Filomat 35:10 (2021), 3303–3317
https://doi.org/10.2298/FIL2110303Z

Published by Faculty of Sciences and Mathematics,
University of Niš, Serbia
Available at: http://www.pmf.ni.ac.rs/filomat

Deviations for Random Sums Indexed by the Generations of a Branching Process

Yanjiao Zhu[a], Zhenlong Gao[a]

[a] *School of Statistics, Qufu Normal University, Qufu 273165, P. R. China*

Abstract. Applying the results about harmonic moments of classical Galton–Watson process, we obtain the deviations for random sums indexed by the generations of a branching process. Our results show that the decay rates of large deviations and moderate deviations depend heavily on the degree of the heavy tail and the asymptotic distributions depend heavily on the normalizing constants. If the underlying Galton–Watson process belongs to the Schröder case, both large deviation and moderate deviation probabilities show three decay rates, where the critical case depends heavily on the Schröder index. Else if the Galton–Watson process belongs to the Böttcher case, there are only two decay rate for both large deviation and moderate deviation probabilities. Simulations are also given to illustrate our results.

1. Introduction

In the last two decades, large deviation theory has been used as an important tool to measure deviations between the offspring mean and its Lotka–Nagaev estimator for a supercritical Galton-Watson process(GW) with or without immigration (see [1], [3],[6],[7],[9],[10], [13], [15], etc).

Formally, consider a supercritical GW $(Z_n, n \geq 0)$ with offspring distribution $\{p_k\}_{k\geq 0}$. The offspring mean is defined as $m = \sum_j jp_j$. A basic task in statistical inference of GW is the estimation of the offspring mean m. The well known Lotka–Nagaev estimator of m, proposed by Nagaev([11]), is defined by

$$R_n = \begin{cases} Z_{n+1}/Z_n, & \text{if } Z_n > 0; \\ 1, & \text{if } Z_n = 0. \end{cases}$$

One of the interesting topics is to consider the convergence rate of

$$\mathbb{P}\left(|R_n - m| > \varepsilon_n\right), \tag{1.1}$$

for some consequence of positive random variables $\{\varepsilon_n\}$ as $n \to \infty$.

2020 *Mathematics Subject Classification*. Primary 60J80; Secondary 60F10
Keywords. large deviation, moderate deviation, normal deviation, Galton–Watson process, Lotka–Nagaev estimator
Received: 21 February 2020; Revised: 07 March 2021; Accepted: 05 July 2021
Communicated by Marija Milošević
Corresponding author: Zhenlong Gao
Research supported by Natural Science Foundation of Shandong Province of China (ZR2021MA32 and ZR2020MA023) and NSFC(11601260).
Email addresses: zhuyanjiao258@163.com (Yanjiao Zhu), gzlkygz@163.com (Zhenlong Gao)

If $\varepsilon_n = O(1)$ *a.s.*, (1.1) is said to be a large deviation probability. See [1] and [13] for $\varepsilon_n \equiv \varepsilon > 0$, and [3] for

$$\varepsilon_n = \varepsilon \sqrt{\frac{\sum_{i=1}^{Z_n} (X_{n,i} - R_n)^2}{Z_n}} = O(1) \;\; a.s.,$$

where $X_{n,i}$ stands for the offspring number of the ith individual in the nth generation.

If $\varepsilon_n = O(1/\sqrt{Z_n})$ *a.s.*, we call (1.1) a normal deviation probability. See [11] for $\varepsilon_n = \varepsilon\sigma m^{-n/2} = O(1/\sqrt{Z_n})$ *a.s.*, and [8] for $\varepsilon_n = \varepsilon\sigma/\sqrt{Z_n}$.

If ε_n satisfies $\varepsilon_n \to 0$ *a.s.* and $\sqrt{Z_n}\varepsilon_n \to \infty$, then (1.1) is said to be a moderate deviation probability. [6] and [7] deal with the case that $\{\varepsilon_n\}$ is a sequence of positive constants which satisfy $\varepsilon_n \to 0$ and $m^{n/2}\varepsilon_n \to \infty$. Particularly, one can choose $\varepsilon_n = \varepsilon m^{-\delta n}$, $\delta \in (0, 0.5)$.

In the rest of this paper, we assume that

$$p_0 = 0, \;\; \forall j, p_j < 1, \;\; \sigma^2 = \mathbb{V}\mathrm{ar}(Z_1) \in (0, \infty).$$

In this paper, we deal with random sums indexed by the generations of a branching process. Let $\{X_n\}$ be a sequence of i.i.d. random variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{V}\mathrm{ar}(X_1) = 1$. Define $S_n = X_1 + \cdots + X_n$ and $L_n = S_{Z_n}/Z_n$, the main task of this manuscript is the estimation of the rates for

$$\mathbb{P}(|L_n| > \varepsilon_n) \tag{1.2}$$

as $n \to \infty$. Moderate deviations were given in [6] and [7] in the case that $\{\varepsilon_n\}$ is a sequence of positive constants which satisfy $\varepsilon_n \to 0$ and $m^{n/2}\varepsilon_n \to \infty$.

Note that

$$R_n - m = \frac{(X_{n,1} - m) + \cdots + (X_{n,Z_n} - m)}{Z_n},$$

we know the deviations for Lotka–Nagaev estimator belong to the model (1.2).

In addition, it is well known that the martingale $\{W_n = Z_n/m^n\}$ convergent to a nonnegative random variable W and

$$\frac{W - W_n}{W_n} = \frac{Y_1 + \cdots + Y_{Z_n}}{Z_n},$$

where $\{Y_n\}$ are independent and have the same distribution as $W - 1$. Since $\mathbb{V}\mathrm{ar}(Z_1) < \infty$, one has $\mathbb{E}(W) = 1$ and $\mathbb{V}\mathrm{ar}(W) \in (0, \infty)$.

Model (1.2) can also be used to estimate the deviations between p_k and its nonparametric estimation $\widehat{p_k}$, see [14], where

$$\widehat{p_k} = \frac{I(X_{n,1} = k) + \cdots + I(X_{n,Z_n} = k)}{Z_n},$$

and $I(A)$ is the indictor function of set A.

We distinguish our GW between the Schröder case and the Böttcher case depending on whether $p_0 + p_1 > 0$ or $p_0 + p_1 = 0$. Note that $p_0 = 0$ in this paper, the Schröder index is defined as $\alpha = -\log_m p_1 \in (0, +\infty]$. If $\alpha \in (0, +\infty)$, GW belongs to the Schröder case, else if $\alpha = \infty$, GW belongs to the Böttcher case.

Firstly, we give the results of large deviations in the case that $\varepsilon_n \equiv \varepsilon > 0$.

Theorem 1.1 (Schröder case, light tail). *If there exists a constant $\theta_0 > 0$, such that $\mathbb{E}(\exp(\theta_0|X_1|)) < \infty$, then $\forall\, \varepsilon > 0$, one has*

$$\frac{1}{p_1^n}\mathbb{P}(|L_n| > \varepsilon) \to \sum_{k=1}^{\infty} \Psi(k, \varepsilon) q_k < \infty, \tag{1.3}$$

where $\Psi(k, \varepsilon) = P(|\overline{X}_k| > \varepsilon)$, $\overline{X}_k = k^{-1}S_k$ and $\{q_k\}$ are defined in Lemma Appendix A.1.

