(\mathbb{E}\exp\big\{\theta\varepsilon\big\}\Big)^{n+1}$$
$$\geq \mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=n-m}^{n} N_j\right\}1_{\{\tau>n\}} \geq \exp\Big\{m\delta\big(\theta - L(\theta)\big)n\Big\}\mathbb{P}\{\tau > n\}.$$

Consequently,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{P}\{\tau > n\} \leq -m\delta\big(\theta - L(\theta)\big) + \log\mathbb{E}\exp\big\{\theta\varepsilon\big\}. \tag{3.11}$$

Given $A > 0$, the right hand side is less than $-A$ for a sufficiently large m.

Pick $\tilde{\theta} \in (\theta, \theta_0)$ and $p > 1$ such that $p\phi(\theta) \leq \phi(\tilde{\theta})$. Further, one can make $p\theta \leq \tilde{\theta}$. Let $q > 1$ be the conjugate of p. By Hölder inequality and (3.11)

$$B(n, m) \leq \left(\mathbb{E}\exp\left\{p\theta N_{n+1} + p\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\}\right)^{1/p}\Big(\mathbb{P}\{\tau > n\}\Big)^{1/q}$$
$$\leq \left(\mathbb{E}\exp\left\{\tilde{\theta}N_{n+1} + \big(\tilde{\theta} - L(\tilde{\theta})\big)\sum_{j=1}^{n} N_j\right\}\right)^{1/p}\exp\Big\{-q^{-1}An\Big\}$$
$$= e^{L(\tilde{\theta})N_0/p}\Big(\mathbb{E}\exp\{\tilde{\theta}\varepsilon\}\Big)^{\frac{n+1}{p}}\exp\Big\{-q^{-1}An\Big\},$$

where the last step follows from (3.5).

Thus

$$\limsup_{n\to\infty}\frac{1}{n}\log B(n, m) \leq \frac{1}{p}\log\mathbb{E}\exp\{\tilde{\theta}\varepsilon\} - q^{-1}A.$$

One may make the right hand side smaller than

$$\log\mathbb{E}\exp\{\theta\varepsilon\}$$

by making A sufficiently large. In view of (3.9) and (3.10), we conclude that

$$\delta\theta + \liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\} \geq \log\mathbb{E}\exp\{\theta\varepsilon\}.$$

Letting $\delta \to 0^+$ on the left hand side leads to the lower bound

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\} \geq \log\mathbb{E}\exp\{\theta\varepsilon\}. \tag{3.12}$$

Equations (3.8) and (3.12) together lead to (3.3) for $0 < \theta < \theta_0$.

Finally, we consider the case when $\theta < 0$. By (3.5), we have

$$\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\} \geq \exp\left\{L(\theta)N_0\right\}\Big(\mathbb{E}\exp\left\{\theta\varepsilon\right\}\Big)^{n+1},$$

which leads to

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\} \geq \log\mathbb{E}\exp\{\theta\varepsilon\}. \tag{3.13}$$

We now come to the upper bound. Let τ be defined as before.

$$\begin{aligned}&\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\}\\ &= \mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\}1_{\{\tau\leq n\}} + \mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\}1_{\{\tau> n\}}.\end{aligned}$$

Notice that $\theta - L(\theta) < 0$. By the previous estimate the second term is bounded by

$$\mathbb{P}\{\tau > n\} \leq \exp\{-An\}$$

and A can be sufficiently large as one makes m large. So the second term is negligible.

As for the first term, by (3.5)

$$\begin{aligned}&\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{n} N_j\right\}1_{\{\tau\leq n\}}\\ &\leq \sum_{k=n-m}^{n}\mathbb{E}\exp\left\{\big(\theta - L(\theta)\big)\sum_{j=1}^{k-1} N_j\right\}1_{\{\tau=k\}}\\ &\leq e^{-\theta\delta n}\sum_{k=n-m}^{n}\mathbb{E}\exp\left\{\theta N_k + \big(\theta - L(\theta)\big)\sum_{j=1}^{k-1} N_j\right\}\\ &= e^{-\theta\delta n}\exp\{L(\theta)N_0\}\sum_{k=n-m}^{n}\Big(\mathbb{E}\exp\{\theta\varepsilon\}\Big)^k\\ &\leq e^{-\theta\delta n}\exp\{L(\theta)N_0\}\sum_{k=n-m}^{\infty}\Big(\mathbb{E}\exp\{\theta\varepsilon\}\Big)^k\\ &= e^{-\theta\delta n}\exp\{L(\theta)N_0\}\Big(1-\Big(\mathbb{E}\exp\{\theta\varepsilon\}\Big)\Big)^{-1}\Big(\mathbb{E}\exp\{\theta\varepsilon\}\Big)^{n-m}\end{aligned}$$

given the fact that

$$\mathbb{E}\exp\{\theta\varepsilon\} < 1, \quad as \quad \theta < 0.$$

In summary of our computation,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\Big(\theta - L(\theta)\Big)\sum_{j=1}^{n}N_j\right\}1_{\{\tau\le n\}} \le -\delta\theta + \log\mathbb{E}\exp\{\theta\varepsilon\}.$$

Thus,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\Big(\theta - L(\theta)\Big)\sum_{j=1}^{n}N_j\right\} \le -\delta\theta + \log\mathbb{E}\exp\{\theta\varepsilon\}.$$

Letting $\delta \to 0^+$ on the right leads to the upper bound

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\Big(\theta - L(\theta)\Big)\sum_{j=1}^{n}N_j\right\} \le \log\mathbb{E}\exp\{\theta\varepsilon\}.$$