Theorem 1.1 shows that if the Cramér condition for X_1 is true and the GW belongs to Schröder case, then the large deviation probability has an exponential rate of decay. From the proof of Theorem 1.1, we know that these conditions lead to $\Psi(k,\varepsilon) = O(\lambda^k)$ for some $\lambda \in (0,1)$ as $k \to \infty$. It turns out that if $\Psi(k,\varepsilon) = O(k^{-r})$ and r is large enough, then (1.3) is also true.

Consider X_1 with the following heavy tails,

$$\mathbb{P}(X_1 \geq x) \sim \mathbb{P}(X_1 \leq -x) \sim \theta x^{-(1+r)}, \tag{1.4}$$

where θ is positive constant, $l(x) \sim s(x)$ stands for $l(x)/s(x) \to 1,\ x \to +\infty,$.

Theorem 1.2 (Schröder case, heavy tail). *If there exists a constant* $r > \max(\alpha, 1)$ *such that* (1.4) *is satisfied, then* (1.3) *is true.*

It turns out that if $\alpha \leq 1$ and the tail probability of X_1 satisfies (1.4), then (1.3) is always true for $r > 1$. But if $\alpha > 1$, (1.3) is only true for $r > \alpha$. We find that if $1 < r \leq \alpha$, then large deviation probabilities have different decay rates (since $\mathbb{E}(X_1^2) < \infty$, r can only be greater than 1).

Theorem 1.3 (Schröder case, heavy tail). *If* $\alpha \in (1,\infty)$ *and the tail probability of* X_1 *satisfies* (1.4) *with* $r = \alpha$, *then we have*

$$\frac{1}{np_1^n}\mathbb{P}(|L_n| > \varepsilon) \to \frac{\theta}{\Gamma(\alpha)}\varepsilon^{-(1+\alpha)}\int_1^m Q(\phi(v))v^{\alpha-1}\mathrm{d}v,$$

where $\Gamma(\cdot)$ *is the* Γ *function,* $Q(s)$ *is defined in Lemma Appendix A.1 and* $\phi(\cdot)$ *is the Laplace transformation of* W.

Theorem 1.4 (heavy tail). *If* $\alpha \in (1,\infty)$ *and the tail probability of* X_1 *satisfies* (1.4) *with* $1 < r < \alpha$, *then we have*

$$m^{rn}\mathbb{P}(|L_n| > \varepsilon) \to \frac{\theta}{\Gamma(r)}\varepsilon^{-(1+r)}\int_0^\infty \phi(v)v^{r-1}\mathrm{d}v.$$

If GW belongs to the Böttcher case, then $\alpha = \infty$. Theorem 1.4 shows that for all $r > 1$, if the tail probability of X_1 satisfies (1.4), then $\mathbb{P}(|L_n| > \varepsilon) = O(m^{-rn})$. The next theorem shows that if X_1 has light tail, then the decay rate of large deviation probabilities are supergeometric.

Theorem 1.5 (Böttcher case, light tail). *If there exists a constant* $\theta_0 > 0$, *such that* $\mathbb{E}(\exp(\theta_0|X_1|)) < \infty$, *then* $\forall\ \varepsilon > 0$, *we obtain*

$$\mathbb{P}(|L_n| > \varepsilon) \leq C(\varepsilon)(\lambda(\varepsilon))^{\mu^n},$$

where $C(\varepsilon) > 0,\ \lambda(\varepsilon) \in (0,1)$ *are two constants and* $\mu = \min\{k : p_k > 0\}$.

It is obvious that we have not found the exact rates of decay for large deviation probabilities. However, we can obtain the following large deviation principle(LDP). Some notes are needed to illustrate the LDP.

$$\gamma = (1-p_\mu)/p_\mu, \quad g(s) = \sum_{k=\mu+1}^{\infty} p_k s^{k-\mu}/(1-p_\mu), \quad R(s) = \prod_{k=0}^{\infty}(1+\gamma g(f_k(s)))^{1/\mu^{k+1}}, \quad B(s) = p_\mu^{1/(\mu-1)} sR(s),$$

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda X_1}), \quad \Lambda^*(x) = \sup_{\lambda\in R}\{\lambda x - \Lambda(\lambda)\}, \quad I(x) = -\log(B(\exp(-\Lambda^*(x)))).$$

Theorem 1.6 (Böttcher case, light tail). *If there exists a constant* $\theta_0 > 0$, *such that* $\mathbb{E}(\exp(\theta_0|X_1|)) < \infty$, *then* L_n *satisfies LDP with speed* μ^{-n} *and rate function* $I(x)$, *that is, for all Borel set* A *in* $\mathbb{R}$, *we have*

$$\begin{aligned}
-\inf_{x\in A^o} I(x) &\leq \liminf_{n\to\infty}\frac{1}{\mu^n}\log\mathbb{P}(L_n \in A) \\
&\leq \limsup_{n\to\infty}\frac{1}{\mu^n}\log\mathbb{P}(L_n \in A) \\
&\leq -\inf_{x\in\overline{A}} I(x),
\end{aligned}$$

where A^o, $\overline{A}$ *are the interior and closure of* A *respectively.*

Next, we give the results of moderate deviation. Let $l(n) : \mathbb{N} \mapsto (0, \infty)$ satisfy $l(n) \to 0$ and $\sqrt{n}l(n) \to \infty$, if X_1 has light tails, we choose $\varepsilon_n = l(Z_n)$.

Theorem 1.7 (Schröder case, light tail). *If there exists a constant $\theta_0 > 0$, such that $\mathbb{E}(\exp(\theta_0|X_1|)) < \infty$, then*

$$\frac{1}{p_1^n}\mathbb{P}(|L_n| > \varepsilon_n) \to \sum_{k=1}^{\infty} \psi_k q_k < \infty, \tag{1.5}$$

where $\psi_k = \mathbb{P}(|\overline{X}_k| > l(k))$.

If X_1 has heavy tails (1.4), the phase transition results similar to those in large deviations cannot be obtained in moderate deviations for general $l(Z_n)$, so we consider the decay rates of moderate deviation probabilities for $\varepsilon_n = \varepsilon Z_n^{-\delta}$, $\delta \in (0, 0.5)$. Define

$$\tau = \frac{\alpha + \delta}{1 - \delta}.$$

Theorem 1.8 (Schröder case, heavy tail). *Let $\varepsilon_n = \varepsilon Z_n^{-\delta}, \delta \in (0, 0.5)$. If there exists a constant $r > \max(\tau, 1)$ such that* (1.4) *is satisfied, then* (1.5) *is true.*

It turns out that if $\alpha \le 1 - 2\delta$ and the tail probability of X_1 satisfies (1.4), then (1.5) is always true for $r > 1$. But if $\alpha > 1 - 2\delta$, (1.5) is only true for $r > \tau$. We will show that if $1 < r \le \tau$, then moderate deviation probabilities have different decay rates.

Theorem 1.9 (Schröder case, heavy tail). *Let $\varepsilon_n = \varepsilon Z_n^{-\delta}, \delta \in (0, 0.5)$. If $\alpha \in (1 - 2\delta, \infty)$ and the tail probability of X_1 satisfies* (1.4) *with $r = \tau$, then we have*

$$\frac{1}{np_1^n}\mathbb{P}(|L_n| > \varepsilon_n) \to \frac{\theta}{\Gamma(\alpha)}\varepsilon^{-(1+\tau)}\int_1^m Q(\phi(v))v^{\alpha-1}\mathrm{d}v,$$

where $Q(s)$ is defined in Lemma Appendix A.1 and $\phi(\cdot)$ is the Laplace transformation of W.

Theorem 1.10 (heavy tail). *Let $\varepsilon_n = \varepsilon Z_n^{-\delta}, \delta \in (0, 0.5)$. If $\alpha \in (1 - 2\delta, \infty)$ and the tail probability of X_1 satisfies* (1.4) *with $1 < r < \tau$, then we have*

$$m^{\varsigma n}\mathbb{P}(|L_n| > \varepsilon_n) \to \frac{\theta}{\Gamma(\varsigma)}\varepsilon^{-(1+r)}\int_0^{\infty} \phi(v)v^{\varsigma-1}\mathrm{d}v,$$

where $\varsigma = r(1 - \delta) - \delta$.

If GW belongs to the Böttcher case, then $\alpha = \infty$ and $\tau = \infty$. Theorem 1.10 shows that for all $r > 1$, if the tail probability of X_1 satisfies (1.4), then $\mathbb{P}(|L_n| > \varepsilon_n) = O(m^{-\varsigma n})$. The next theorem shows that if X_1 has light tail, then the decay rate of moderate deviation probabilities are supergeometric.

Theorem 1.11 (Böttcher case, light tail). *Let $\delta \in (0, 0.5)$, $\varepsilon_n = \varepsilon Z_n^{-\delta}$. If there exists a constant $\theta_0 > 0$, such that $\mathbb{E}(\exp(\theta_0|X_1|)) < \infty$, then for some $C(\varepsilon) > 0$, one has*

$$\mathbb{P}(|L_n| > \varepsilon_n) \le \exp\left(-C(\varepsilon)\mu^{(1-2\delta)n}\right).$$

Finally, we consider the normal deviations. Two cases of ε_n are chosen, $\varepsilon_n = \varepsilon m^{-n/2}$ and $\varepsilon_n = \varepsilon Z_n^{-1/2}$.

Theorem 1.12. *Assume that $\varepsilon_n = \varepsilon m^{-n/2}$, we obtain*

$$\mathbb{P}(|L_n| > \varepsilon_n) \to 2\int_0^{\infty}\left(1 - \Phi\left(\varepsilon x^{1/2}\right)\right)w(x)dx,$$

where $\Phi(x)$ is the cumulative distribution function of a standard normal random variable and $w(x)$ is the probability density function of W.

Theorem 1.13. *Assume that $\varepsilon_n = \varepsilon Z_n^{-1/2}$, we have*

$$\mathbb{P}(|L_n| > \varepsilon_n) \to 2(1 - \Phi(x)).$$

The rest of this paper is organized as follows. In Section 2, we give the proofs of main results. In Section 3, we give some examples and conduct simulation studies to illustrate the performance of our results. Axillary results needed in the proofs are given in Appendix.

2. Proofs of main results

2.1. Proofs of large deviation results

The proofs are heavily dependent on the decay rates of generating function $f_n(s) = \mathbb{E}(s^{Z_n})$ and that of harmonic moments of Z_n, see Lemma Appendix A.1 – Lemma Appendix A.3.

The proof of Theorem 1.1.

Choose $c \in (0, \theta_0)$, $\beta < 0$, by Markov inequality, we obtain

$$\begin{aligned}
\Psi(k, \varepsilon) &= \mathbb{P}(\overline{X}_k > \varepsilon) + \mathbb{P}(\overline{X}_k < -\varepsilon) \\
&= \mathbb{P}(S_k > k\varepsilon) + \mathbb{P}(S_k < -k\varepsilon) \\
&= \mathbb{P}(\exp(cS_k) > \exp(ck\varepsilon)) + \mathbb{P}(\exp(\beta S_k) > \exp(-\beta k\varepsilon)) \\
&\le \mathbb{E}(\exp(cS_k))/\exp(ck\varepsilon) + \mathbb{E}(\exp(\beta S_k))/\exp(-\beta k\varepsilon) \\
&= (g(c)/\exp(c\varepsilon))^k + (g(\beta)/\exp(-\beta\varepsilon))^k,
\end{aligned}$$

where $g(s) = \mathbb{E}(\exp(sX_1))$ is the generating function of X_1. Denote

$$u(s) = g(s) - \exp(s\varepsilon), \quad v(s) = g(s) - \exp(-s\varepsilon),$$

one has $u(0) = 0$, $v(0) = 0$, $u'(0) = \mathbb{E}(X_1) - \varepsilon < 0$ and $v'(0) = \mathbb{E}(X_1) + \varepsilon > 0$. Thus, we can choose $c_0 \in (0, \theta_0), \beta_0 < 0$ such that $u(c_0) < 0, v(\beta_0) < 0$, which means

$$0 < g(c_0)/\exp(c_0\varepsilon) < 1, \quad 0 < g(\beta_0)/\exp(-\beta_0\varepsilon) < 1.$$

Define $\lambda = [g(c_0)/\exp(c_0\varepsilon)] \vee [g(\beta_0)/\exp(-\beta_0\varepsilon)]$, we know

$$\Psi(k, \varepsilon) \le 2\lambda^k.$$

According to the formula of total probability, we obtain

$$\begin{aligned}
\mathbb{P}(|L_n| > \varepsilon) &= \sum_{k=1}^{\infty} \Psi(k, \varepsilon) P(Z_n = k) \\
&\le \sum_{k=1}^{\infty} 2\lambda^k P(Z_n = k) \\
&= 2f_n(\lambda).
\end{aligned}$$

We complete the proof of Theorem 1.1 via Lemma Appendix A.1 and Lebesgue's dominated convergence theorem. □

The proof of Theorem 1.2 depends on the relation of deviation for S_n and its individuals, see Lemma Appendix B.1.

The proof of Theorem 1.2.

According to Lemma Appendix B.1,

$$\begin{aligned}
\Psi(k, \varepsilon) &= \mathbb{P}(|\overline{X}_k| > \varepsilon) \\
&= \mathbb{P}(|S_k| > k\varepsilon) \\
&\sim k\mathbb{P}(|X_1| > k\varepsilon), \ k \to \infty
\end{aligned}$$

Since (1.4) is satisfied, we know that $\mathbb{P}(|X_1| > k\varepsilon) = O(k^{-(1+r)})$. Then there exists a constant $C(\varepsilon) > 0$ such that

$$\Psi(k, \varepsilon) \le C(\varepsilon)k^{-r}.$$

The formula of total probability implies that

$$\begin{aligned}
\mathbb{P}(|L_n| > \varepsilon) &= \sum_{k=1}^{\infty} \Psi(k,\varepsilon)P(Z_n = k) \\
&\le \sum_{k=1}^{\infty} C(\varepsilon)k^{-r}P(Z_n = k) \\
&= C(\varepsilon)\mathbb{E}(Z_n^{-r}).
\end{aligned}$$

Note that $r > \max\{\alpha, 1\}$, we complete the proof of Theorem 1.2 via Lemma Appendix A.3 and Lebesgue's dominated convergence theorem. □

The proof of Theorem 1.3.