Together with (3.13), this proves (3.3) for $\theta < 0$. □

In Case 1, let $\phi^{-1}(\kappa)$ be the inverse function of $\phi(\theta)$ with $\phi(\theta)$ being viewed as a function on $(-\infty,\infty)$. In view of (3.1) and 3.2), the function $\phi^{-1}(\kappa)$ has the domain $(-\infty, \log 1/\alpha)$ and the range $(-\infty,\infty)$. Set

$$\Lambda(\kappa) \triangleq \begin{cases} \log\mathbb{E}\exp\left(\phi^{-1}(\kappa)\varepsilon\right) & \kappa < \log 1/\alpha \\ \infty & \kappa \ge \log 1/\alpha. \end{cases}$$

In Case 2, $\phi^{-1}(\kappa)$ is defined as the inverse of $\phi(\theta)$ with $\phi(\theta)$ being viewed as a function limited on $(-\infty,\theta_0]$. Therefore, the function $\phi^{-1}(\kappa)$ has the domain $(-\infty,\phi(\theta_0)]$ and the range $(-\infty,\theta_0]$. Set

$$\Lambda(\kappa) \triangleq \begin{cases} \log\mathbb{E}\exp\left(\phi^{-1}(\kappa)\varepsilon\right) & \kappa \le \phi(\theta_0) \\ \infty & \kappa > \phi(\theta_0). \end{cases}$$

A subtle difference between Case 1 and Case 2 is that $\Lambda(\kappa)$ is left continuous at the intersection point $\phi(\theta_0)$ in Case 2.

Lemma 3.2 *For any $\kappa \in \mathbb{R}$,*

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\kappa\sum_{k=1}^{n}N_k\right\} = \Lambda(\kappa). \tag{3.14}$$

Proof In light of Lemma 3.1, we need only to prove:

1. In Case 1,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\kappa\sum_{j=1}^{n}N_j\right\}=\infty \tag{3.15}$$

   for any $\kappa\geq\log 1/\alpha$.
2. In Case 2,

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{(\theta_0-L(\theta_0))\sum_{j=1}^{n}N_j\right\}=\log\mathbb{E}\exp\{\theta_0\varepsilon\}, \tag{3.16}$$

   and

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\kappa\sum_{j=1}^{n}N_j\right\}=\infty,\quad \kappa>\phi(\theta_0). \tag{3.17}$$

We now prove (3.15). By monotonicity we only consider the case when $\kappa=\phi(\infty)=\log 1/\alpha$. Since $\theta_0=\infty$ and $\kappa=\phi(\theta_0)>\phi(\theta)$ for all $\theta>0$. Consequently, by (3.3), we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\kappa\sum_{j=1}^{n}N_j\right\}\geq\log\mathbb{E}\exp\{\theta\varepsilon\}$$

for every $\theta>0$. Letting $\theta\to\infty$ on the right hand side leads to (3.15).

By letting $\theta\to\theta_0^-$ in (3.3), we conclude in Case 2 that

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\big(\theta_0-L(\theta_0)\big)\sum_{j=1}^{n}N_j\right\}\geq\log\mathbb{E}\exp\{\theta_0\varepsilon\}.$$

On the other hand, by (3.5), we have

$$\begin{aligned}\mathbb{E}\exp\left\{\big(\theta_0-L(\theta_0)\big)\sum_{j=1}^{n}N_j\right\}&\leq\mathbb{E}\exp\left\{\theta_0N_{n+1}+\big(\theta_0-L(\theta_0)\big)\sum_{j=1}^{n}N_j\right\}\\&=\exp\{L(\theta_0)N_0\}\Big(\mathbb{E}\exp\left\{\theta_0\varepsilon\right\}\Big)^{n+1}.\end{aligned}$$

Consequently,

$$\limsup_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\big(\theta_0-L(\theta_0)\big)\sum_{j=1}^{n}N_j\right\}\leq\log\mathbb{E}\exp\{\theta_0\varepsilon\}.$$

So we have proved (3.16). From (3.16) and Lemma 3.1, we have that

$$\lim_{n\to\infty}\frac{1}{n}\log\mathbb{E}\exp\left\{\kappa\sum_{j=1}^{n}N_j\right\}=\mathbb{E}\exp\big(\phi^{-1}(\kappa)\varepsilon\big),\quad \kappa\le\phi(\theta_0).$$

It remains to prove (3.17). Let $\kappa>\phi(\theta_0)$ be fixed and we use the argument by contradiction. For otherwise there would be a subsequence $\{n_k\}$ such that the limit

$$\tilde{\Lambda}(\kappa)=\lim_{k\to\infty}\frac{1}{n_k}\log\mathbb{E}\exp\left\{\kappa\sum_{j=1}^{n_k}N_j\right\}$$

exists and finite. For any $\kappa'<\phi(\theta_0)\equiv\kappa_0$, by Hölder inequality

$$\Lambda(\kappa_0)=\Lambda\Big(\frac{\kappa-\kappa_0}{\kappa-\kappa'}\kappa'+\frac{\kappa_0-\kappa'}{\kappa-\kappa'}\kappa\Big)\le\frac{\kappa-\kappa_0}{\kappa-\kappa'}\Lambda(\kappa')+\frac{\kappa_0-\kappa'}{\kappa-\kappa'}\tilde{\Lambda}(\kappa),$$