Note that $\Psi(k, \varepsilon) \sim k\mathbb{P}(|X_1| > k\varepsilon)$, $k \to \infty$, from (1.4), we know that for any $\delta > 0$, there exists a constants $\theta > 0, k_0 = k_0(\delta)$ such that for all $k \ge k_0$, one has

$$\theta k^{-\alpha}\varepsilon^{-(1+\alpha)}(1-\delta) \le \Psi(k,\varepsilon) \le \theta k^{-\alpha}\varepsilon^{-(1+\alpha)}(1+\delta). \tag{2.1}$$

Conditioning on Z_n, we obtain

$$\begin{aligned}
\mathbb{P}(|L_n| > \varepsilon) &= \sum_{k=1}^{\infty} \Psi(k,\varepsilon)P(Z_n = k) \\
&= \sum_{k=1}^{k_0-1} \Psi(k,\varepsilon)P(Z_n = k) + \sum_{k=k_0}^{\infty} \Psi(k,\varepsilon)P(Z_n = k) \\
&\ge \sum_{k=1}^{k_0-1} \Psi(k,\varepsilon)P(Z_n = k) + \theta\varepsilon^{-(1+\alpha)}(1-\delta)\sum_{k=k_0}^{\infty} k^{-\alpha}P(Z_n = k) \\
&= \theta\varepsilon^{-(1+\alpha)}(1-\delta)\mathbb{E}(Z_n^{-\alpha}) + \sum_{k=1}^{k_0-1} \Psi(k,\varepsilon)P(Z_n = k) \\
&- \theta\varepsilon^{-(1+\alpha)}(1-\delta)\sum_{k=1}^{k_0-1} k^{-\alpha}P(Z_n = k) \\
&= \theta\varepsilon^{-(1+\alpha)}(1-\delta)\mathbb{E}(Z_n^{-\alpha}) + I_0(n),
\end{aligned}$$

where

$$I_0(n) = \sum_{k=1}^{k_0-1} \Psi(k,\varepsilon)P(Z_n = k) - \theta\varepsilon^{-(1+\alpha)}(1-\delta)\sum_{k=1}^{k_0-1} k^{-\alpha}P(Z_n = k).$$

According to Lemma Appendix A.1, for any $k \le k_0$, we have

$$\frac{\mathbb{P}(Z_n = k)}{np_1^n} = \frac{1}{n}\frac{\mathbb{P}(Z_n = k)}{p_1^n} = O(n^{-1}) \to 0, \ \ n \to \infty,$$

Consequently, $I_0(n)/(np_1^n) \to 0$, $n \to \infty$. By Lemma Appendix A.3 and the arbitrariness of δ, we obtain

$$\liminf_{n\to\infty} \frac{1}{np_1^n}\mathbb{P}(|L_n| > \varepsilon) \le \frac{\theta}{\Gamma(\alpha)}\varepsilon^{-(1+\alpha)}\int_1^m Q(\phi(v))v^{r-1}\mathrm{d}v. \tag{2.2}$$

Similarly, if we use the right hand of (2.1), we have

$$\limsup_{n\to\infty} \frac{1}{np_1^n}\mathbb{P}(|L_n| > \varepsilon) \geq \frac{\theta}{\Gamma(\alpha)}\varepsilon^{-(1+\alpha)} \int_1^m Q(\phi(v))v^{r-1}\mathrm{d}v. \tag{2.3}$$

We complete the proof of Theorem 1.3 by (2.2) and (2.3). □

The proof of Theorem 1.4.

The proof of Theorem 1.4 is similar to that Theorem 1.3. It is only need to note that when $r < \alpha, p_1 m^r < 1$, then for any k, we have

$$m^{rn}\mathbb{P}(Z_n = k) = (p_1 m^r)^n \frac{\mathbb{P}(Z_n = k)}{p_1^n} = O((p_1 m^r)^n) \to 0, \quad n \to \infty.$$

When GW belongs to the Böttcher case, $f_n(s)$ has a supergeometric decay rate, see Lemma Appendix A.2.

The proof of Theorem 1.5.

According to the proof of Theorem 1.1, there exists a constant $\lambda_0 = \lambda_0(\varepsilon) \in (0,1)$ such that

$$\mathbb{P}(|L_n| > \varepsilon) \leq 2f_n(\lambda_0).$$

By Lemma Appendix A.2, we know for some positive constant $C = C(\varepsilon)$, we have

$$\mathbb{P}(|L_n| > \varepsilon) \leq C(B(\lambda_0))^{\mu^n}.$$

Denote $\lambda = \lambda(\varepsilon) := B(\lambda_0)$, if $\lambda \in (0,1)$, then Theorem 1.5 is true. In fact, according to Lemma Appendix A.2, one has

$$f(s)R(f(s)) = p_\mu (sR(s))^\mu,$$

By iteration, we know for any $n \geq 1$, one has

$$\begin{aligned} f_n(s)R(f_n(s)) &= p_\mu(f_{n-1}(s)R(f_{n-1}(s)))^\mu \\ &= p_\mu^{\mu+1}(f_{n-2}(s)R(f_{n-2}(s)))^{\mu^2} \\ &= \cdots \\ &= p_\mu^{\mu^{n-1}+\cdots+\mu+1}(sR(s))^{\mu^n} \\ &= p_\mu^{-1/(\mu-1)}(p_\mu^{1/(\mu-1)}sR(s))^{\mu^n}. \end{aligned}$$

Consequently,

$$\begin{aligned} B(s) &= p_\mu^{1/(\mu-1)}sR(s) \\ &= p_\mu^{1/(\mu^n(\mu-1))}(f_n(s)R(f_n(s)))^{1/\mu^n}. \end{aligned}$$

Note that for any $s \in [0,1)$, $f_n(s) \to 0$, according to the continuity of $R(s)$ at $s = 0$ and $R(0) = 1$, we obtain that for any $s \in [0,1)$, $f_n(s)R(f_n(s)) \to 0$. Consequently, one can choose some positive n_0 such $f_{n_0}(\lambda_0)R(f_{n_0}(\lambda_0)) < 1$. Then

$$\lambda = B(\lambda_0) = p_\mu^{1/(\mu^{n_0}(\mu-1))}(f_{n_0}(\lambda_0)R(f_{n_0}(\lambda_0)))^{1/\mu^{n_0}} < 1.$$

we complete the proof of Theorem 1.5 □

The proof of Theorem 1.6 depends on that classical deviation results for i.i.d. case, see Lemma Appendix B.3 and Appendix B.4.

The proof of Theorem 1.6.

Upper bound. For any closed set $F \subset \mathbb{R}$, by Lemma Appendix B.3, one has

$$\begin{aligned}\mathbb{P}(L_n \in F) &= \sum_{k=\mu^n}^{\infty} \mathbb{P}(\overline{X}_k \in F)\mathbb{P}(Z_n = k) \\ &\leq \sum_{k=\mu^n}^{\infty} 2\mathbb{P}(Z_n = k)\exp(-k\inf_{x\in F}\Lambda^*(x)) \\ &= 2f_n(\exp(-\inf_{x\in F}\Lambda^*(x))).\end{aligned}$$

Note that $B(s)$ increases as s increases, by Lemma Appendix A.2, we know

$$\begin{aligned}\limsup_{n\to\infty}\frac{1}{\mu^n}\log\mathbb{P}(L_n \in F) &\leq \log(B(\exp(-\inf_{x\in F}\Lambda^*(x)))) \\ &= \log(\sup_{x\in F} B(\exp(-\Lambda^*(x)))) \\ &= -\inf_{x\in F}(-\log(B(\exp(-\Lambda^*(x))))) \\ &= -\inf_{x\in F} I(x).\end{aligned}$$

Lower bound. According to Lemma Appendix B.4, for any open set $G \subset \mathbb{R}$ and $\delta > 0$, there exists a positive constant $N = N(\delta)$ such that for any $k \geq N$, we have

$$\mathbb{P}(\overline{X}_k \in G) \geq \exp(-k(\inf_{x\in G}\Lambda^*(x) + \delta)).$$

For n large enough, one has $\mu^n \geq N$, then

$$\mathbb{P}(L_n \in G) \geq 2f_n(\exp(-(\inf_{x\in G}\Lambda^*(x) + \alpha))).$$

Applying Lemma Appendix A.2, the monotonicity of$B(s)$ and the arbitrariness of δ, we obtain

$$\liminf_{n\to\infty}\frac{1}{\mu^n}\log\mathbb{P}(L_n \in G) \geq -\inf_{x\in G} I(x).$$

We complete the proof of Theorem 1.6. □

2.2. *Proofs of moderate deviation results*

The proof of Theorem 1.7 is similar to that of Theorem 1.1, we omit it.

The proof of Theorem 1.8.

Note that for any $\delta \in (0, 0.5)$,

$$\frac{k^{1-\delta}}{(1-\delta)\log k} \to \infty, \quad k \to \infty,$$

according to Lemma Appendix B.1, we have

$$\begin{aligned}\psi_k &= \mathbb{P}(|\overline{X}_k| > k^{-\delta}) \\ &\sim k\mathbb{P}(|X_1| > k^{1-\delta}) \\ &= O(k \cdot (k^{1-\delta})^{-(1+r)}) \\ &= O(k^{1-(1-\delta)(1+r)}),\ k \to \infty.\end{aligned}$$

Then, there exists a positive constant C such that

$$\psi_k \leq Ck^{1-(1-\delta)(1+r)}.$$

Conditioning on Z_n, one has

$$\mathbb{P}(|L_n| > \varepsilon_n) \le \sum_{k=1}^{\infty} Ck^{1-(1-\delta)(1+r)}P(Z_n = k) = C\mathbb{E}(Z_n^{1-(1-\delta)(1+r)}).$$

Since $r > \max\{\tau, 1\}$, $1-(1-\delta)(1+r) < -\alpha$, we complete the proof of Theorem 1.8 via Lemma Appendix A.1, Lemma Appendix A.3 and Lebesgue's dominated convergence theorem. □

The proof of Theorem 1.9.

According to Lemma Appendix B.1,

$$\psi_k \sim k\mathbb{P}(|X_1| > \varepsilon k^{1-\delta}),\ k \to \infty.$$

From (1.4), we know that for any $\vartheta > 0$, there exists a positive constant $\theta > 0, k_0 = k_0(\delta)$ such that for any $k \ge k_0$, one has

$$\begin{aligned} \psi_k &\ge \theta k \cdot k^{-(1-\delta)(1+\tau)}(1-\vartheta)\varepsilon^{-(1+\tau)} \\ &= \theta(1-\vartheta)k^{\delta-\tau(1-\delta)}\varepsilon^{-(1+\tau)} \\ &= \theta(1-\vartheta)k^{-\alpha}\varepsilon^{-(1+\tau)}, \end{aligned} \tag{2.4}$$

and similarly

$$\psi_k \le \theta(1+\vartheta)k^{-\alpha}\varepsilon^{-(1+\tau)}. \tag{2.5}$$

Conditioning on Z_n, by (2.4), one has

$$\mathbb{P}(|L_n| > \varepsilon_n) \ge \theta(1-\vartheta)\varepsilon^{-(1+\tau)}\mathbb{E}(Z_n^{-\alpha}) + I_0(n),$$

where

$$I_0(n) = \sum_{k=1}^{k_0-1} \psi_k P(Z_n = k) - \theta(1-\vartheta)\varepsilon^{-(1+\tau)} \sum_{k=1}^{k_0-1} k^{-\alpha}P(Z_n = k).$$

Applying Lemma Appendix A.1, we know that for any $k \le k_0$, we have

$$\frac{\mathbb{P}(Z_n = k)}{np_1^n} = \frac{1}{n}\frac{\mathbb{P}(Z_n = k)}{p_1^n} = O(n^{-1}),\quad n \to \infty,$$

Consequently, $I_0(n)/(np_1^n) \to 0$, $n \to \infty$. Applying Lemma Appendix A.3 and the arbitrariness of ϑ, one has

$$\liminf_{n\to\infty} \frac{1}{np_1^n}\mathbb{P}(|L_n| > \varepsilon_n) \ge \frac{\theta}{\Gamma(\alpha)}\varepsilon^{-(1+\tau)} \int_1^m Q(\phi(v))v^{\alpha-1}\mathrm{d}v. \tag{2.6}$$

Similarly, by (2.5), we have

$$\limsup_{n\to\infty} \frac{1}{np_1^n}\mathbb{P}(|L_n| > \varepsilon_n) \le \frac{\theta}{\Gamma(\alpha)}\varepsilon^{-(1+\tau)} \int_1^m Q(\phi(v))v^{r-1}\mathrm{d}v. \tag{2.7}$$

We complete the proof of Theorem 1.9 via (2.6) and (2.7). □

The proof of Theorem 1.10 is similar to that of Theorem 1.9, we omit it.

The proof of Theorem 1.11.

Define

$$\phi(k, \varepsilon) = \log(\mathbb{P}(|\overline{X}_k| > \varepsilon k^{-\delta})).$$

Letting $a_n = n^{2\delta-1}$ in Lemma Appendix B.5, one has $\sqrt{na_n} = n^\delta$ and for any $\vartheta > 0$, there exists a positive constant $k_0 = k_0(\vartheta, \varepsilon, \delta)$ such that if $k \ge k_0$, one has

$$\phi(k, \varepsilon) \le \left(-\frac{\varepsilon^2}{2} + \vartheta\right)k^{1-2\delta}.$$

Note that for n large enough, one has $\mu^n \geq k_0$. Conditioning on Z_n, we obtain

$$\begin{aligned}
\mathbb{P}(|L_n| \geq \varepsilon_n) &= \sum_{k=\mu^n}^{\infty} \mathbb{P}(|\overline{X}_k| > \varepsilon k^{-\delta}) P(Z_n = k) \\
&\leq \sum_{k=\mu^n}^{\infty} \exp\left(\left(-\frac{\varepsilon^2}{2} + \vartheta\right) k^{1-2\delta}\right) P(Z_n = k) \\
&\leq \sum_{k=\mu^n}^{\infty} \exp\left(\left(-\frac{\varepsilon^2}{2} + \vartheta\right) \mu^{(1-2\delta)n}\right) P(Z_n = k) \\
&= \exp\left(\left(-\frac{\varepsilon^2}{2} + \vartheta\right) \mu^{(1-2\delta)n}\right).
\end{aligned}$$