or

$$\tilde{\Lambda}(\kappa)\ge\Lambda(\kappa_0)+(\kappa-\kappa_0)\frac{\Lambda(\kappa_0)-\Lambda(\kappa')}{\kappa_0-\kappa'}=\Lambda(\kappa_0)+(\kappa-\kappa_0)\Lambda'(\hat{\kappa}),$$

where $\hat{\kappa}\in(\kappa',\kappa_0)$ and the second step follows from the mean value theorem. Notice that

$$\Lambda'(\hat{\kappa})=\Big(\mathbb{E}\exp\big\{\hat{\kappa}\varepsilon\big\}\Big)^{-1}\mathbb{E}\varepsilon\exp\big\{\hat{\kappa}\varepsilon\big\}\frac{1}{\phi'(\hat{\theta})}\longrightarrow\infty$$

as $\kappa'\to\kappa_0^-$ and therefore $\phi'(\hat{\theta})\to\phi'(\theta_0)=0$. Therefore, we must have $\tilde{\Lambda}(\kappa)=\infty$ that leads to contradiction. □

Proof of Theorem 1.1 We need to do two things in the proof: Identify the rate function given in Theorem 1.1 with the function given as

$$I_L(x)=\sup_{\kappa\in\mathbb{R}}\Big\{\kappa x-\Lambda(\kappa)\Big\}. \tag{3.18}$$

and apply Gärtner-Ellis theorem (see. e.g., Theorem 2.3.6, [8]) based on Lemma 3.2.

We first consider Case 1. By the fact that $\Lambda(\kappa)=\infty$ for $\kappa\ge\log 1/\alpha$,

$$\sup_{\kappa\in\mathbb{R}}\Big\{\kappa x-\Lambda(\kappa)\Big\}=\sup_{\kappa<\log 1/\alpha}\Big\{\kappa x-\Lambda(\kappa)\Big\}.$$

The substitution $\kappa=\phi(\theta)$ leads to

$$\sup_{\kappa<\log 1/\alpha}\Big\{\kappa x-\Lambda(\kappa)\Big\}=\sup_{\theta\in\mathbb{R}}\Big\{\phi(\theta)x-\Lambda\big(\phi(\theta)\big)\Big\}=\sup_{\theta\in\mathbb{R}}\Big\{\phi(\theta)x-\log\mathbb{E}\exp\{\theta\varepsilon\}\Big\}$$

where the last step follows from the definition of $\Lambda(\kappa)$.

We now come to Case 2. Notice that $\Lambda(\kappa) = \infty$ for $\kappa > \phi(\theta_0)$. Thus,

$$\sup_{\kappa\in\mathbb{R}} \Big\{\kappa x - \Lambda(\kappa)\Big\} = \sup_{\kappa\le\phi(\theta_0)} \Big\{\kappa x - \Lambda(\kappa)\Big\}.$$

Consequently, the substitution $\kappa = \phi(\theta)$ $(\theta \le \theta_0)$ leads to

$$\sup_{\kappa\in\mathbb{R}} \Big\{\kappa x - \Lambda(\kappa)\Big\} = \sup_{\theta\le\theta_0} \Big\{\phi(\theta)x - \log\mathbb{E}\exp\{\theta\varepsilon\}\Big\}.$$

Therefore, all we need is to show that for any $x \ge 0$, the function

$$f(\theta) = \phi(\theta)x - \log\mathbb{E}\exp\{\theta\varepsilon\}$$

can not approximate its global supremum on $(-\infty, \infty)$ when θ is limited to (θ_0, ∞). Indeed, this is obvious as on (θ_0, ∞), $\phi(\theta)$ is decreasing while $\log\mathbb{E}\exp\{\theta\varepsilon\}$ is increasing.

We are in the position to apply Gärtner-Ellis Theorem with the rate function $I_L(x)$ given in (3.18). By Part (a) of Theorem 2.3.6, p. 44 in Dembo-Zeitouni [8], Lemma 3.2 leads to the upper bound

$$\limsup_{n\to\infty} \frac{1}{n}\log\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} N_k \in F\right) \leqslant -\inf_{x\in F} I_L(x)$$

for each close set $F \subset \mathbb{R}$.

The lower bound is harder to get. In addition to the exponential asymptotics given in Lemma 3.2, it requires some extra conditions on $\Lambda(\kappa)$ known as essential smoothness (Definition 2.3.5, p. 44, [8]). Recall that a convex function $\Lambda\colon R \longrightarrow (-\infty, \infty]$ is called essentially smooth if its domain $D_\lambda = \{\kappa \in \mathbb{R};\quad \Lambda(\kappa) < \infty\}$ has a non-empty interior $D_\lambda^0 = (a, b)$, if $\Lambda(\kappa)$ is differentiable on D_λ^0, and if

$$\lim_{\kappa\to a^+} |\Lambda'(\kappa)| = \infty \quad \text{and} \quad \lim_{\kappa\to b^-} |\Lambda'(\kappa)| = \infty. \tag{3.19}$$

The property in (3.19) is called the steepness at the domain boundary. Back to our setting, it is not hard to see that $D_\lambda^0 = \big(-\infty, \phi(\theta_0)\big)$. Here we recall our convention that $\theta_0 = \infty$ and $\phi(\theta_0) = \log 1/\alpha$ in Case 1. Further, $\Lambda(\kappa)$ is differentiable on D_λ^0 with

$$\Lambda'(\kappa) = \frac{\mathbb{E}\varepsilon\exp\{\phi^{-1}(\kappa)\varepsilon\}}{\mathbb{E}\exp\{\phi^{-1}(\kappa)\varepsilon\}}\frac{1}{\phi'\big(\phi^{-1}(\kappa)\big)}. \tag{3.20}$$