We complete the proof of Theorem 1.11. □

2.3. Proofs of normal deviation results

The proof of Theorem 1.12.
Conditioning on Z_n,

$$\mathbb{P}(|L_n| > \varepsilon_n) = \sum_{k=1}^{\infty} \mathbb{P}(|S_k| > \varepsilon_n k) \mathbb{P}(Z_n = k).$$

For any $1 > \delta > 0$, we divide $\mathbb{P}(|L_n| > \varepsilon_n)$ into the following three parts,

$$J_1(n, \varepsilon) = \sum_{k < \varepsilon m^n} \mathbb{P}(|S_k| > \varepsilon_n k) \mathbb{P}(Z_n = k),$$

$$J_2(n, \varepsilon) = \sum_{\varepsilon m^n \leq k \leq \varepsilon^{-1} m^n} \mathbb{P}(|S_k| > \varepsilon_n k) \mathbb{P}(Z_n = k),$$

$$J_3(n, \varepsilon) = \sum_{k > \varepsilon^{-1} m^n} \mathbb{P}(|S_k| > \varepsilon_n k) \mathbb{P}(Z_n = k).$$

Note that $Z_n/m^n \to W$ *a.s.*, see [2] for example, one has

$$\begin{aligned}
|J_1(n, \varepsilon)| &\leq \sum_{k < \varepsilon m^n} \mathbb{P}(Z_n = k) \\
&= \mathbb{P}(Z_n < \varepsilon m^n) \\
&= \mathbb{P}(W_n < \varepsilon) \\
&\to \mathbb{P}(W < \varepsilon).
\end{aligned}$$

Similarly, $|J_3(n, \varepsilon)| \leq \mathbb{P}(W_n > \varepsilon^{-1}) \to \mathbb{P}(W > \varepsilon^{-1})$. If $p_0 = 0$, one has $\mathbb{P}(W > 0) = 1$, see [2]. Letting $\varepsilon \to 0$, we have

$$\lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} J_1(n, \varepsilon) = \lim_{\varepsilon \downarrow 0} \lim_{n \to \infty} J_3(n, \varepsilon) = 0.$$

Next, according to Lemma Appendix B.2, we have

$$\begin{aligned} J_2(n,\varepsilon) &= \sum_{\varepsilon m^n \le k \le \varepsilon^{-1} m^n} \mathbb{P}\left(|S_k| > \varepsilon m^{-n/2} k\right) \mathbb{P}(Z_n = k) \\ &= \sum_{\varepsilon m^n \le k \le \varepsilon^{-1} m^n} \mathbb{P}\left(k^{-1/2}|S_k| > \varepsilon (k m^{-n})^{-1/2}\right) \mathbb{P}(Z_n m^{-n} = k m^{-n}) \\ &= 2\int_{\varepsilon}^{\varepsilon^{-1}} \left(1 - \Phi\left(\varepsilon x^{1/2}\right)\right) d\mathbb{P}(W_n \le x) + o(1) \\ &\to 2\int_{\varepsilon}^{\varepsilon^{-1}} \left(1 - \Phi\left(\varepsilon x^{1/2}\right)\right) w(x) dx, \end{aligned}$$

where the convergence is guaranteed by the continuity of $\Phi(\varepsilon x^{1/2})$ for x and the absolute continuity of the distribution of W, see [2]. Letting $\varepsilon \downarrow 0$, we complete the proof of Theorem 1.12. □

The proof of Theorem 1.13 is similar to that of Theorem 1.12, we omit it.

3. Examples and simulations

The only one non–trivial example for which the generating functions $f_n(s), n \ge 1$ have been explicitly computed is the linear fractional case. Note that $p_0 = 0$ in our paper, we can choose our offspring distribution satisfies the following geometric distribution,

$$p_k = p(1-p)^{k-1}, k = 1, 2, \cdots,$$

where $p \in (0,1)$. It is obvious that $p_1 = p,\ m = 1/p$ and the Schröder index $\alpha = 1$. In addition,

$$f_n(s) = \frac{s}{m^n - (m^n - 1)s}, \quad Q(s) = \frac{1}{1-s}, \quad q_k \equiv 1, \forall k \ge 1, \quad \phi(v) = \frac{1}{1+v}, \quad w(x) = e^{-x}.$$

We choose $p = 0.5$ in our simulations. The simulations include the following two steps.

Step I: Simulate the probability of deviations.

We conduct 10000 simulations and the frequency (FQ) is used to instead of the probability of deviations. According to our assumption, the distribution of Z_n is a geometric distribution with parameter 2^{-n}. One can generate Z_n from this population. Samples $(X_1, \cdots, X_{Z_n})$ are generated from the following two populations, a light tail distribution $N(0,1)$ and a heavy tail distribution (Pareto distribution with parameters $(3,1)$). Since $Z_n = O_p(2^n)$ as $n \to \infty$, we choose the following three sample size: 1. $n = 6 (2^6 = 64)$ stands for relatively small sample case; 2. $n = 8 (2^8 = 256)$ stands for middle sample case; 3. $n = 10 (2^{10} = 1024)$ stands for relatively large sample case.

Step II: Simulate the values given in our results.

For every k, we conduct 1000 simulations for $\Psi(k,x)$ and ψ_k and use $\sum_{k=1}^{100} q_k \Psi(k,x)$, $\sum_{k=1}^{100} q_k \psi_k$ instead of $\sum_{k=1}^{\infty} q_k \Psi(k,x)$ and $\sum_{k=1}^{\infty} q_k \psi_k$ respectively. We choose $x = 0.3, 0.4, 0.5, 0.6$ in our simulations as examples.

Let LD and MD stand for our large and moderate deviation results respectively. Compares of LD and FQ are given in Table 1 and Table 2. For moderate deviations, we choose $\delta = 0.1 \in (0, 0.5)$ as an example. Compares of MD and FQ are given in Table 3 and Table 4. It turns out that both large and moderate deviation results can work effectively for large samples. In addition, it is obvious that for n large enough, the decay rates for moderate deviations are slower than that of moderate deviations.

Let ND stands for our normal deviation result in Theorem 1.11. Compares of ND and empirical distribution functions are given in Figure 1 and Figure 2, where we choose $n = 10$. It turn out that if we choose $\varepsilon_n = \varepsilon m^{-n/2}$, then the degree of dispersion of the data is larger than a standard normal random variable.

Since the result of Theorem 1.12 does not need the explicit expression of $f_n(s)$, for each population, we can choose any non–trivial offspring distributions. Except for $Z_1 \sim \text{Geom}(0.5)$, we choose $Z_1 - 1 \sim Pois(1)$ as an example.

Compares of $\Phi(x)$ and empirical distribution functions are given in Figure 3 –Figure 6. It turns out that both Theorem 1.11 and Theorem 1.12 can work effectively for samples large enough.

Table 1: $\{p_k\} \sim$ Geom(0.5), $X_1 \sim N(0,1)$

Size	Method	$\varepsilon = 0.3$	$\varepsilon = 0.4$	$\varepsilon = 0.5$	$\varepsilon = 0.6$
$n = 6$	FQ	0.1366	0.0833	0.0538	0.0377
	LD	0.1672	0.0901	0.0579	0.0368
$n = 8$	FQ	0.0398	0.0234	0.0128	0.0093
	LD	0.0408	0.0231	0.0138	0.0093
$n = 10$	FQ	0.0103	0.0064	0.0040	0.0023
	LD	0.0104	0.0058	0.0035	0.0023

Table 2: $\{p_k\} \sim$ Geom(0.5), $\sqrt{3}X_1/2 + 0.5 \sim Pareto(3,1)$

Size	method	$\varepsilon = 0.3$	$\varepsilon = 0.4$	$\varepsilon = 0.5$	$\varepsilon = 0.6$
$n = 6$	FQ	0.0934	0.0507	0.0276	0.0163
	LD	0.1150	0.0581	0.0320	0.0197
$n = 8$	FQ	0.0292	0.0138	0.0072	0.0045
	LD	0.0288	0.0145	0.0077	0.0048
$n = 10$	FQ	0.0073	0.0033	0.0021	0.0014
	LD	0.0072	0.0036	0.0020	0.0012

Table 3: $\{p_k\} \sim$ Geom(0.5), $X_1 \sim N(0,1), \delta = 0.1$

Size	Method	$\varepsilon = 0.3$	$\varepsilon = 0.4$	$\varepsilon = 0.5$	$\varepsilon = 0.6$
$n = 6$	FQ	0.2193	0.1366	0.0829	0.0590
	MD	0.3242	0.1758	0.1016	0.0642
$n = 8$	FQ	0.0798	0.0439	0.0258	0.0149
	MD	0.0811	0.0441	0.0260	0.0155
$n = 10$	FQ	0.0223	0.0115	0.0058	0.0039
	MD	0.0203	0.0110	0.0063	0.0040

Table 4: $\{p_k\} \sim$ Geom(0.5), $\delta = 0.1$, $\sqrt{3}X_1/2 + 0.5 \sim Pareto(3,1)$

Size	method	$\varepsilon = 0.3$	$\varepsilon = 0.4$	$\varepsilon = 0.5$	$\varepsilon = 0.6$
$n = 6$	FQ	0.1683	0.0857	0.0494	0.0256
	LD	0.2288	0.1128	0.0592	0.0353
$n = 8$	FQ	0.0569	0.0287	0.0145	0.0084
	LD	0.0577	0.0281	0.0152	0.0090
$n = 10$	FQ	0.0165	0.0066	0.0032	0.0023
	LD	0.0143	0.0071	0.0038	0.0022

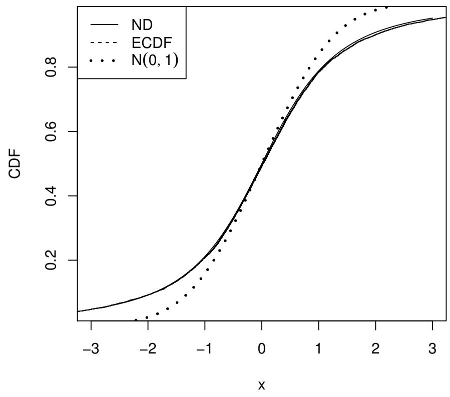

Figure 1: $\varepsilon_n = \varepsilon m^{-n/2}$, $X_1 \sim N(0,1)$

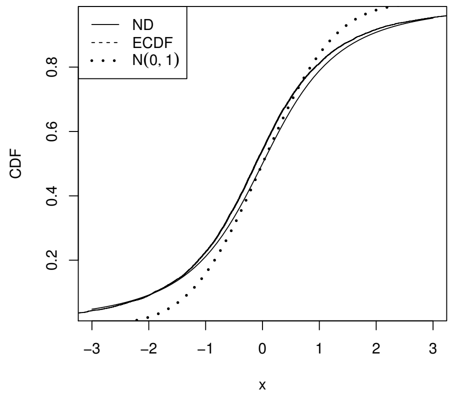

Figure 2: $e_n = \varepsilon m^{-n/2}$, $\sqrt{3}X_1/2 + 0.5 \sim \mathit{Pareto}(3,1)$

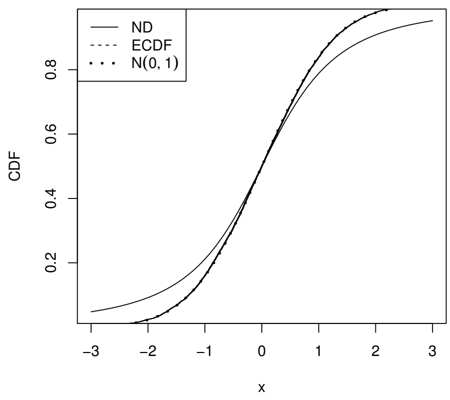

Figure 3: $\varepsilon_n = \varepsilon Z_n^{-1/2}$, $X_1 \sim N(0,1)$

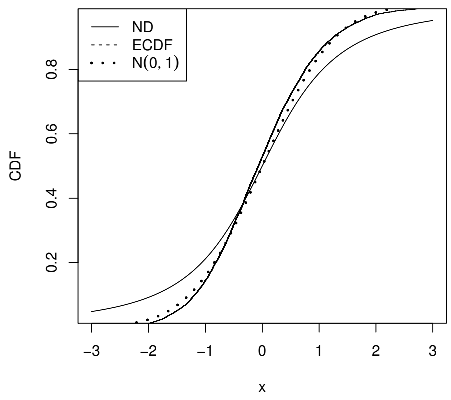

Figure 4: $\varepsilon_n = \varepsilon Z_n^{-1/2}$, $\sqrt{3}X_1/2 + 0.5 \sim \mathit{Pareto}(3,1)$

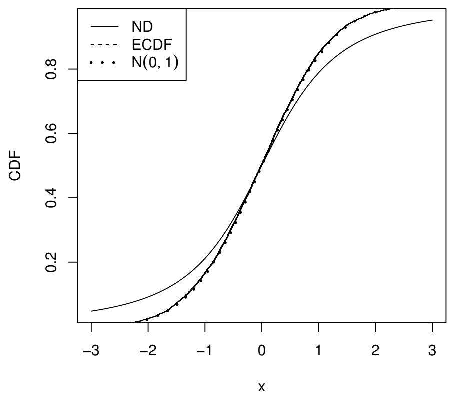

Figure 5: $Z_1 - 1 \sim \text{Pois}(0.5)$, $\varepsilon_n = \varepsilon Z_n^{-1/2}$, $X_1 \sim N(0,1)$

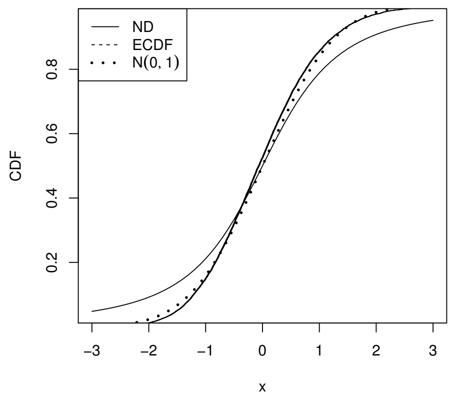

Figure 6: $Z_1 - 1 \sim \text{Pois}(0.5)$, $\varepsilon_n = \varepsilon Z_n^{-1/2}$, $\sqrt{3}X_1/2 + 0.5 \sim \mathit{Pareto}(3,1)$