By the fact that $\lim_{\theta\to\theta_0^-}\phi'(\theta) = 0$, $\Lambda(\kappa)$ is steep at its right boundary. More precisely, one can see that

$$\lim_{\kappa\to\phi(\theta_0)^-} \Lambda'(\kappa) = \infty. \tag{3.21}$$

By (3.1), on the other hand, $\phi^{-1}(\kappa) \to -\infty$ and $\phi'\big(\phi^{-1}(\kappa)\big) \to 1$ as $\kappa \to -\infty$. In addition, by L'Hôpital's rule

$$\lim_{q\to\infty} \frac{\mathbb{E}\varepsilon \exp\{-q\varepsilon\}}{\mathbb{E}\exp\{-q\varepsilon\}} = -\lim_{q\to\infty} \frac{1}{q} \log \mathbb{E}\exp\{-q\varepsilon\} = k_0$$

where $k_0 \geq 0$ is the least integer with $\mathbb{P}\{\varepsilon = k_0\} > 0$. Summarizing our computation,

$$\lim_{\kappa\to-\infty} \Lambda'(\kappa) = k_0. \tag{3.22}$$

This implies that $\Lambda(\kappa)$ is not steep at the left boundary of its domain.

To fix this problem, we adopt a strategy that goes back to at least Varadhan (p. 11, [15]). Let $\{g_k, k \in \mathbb{N}^+\}$ be an i.i.d. sequence of $N(0, \gamma^2)$-random variables that are independent of $\{N_k\}$. Here $\gamma^2 > 0$ is small but fixed.

$$\begin{aligned}\mathbb{E}\exp\left\{\kappa \sum_{k=1}^n (N_k + g_k)\right\} &= \Big\{\mathbb{E}\exp(\kappa g_1)\Big\}^n \mathbb{E}\exp\left\{\kappa \sum_{k=1}^n N_k\right\} \\ &= \exp\{n\kappa^2\gamma^2/2\}\mathbb{E}\exp\left\{\kappa \sum_{k=1}^n N_k\right\}.\end{aligned}$$

Therefore,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{E}\exp\left\{\kappa \sum_{k=1}^n (N_k + g_k)\right\} = \frac{\gamma^2}{2}\kappa^2 + \Lambda(\kappa) \triangleq \Lambda_\gamma(\kappa).$$

Clearly, $\mathcal{D}_{\Lambda_\gamma} = \mathcal{D}_\Lambda$, $\Lambda_\gamma(\kappa)$ is differentiable throughout $\mathcal{D}^0_{\Lambda_\gamma}$ with $\Lambda'_\gamma(\kappa) = \kappa\gamma^2 + \Lambda'(\kappa)$. Most importantly, in connection to (3.21) and (3.22), $\Lambda_\gamma(\kappa)$ is steep at the left and right boundaries of its domain! Hence $\Lambda_\gamma(\kappa)$ is essentially smooth.

By Gärtner-Ellis Theorem (Part (c), Theorem 2.3.6, p.44, [8]), the auxiliary random sequence

$$\frac{1}{n} \sum_{k=1}^n (N_k + g_k)$$

satisfies the large deviation principle with the rate function $I_\gamma(x) = \sup_{\kappa\in\mathbb{R}} \Big\{\kappa x - \Lambda_\gamma(\kappa)\Big\}$. In particular, the lower bound claims that

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\frac{1}{n} \sum_{k=1}^n (N_k + g_k) \in G\right) \geq -\inf_{x\in G} I_\gamma(x)$$

for every open set $G \subset \mathbb{R}^+$.

The obvious relation $\Lambda_\gamma(\cdot) \geq \Lambda(\cdot)$ implies that $I_\gamma(\cdot) \leq I_L(\cdot)$. Given $x_0 > 0$ and $\delta > 0$, applying the above lower bound to the open set $G = (x - 2^{-1}\delta, x + 2^{-1}\delta)$ leads to

$$\liminf_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(\left|\frac{1}{n}\sum_{k=1}^{n}(N_k + g_k) - x_0\right| < \frac{\delta}{2}\right) \geq - \inf_{|x-x_0|<\frac{\delta}{2}} I_L(x) \geq -I_L(x_0). \tag{3.23}$$

On the other hand,

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} N_k - x_0\right| < \delta\right\} &\geq \mathbb{P}\{\left|\frac{1}{n}\sum_{k=1}^{n}(N_k + g_k) - x_0\right| < \frac{\delta}{2}, \quad |\frac{1}{n}\sum_{k=1}^{n} g_k| < \frac{\delta}{2}\} \\ &\geq \mathbb{P}\{|\frac{1}{n}\sum_{k=1}^{n}(N_k + g_k) - x_0| < \frac{\delta}{2}\} - \mathbb{P}\{|\frac{1}{n}\sum_{k=1}^{n} g_k| \geq \frac{\delta}{2}\}\end{aligned}$$

and

$$\mathbb{P}\left\{\left|\frac{1}{\sqrt{n}}\sum_{k=1}^{n} g_k\right| \geq x\right\} = \exp\left\{-\frac{1+o(1)}{2\gamma^2}x^2\right\} \quad (x \to \infty).$$

Thus, we have

$$\begin{aligned}\mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n}(N_k + g_k) - x_0\right| < \frac{\delta}{2}\right\} &\leq \mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} N_k - x_0\right| < \delta\right\} \\ &+ \exp\{-\frac{1+o(1)}{2\gamma^2}\left(\frac{\sqrt{n}\delta}{2}\right)^2\}.\end{aligned} \tag{3.24}$$

Equations (3.23) and (3.24) together prove that

$$\max\left\{\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{|\frac{1}{n}\sum_{k=1}^{n} N_k - x_0| < \delta\right\}, \quad -\frac{1}{2\gamma^2}(\delta/2)^2\right\} \geq -I_L(x_0).$$

Letting $\gamma^2 \to 0^+$ on the left hand side, we have

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{|\frac{1}{n}\sum_{k=1}^{n} N_k - x_0| < \delta\right\} \geq -I_L(x_0). \tag{3.25}$$

Let $G \subset \mathbb{R}^+$ be an arbitrary open set again. For any $x_0 \in G$ there is a $\delta > 0$ such that $(x_0 - \delta, x_0 + \delta) \subset G$. Hence

$$\mathbb{P}\left\{\frac{1}{n}\sum_{k=1}^{n} N_k \in G\right\} \geq \mathbb{P}\left\{\left|\frac{1}{n}\sum_{k=1}^{n} N_k - x_0\right| < \delta\right\}.$$

Applying (3.25) to the right hand side

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{\frac{1}{n}\sum_{k=1}^{n}N_k\in G\right\}\geq -I_L(x_0).$$

Taking supremum over $x_0\in G$, we finally establish the desired lower bound

$$\liminf_{n\to\infty}\frac{1}{n}\log\mathbb{P}\left\{\frac{1}{n}\sum_{k=1}^{n}N_k\in G\right\}\geq -\inf_{x\in G}I_L(x).$$ □

Proof of Lemma 1.3: Some of the approaches used below may be standard. We include them here for the reader's convenience. Recall that

$$I_L(x)=\sup_{\theta\in\mathbb{R}}\left\{x\left(\theta-\log\mathbb{E}e^{\theta\xi}\right)-\log\mathbb{E}\exp\{\theta\varepsilon\}\right\}\quad x\geq 0.$$

Since the quantity inside the supremum is equal to zero as $\theta=0$, so we have $I_L(x)\geq 0$. In addition, by Jensen's inequality,

$$\log\mathbb{E}\exp\{\theta\xi\}\geq\mathbb{E}\log e^{\theta\xi}=\theta\mathbb{E}\xi,\quad \log\mathbb{E}\exp\{\theta\varepsilon\}\geq\mathbb{E}\log e^{\theta\varepsilon}=\theta\mathbb{E}\varepsilon$$

Thus,

$$\frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}\left(\theta-\log\mathbb{E}e^{\theta\xi}\right)-\log\mathbb{E}\exp\{\theta\varepsilon\}\leq 0$$

for any $\theta\in\mathbb{R}$. Consequently, $I_L\left(\frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}\right)=0$.

Assume, on the other hand, that $x\in\mathbb{R}$ satisfies $I_L(x)=0$. Then we have that

$$x\left(\theta-\log\mathbb{E}e^{\theta\xi}\right)-\log\mathbb{E}\exp\{\theta\varepsilon\}\leq 0.$$

Or

$$x\left(\theta-\log\mathbb{E}e^{\theta\xi}\right)\leq\log\mathbb{E}\exp\{\theta\varepsilon\}$$

for every $\theta\in\mathbb{R}$. In particular,

$$x\left(1-\frac{1}{\theta}\log\mathbb{E}e^{\theta\xi}\right)\leq\frac{1}{\theta}\log\mathbb{E}\exp\{\theta\varepsilon\}$$

for $\theta>0$; and

$$x\left(1-\frac{1}{\theta}\log\mathbb{E}e^{\theta\xi}\right)\geq\frac{1}{\theta}\log\mathbb{E}\exp\{\theta\varepsilon\}$$

for $\theta < 0$. Letting $\theta \to 0^+$ and $\theta \to 0^-$, respectively, we obtain that $x(1-\mathbb{E}\xi) \le \mathbb{E}\varepsilon$ and $x(1-\mathbb{E}\xi) \ge \mathbb{E}\varepsilon$. So we have $x = \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}$. We have completed the proof of (1).

We now come to (2). By the assumption we have that

$$\log \mathbb{E}_0 \exp\{\theta\varepsilon\} \le k_0\theta \quad \theta < 0.$$

Combining this with the first relation in (3.1), therefore,

$$x\phi(\theta) - \log \mathbb{E}_0 \exp\{\theta\varepsilon\} \longrightarrow \infty \quad (\theta \to \infty)$$

for $x < k_0$. Consequently, $I_L(x) = \infty$.

When $k_0 = 0$, the function

$$f(\theta) = -\log \mathbb{E} \exp\{\theta\varepsilon\} \quad \theta \in \mathbb{R}$$

is decreasing with the limit $\log\big(\mathbb{P}\{\varepsilon = 0\}\big)^{-1}$ as $\theta \to -\infty$. So $I_L(0) = \log\big(\mathbb{P}\{\varepsilon = 0\}\big)^{-1}$.

We now prove (3). The lower-semi-continuity follows directly from the fact that $I_L(x)$ is a convex conjugate of $\Lambda(\kappa)$ (see (3.18)). By the lower-semi-continuity the level set $\{x;\quad I_L(x) \le l\}$ is a close set. We just need to prove that the level set is bounded. First notice that there is a $M > 0$ such that

$$c \equiv \sup_{|\kappa| \le M} \Lambda(\kappa) < \infty.$$

From (3.18),

$$I_L(x) \ge \kappa x - \Lambda(\kappa) \ge \kappa x - c \quad \text{for any } \kappa \text{ satisfying} \quad |\kappa| \le M.$$

In particular, $I_L(x) \ge Mx - c$ for all $x \ge 0$. This implies that

$$\big\{x;\quad I_L(x) \le l\big\} \subset \big\{x;\quad 0 \le x \le M^{-1}(l+c)\big\}.$$

In other words, $\big\{x;\quad I_L(x) \le l\big\}$ is bounded. □

4 Proof of Theorem 1.2

Write $\lambda = \frac{\mathbb{E}\varepsilon}{1-\mathbb{E}\xi}$. By Gärtner-Ellis theorem (Theorem 2.3.6, p.44, [8]), all we need to show is that

$$\lim_{n\to\infty} \frac{1}{b_n} \log \mathbb{E} \exp\left\{\beta\sqrt{\frac{b_n}{n}}\sum_{k=1}^n (N_k - \lambda)\right\} = \frac{1}{2}\sigma^2\beta^2 \quad \beta \in \mathbb{R}. \tag{4.1}$$

Let $\beta \in \mathbb{R}$ be fixed but arbitrary and write

$$l_n = \log \mathbb{E} \exp\Big\{\theta_n(\xi - \mathbb{E}\xi)\Big\},$$

where $\theta_n = \sqrt{\frac{b_n}{n}}\frac{\beta}{1-\mathbb{E}\xi}$. Notice that

$$N_k - \lambda = \sum_{j=1}^{N_{k-1}}(\xi_{k-1,j} - \mathbb{E}\xi) + (\mathbb{E}\xi)(N_{k-1} - \lambda) + (\varepsilon_k - \mathbb{E}\varepsilon) \quad k = 1, 2, \cdots .$$

Recall that $\mathcal{F}_k = \sigma\{N_1, \cdots, N_k\}$ and observe that for any $k \geq 1$,

$$\begin{aligned}
&\mathbb{E}\Big[\exp\Big\{\theta_n\Big((N_k - \lambda) - (\mathbb{E}\xi)(N_{k-1} - \lambda)\Big) - l_n N_{k-1}\Big\}\Big|\mathcal{F}_{k-1}\Big]\\
&= \exp\Big\{-l_n N_{k-1}\Big\}\mathbb{E}\Big[\exp\Big\{\theta_n\Big(\sum_{j=1}^{N_{k-1}}(\xi_{k-1,j} - \mathbb{E}\xi) + (\varepsilon_k - \mathbb{E}\varepsilon)\Big)\Big\}\Big|\mathcal{F}_{k-1}\Big]\\
&= \mathbb{E}\exp\big\{\theta_n(\varepsilon - \mathbb{E}\varepsilon)\big\}.
\end{aligned}$$

Hence,

$$\begin{aligned}
&\mathbb{E}\exp\Big\{\sum_{k=1}^{n+1}\Big\{\theta_n\Big((N_k - \lambda) - (\mathbb{E}\xi)(N_{k-1} - \lambda)\Big) - l_n N_{k-1}\Big\}\Big\}\\
&= \Big(\mathbb{E}\exp\big\{\theta_n\big(\varepsilon - \mathbb{E}\varepsilon\big)\big\}\Big)^{n+1}.
\end{aligned} \tag{4.2}$$

On the other hand,

$$\begin{aligned}
&\mathbb{E}\exp\Big\{\sum_{k=1}^{n+1}\Big\{\theta_n\Big((N_k - \lambda) - (\mathbb{E}\xi)(N_{k-1} - \lambda)\Big) - l_n N_{k-1}\Big\}\Big\}\\
&= \exp\Big\{-(n+1)\lambda l_n\Big\}\mathbb{E}\Big\{\exp\Big\{\theta_n(N_{n+1} - \lambda) - \Big((\mathbb{E}\xi)\theta_n + l_n\Big)(N_0 - \lambda)\Big\}\\
&\qquad\qquad \times \exp\Big\{\Big(\theta_n(1 - \mathbb{E}\xi) - l_n\Big)\sum_{k=1}^{n}(N_k - \lambda)\Big\}\Big\} \qquad (4.3)\\
&= \Big(1 + o(1)\Big)\exp\Big\{-n\lambda l_n\Big\}\mathbb{E}\exp\Big\{\theta_n N_{n+1} + \Big(\theta_n(1 - \mathbb{E}\xi) - l_n\Big)\sum_{k=1}^{n}(N_k - \lambda)\Big\}.
\end{aligned}$$

Here we use "$\big(1 + o(1)\big)$" to absorb all insignificant and deterministic factors.

Combining (4.2) and (4.3) and by the definition of l_n

$$\begin{aligned}
&\Big(\mathbb{E}\exp\big\{\theta_n(\varepsilon - \mathbb{E}\varepsilon)\big\}\Big)^{n+1}\Big(\mathbb{E}\exp\Big\{\theta_n(\xi - \mathbb{E}\xi)\Big\}\Big)^{n\lambda}\\
&= \Big(1 + o(1)\Big)\mathbb{E}\exp\Big\{\theta_n N_{n+1} + \Big(\theta_n(1 - \mathbb{E}\xi) - l_n\Big)\sum_{k=1}^{n}(N_k - \lambda)\Big\}.
\end{aligned}$$

By Taylor expansion, the left hand side is asymptotically equivalent to

$$\exp\left\{\frac{1}{2}\sigma^2\beta^2 b_n\right\}.$$

Thus

$$\lim_{n\to\infty}\frac{1}{b_n}\log\mathbb{E}\exp\left\{\theta_n N_{n+1}+\Big(\theta_n(1-\mathbb{E}\xi)-l_n\Big)\sum_{k=1}^n(N_k-\lambda)\right\}=\frac{1}{2}\sigma^2\beta^2.$$

By the fact that $\theta_n\to 0$ and by Lemma 4.1 below, a standard argument of exponential approximation by Hölder inequality enables us to remove the term $\theta_n N_{n+1}$ from the above equation. So we have

$$\lim_{n\to\infty}\frac{1}{b_n}\log\mathbb{E}\exp\left\{\Big(\theta_n(1-\mathbb{E}\xi)-l_n\Big)\sum_{k=1}^n(N_k-\lambda)\right\}=\frac{1}{2}\sigma^2\beta^2. \tag{4.4}$$

In addition, by Jensen's inequality,

$$\mathbb{E}\exp\Big\{\theta_n(\xi-\mathbb{E}\xi)\Big\}\ge\exp\Big\{\theta_n\mathbb{E}(\xi-\mathbb{E}\xi)\Big\}=1.$$

Consequently, $l_n\ge 0$. On the other hand, $l_n\sim 2^{-1}\mathrm{Var}(\xi)\theta_n^2=o(\theta_n)$. By Hölder inequality, therefore,

$$\begin{aligned}&\mathbb{E}\exp\left\{\Big(\theta_n(1-\mathbb{E}\xi)-l_n\Big)\sum_{k=1}^n(N_k-\lambda)\right\}\\&\le\left(\mathbb{E}\exp\left\{\theta_n(1-\mathbb{E}\xi)\sum_{k=1}^n(N_k-\lambda)\right\}\right)^{\frac{\theta_n(1-\mathbb{E}\xi)-l_n}{\theta_n(1-\mathbb{E}\xi)}}.\end{aligned}$$

By the fact that $\theta_n(1-\mathbb{E}\xi)=\beta\sqrt{\frac{b_n}{n}}$ and by (4.4), we obtain the lower bound

$$\liminf_{n\to\infty}\frac{1}{b_n}\log\mathbb{E}\exp\left\{\beta\sqrt{\frac{b_n}{n}}\sum_{k=1}^n(N_k-\lambda)\right\}\ge\frac{1}{2}\sigma^2\beta^2. \tag{4.5}$$

On the other hand, given a small number $0<\delta<1$,

$$\theta_n(1-\mathbb{E}\xi)-l_n>(1-\delta)\theta_n(1-\mathbb{E}\xi)=(1-\delta)\beta\sqrt{\frac{b_n}{n}}$$

as n is sufficiently large. By Hölder inequality

$$\mathbb{E}\exp\left\{(1-\delta)\beta\sqrt{\frac{b_n}{n}}\sum_{k=1}^{n}(N_k-\lambda)\right\}$$
$$\leq\left(\mathbb{E}\exp\left\{\Big(\theta_n(1-\mathbb{E}\xi)-l_n\Big)\sum_{k=1}^{n}(N_k-\lambda)\right\}\right)^{\frac{(1-\delta)\theta_n(1-\mathbb{E}\xi)}{\theta_n(1-\mathbb{E}\xi)-l_n}}.$$

By (4.4), therefore,

$$\limsup_{n\to\infty}\frac{1}{b_n}\log\mathbb{E}\exp\left\{(1-\delta)\beta\sqrt{\frac{b_n}{n}}\sum_{k=1}^{n}(N_k-\lambda)\right\}\leq\frac{1}{2}\sigma^2\beta^2.$$

Since $\beta\in\mathbb{R}$ can be arbitrary, replacing it by $(1-\delta)^{-1}\beta$ in the above leads to

$$\limsup_{n\to\infty}\frac{1}{b_n}\log\mathbb{E}\exp\left\{\beta\sqrt{\frac{b_n}{n}}\sum_{k=1}^{n}(N_k-\lambda)\right\}\leq\frac{1}{2}\sigma^2\Big(\frac{\beta}{1-\delta}\Big)^2.$$

Letting $\delta\to 0^+$ on the right hand side yields the desired upper bound which, together with the lower bound (4.5), leads to (4.1). □

Recall our notation $L(\theta)=\log\mathbb{E}\exp\{\theta\xi\}$ and our discussion on the function $\phi(\theta)=\theta-L(\theta)$ in the beginning of the previous section. In Case 1 where ξ is Bernoulli, $\phi(\theta)$ is positive and strictly increasing on $(0,\infty)$; while in the remaining setting (Case 2), there is $\theta_0>0$ such that $\phi(\theta)$ is positive and strictly increasing on $(0,\theta_0)$ but is decreasing on (θ_0,∞). To have a uniform statement in the following lemma. We use the convention that $\theta_0=\infty$ in Case 1. Sharply contrary to (3.6) where $\theta>0$ is large, we have

Lemma 4.1 *For any* $0<\theta<\theta_0$,

$$\sup_{k\geq 1}\mathbb{E}\exp\left\{\theta N_k\right\}<\infty.$$

Proof We start with the fact that the relation (1.1) implies that

$$\mathbb{E}\exp\left\{\theta N_k\right\}=\Big(\mathbb{E}\exp\{\theta\varepsilon\}\Big)\mathbb{E}\exp\left\{L(\theta)N_{k-1}\right\}$$

for any $\theta\in\mathbb{R}$. Iterating the above argument gives that

$$\mathbb{E}\exp\left\{\theta N_k\right\}=\left(\prod_{j=0}^{k-1}\mathbb{E}\exp\{L^j(\theta)\varepsilon\}\right)\exp\left\{L^k(\theta)N_0\right\}\tag{4.6}$$

where the notation $L^j(\cdot)$ is for the j-th fold composition of the function $L(\cdot)$.

Fix $0 < \theta < \theta_0$. The derivative function $\phi'(\cdot)$ is strictly positive on $[0, \theta]$. Moreover, there is $\delta > 0$ such that $\phi'(\zeta) \geq \delta$ for any $\zeta \in [0, \theta]$. In other words, $L'(\zeta) \leq 1 - \delta$ on $[0, \theta]$. Notice that $L(0) = 0$. By the mean value theorem, $L(\zeta) \leq (1 - \delta)\zeta$ for any $\zeta \in [0, \theta]$. In particular, $L^j(\theta) \leq (1 - \delta)^j \theta$ for every integer $j \geq 1$. By (4.6), therefore,

$$\mathbb{E} \exp\left\{\theta N_k\right\} = \left(\prod_{j=0}^{k-1} \mathbb{E} \exp\{(1-\delta)^j \theta \varepsilon\}\right) \exp\left\{(1-\delta)^k \theta N_0\right\}.$$

By Hölder inequality,

$$\prod_{j=0}^{k-1} \mathbb{E} \exp\{(1-\delta)^j \theta \varepsilon\} \leq \prod_{j=0}^{k-1} \left(\mathbb{E} \exp\{\theta \varepsilon\}\right)^{(1-\delta)^j} \leq \left(\mathbb{E} \exp\{\theta \varepsilon\}\right)^{\sum_{j=0}^{k-1}(1-\delta)^j}$$
$$\leq \left(\mathbb{E} \exp\{\theta \varepsilon\}\right)^{\sum_{j=0}^{\infty}(1-\delta)^j} = \left(\mathbb{E} \exp\{\theta \varepsilon\}\right)^{1/\delta}.$$

Thus, the requested conclusion follows from the obvious boundedness of the factor $\exp\left\{(1-\delta)^k \theta N_0\right\}$ on k.

Acknowledgements This work is supported by National Natural Science Foundation of China (Nos. 11271155, 11371168, J1310022, 11571138, 11501241, 11571051), Science and Technology Research Program of Education Department in Jilin Province for the 12th Five-Year Plan (440020031139) and Jilin Province Natural Science Foundation (20130101066JC, 20130522102JH, 20150520053JH). Research partially supported by the Simons Foundation (No. 244767) and by the "1000 Talents Plan" from Jilin University, Jilin Province and Chinese government and supported by Science and Technology Research Program of Qiqihar city (N0. RKX-201513).

References

1. Al-Osh, M.A., Alzaid, A.A.: First-order integer-valued autoregressive (INAR(1)) process. J. Time Ser. Anal. **8**, 261–275 (1987)
2. Al-Osh, M.A., Alzaid, A.A.: Frist-order autoregressive time series with negative binomial and geometric marginals. Commun. Stat. Theory Methods **21**, 2483–2492 (1992)
3. Alzaid, A.A., Al-Osh, M.A.: First-order integer-valued autoregressive (INAR(1)) processes: distributional and regression properties. Stat. Neerl. **42**, 53–61 (1988)
4. Athreya, K.B., Ney, P.E.: Branching Processes. Springer-Verlag, New York (1972)
5. Brockwell, P.J., Davis, R.A.: Time Series: Theory and Methods, 2nd edn. Springer-Verlag, New York (1991)
6. Chen, X.: Limit Theorems for Functionals of Ergodic Markov Chains with General State Space, vol. 139(664). Memoirs of American Mathematical Society, Providence (1999)
7. de Acosta, A., Ney, P.: Large Deviations for Additive Functionals of Markov Chains, vol. 228(1070). American Mathematical Society, Providence (2014)
8. Dembo, A., Zeitouni, O.: Large Deviations Techniques and Applications, 2nd edn. Springer-Verlag, New York (1998)
9. Jung, R.C., Tremayne, A.R.: Binomial thinning models for integer time series. Stat. Model. **6**, 81–96 (2006a)
10. Jung, R.C., Tremayne, A.R.: Coherent forecasting in integer time series models. Int. J. Forecast. **22**, 223–238 (2006b)

11. Mckenzie, E.: Some ARMA models for dependence sequence of Poisson counts. Adv. Appl. Probab. **20**, 822–835 (1988)
12. Pakes, A.G.: Branching processes with immigration. J. Appl. Probab. **8**, 32–42 (1971)
13. Ristić, M.M., Bakouch, H.S., Nastić, A.S.: A new geometric first-order integer-valued autoregressive(NGINAR(1)) process. J. Stat. Plan. Inference **139**, 2218–2226 (2009)
14. Steutal, F., Van Harn, K.: Discrete analogues of self-decomposability and stability. Ann. Probab. **7**, 893–899 (1979)
15. Varadhan, S.R.S.: Large Deviations and Applications. CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 46. Society for Industrial and Applied Mathematics (SIAM), Philadelphia (1984)
16. Weiss, C.H.: Thinning operations for modeling time series of counts -a survey. Adv. Stat. Anal. **92**, 319–341 (2008)