Appendix A. Axillary results on Galton–Watson process

Lemma Appendix A.1 ([1] or [2]). *If* GW *belongs to Schröder case,* $p_0 = 0$, $\mathbb{E}(Z_1^2) < \infty$, *then*

$$\frac{f_n(s)}{p_1^n} \uparrow Q(s) = \sum_{k=1}^{\infty} q_k s^k, \quad \frac{\mathbb{P}(Z_n = k)}{p_1^n} \to q_k,$$

where $f_n(s) = \mathbb{E}(s^{Z_n})$ *and* $Q(s)$ *is the unique solution of the following equation,*

$$\begin{cases} Q(f(s)) = p_1 Q(s), & 0 \le s < 1; \\ Q(0) = 0. \end{cases}$$

Lemma Appendix A.2 ([1]). *If* GW *belongs to Böttcher case,* $\mathbb{E}(Z_1^2) < \infty$, *then*

$$f_n(s) \sim \left(p_\mu^{-1/(\mu-1)}\right)(B(s))^{\mu^n}, s \in [0,1], \quad f(s)R(f(s)) = p_\mu (sR(s))^\mu,$$

where $\mu, B(s), R(s)$ *are defined before Theorem 1.6 and* $f(s) = \mathbb{E}(Z_1)$.

Lemma Appendix A.3 ([13]). *If* $\mathbb{E}(Z_1^2) < \infty$, *then for any* $r > 0$, *we have*

$$\lim_{n\to\infty} A_n(r)\mathbb{E}(Z_n^{-r}) = C(r),$$

where

$$A_n(r) = \begin{cases} p_1^{-n}, & p_1 m^r > 1; \\ (np_1^n)^{-1}, & p_1 m^r = 1; \\ (m^r)^n, & p_1 m^r < 1 \end{cases}$$

and

$$C(r) = \begin{cases} \frac{1}{\Gamma(r)} \int_0^\infty Q(\exp\{-v\}) v^{r-1} \mathrm{d}v, & p_1 m^r > 1; \\ \frac{1}{\Gamma(r)} \int_1^m Q(\phi(v)) v^{r-1} \mathrm{d}v, & p_1 m^r = 1; \\ \frac{1}{\Gamma(r)} \int_0^\infty \phi(v) v^{r-1} \mathrm{d}v, & p_1 m^r < 1. \end{cases}$$

Appendix B. Axillary results on deviations of partial sums of i.i.d. case

Let $\{X_n\}$ be a sequence of i.i.d. radom variables with $\mathbb{E}(X_1) = 0$ and $\mathbb{V}\mathrm{ar}(X_1) = 1$. Define

$$S_n = X_1 + \cdots + X_n, \quad \overline{X}_n = S_n/n.$$

Lemma Appendix B.1 ([12]). *Assume that*

$$\mathbb{P}(X_1 \ge x) \sim x^{-r} h_1(x), \ \ \mathbb{P}(X_1 \le -x) \sim x^{-r} h_2(x),$$

for some $r > 2$, *where* $h_1(x), h_2(x)$ *are slowly varying functions. Let* $\{a_n\}$ *be a sequence of positive constants with* $a_n/\log(a_n) \ge \sqrt{n}$, *one has*

$$\frac{\mathbb{P}(|S_n| > a_n)}{n\mathbb{P}(|X_1| > a_n)} \to 1, \ \ n \to \infty.$$

Lemma Appendix B.2 ([5]P105). *Define* $F_n(x) = \mathbb{P}(\sqrt{n}\overline{X}_n \le x)$, $x \in \mathbb{R}$, *one has*

$$\Delta_n := \sup_{x\in\mathbb{R}} |F_n(x) - \Phi(x)| \to 0, \ \ n \to \infty.$$

Lemma Appendix B.3 ([4]P27). *Assume that there exists a constant* $\theta_0 > 0$ *such that* $E(e^{\theta_0|X_1|}) < \infty$. *Define*

$$\Lambda(\lambda) = \log \mathbb{E}(e^{\lambda X_1}), \quad \Lambda^*(x) = \sup_{\lambda \in R}\{\lambda x - \Lambda(\lambda)\}.$$

Then for any closed set $F \subset \mathbb{R}$, *one has*

$$\mathbb{P}(\overline{X}_n \in F) \leq 2e^{-n \inf_{x\in F} \Lambda^*(x)}.$$

Lemma Appendix B.4 ([4]P27). *Assume that there exists a constant* $\theta_0 > 0$ *such that* $E(e^{\theta_0|X_1|}) < \infty$. *Then for any open set* $G \subset \mathbb{R}$, *one has*

$$\frac{1}{n} \log \mathbb{P}(\overline{X}_n \in G) \geq -\inf_{x\in G} \Lambda^*(x).$$

Lemma Appendix B.5 ([4]P109). *Let* $\{a_n\}$ *be a sequence of positive constants with* $a_n \to 0$, $na_n \to \infty$, *if there exists a constant* $\theta_0 > 0$ *such that* $E(e^{\theta_0|X_1|}) < \infty$, *then for any* $\varepsilon > 0$, *one has*

$$a_n \log \mathbb{P}(\sqrt{na_n}|\overline{X}_n| \geq \varepsilon) \to -\frac{\varepsilon^2}{2}.$$

References

[1] K. B. Athreya, Large deviation rates for branching processes I: single type case, Annals of Applied Probability 4(1994) 779–790.
[2] K. B. Athreya, P.E. Ney, Branching Processes, Springer–Verlag, New York, 1972.
[3] W. J. Chu, Self-normalized large deviation for supercritical branching processes, Journal of Applied Probability 55(2018) 450–458.
[4] A. Dembo, O. Zeitouni, Large deviations techniques and applications,(4rd edition), Springer–Verlag, New York, 1998.
[5] R. Durrett, Probabilty: theory and examples, (4rd edition), Springer–Verlag, New York, 2010.
[6] K. Fleischmann, V. Wachtel, Large deviations for sums indexed by the generations of a Galton–Watson process, Probability Theory and Related Fields 141(2008) 445–470.
[7] H. He, On large deviation rates for sums associated with Galton-Watson processes, Advances in Applied Probability 48(2016) 672–690.
[8] C. C. Heyde, B. M. Brown, An invariance principle and some convergence rate results for branching processes, Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete 20(1971) 271–278.
[9] L. Y. Li, J. P. Li, Large deviation rates for supercritical branching processes with immigration, Journal of Theoretical Probability, 34(2021) 162–172.
[10] J. N. Liu, M. Zhang, Large deviation for supercritical branching processes with immigration. Acta Mathematica Sinica 32(2016) 893–900.
[11] A. V. Nagaev, On estimating the expected number of direct descendants of a particle in a branching process, Theory of Probability and its Applications 12(1967) 314–320.
[12] A. V. Nagaev, Limit theorems considering large deviations with Cramér's condition violated, Izv. Akad. Nauk. Uzbek SSR, Ser. Mat. Nauk, 6 (1969) 17–22.
[13] P. E. Ney, A. N. Vidyashankar, Harmonic moments and large deviation rates for supercritical branching processes, Annals of Applied Probability 13(2003) 475–489.
[14] A. G. Pakes, Non–parametric estimation in the Golton–Watson process, Mathematical Biosciences 26(1975) 1–18.
[15] Q.Sun, M.Zhang, Harmonic moments and large deviations for supercritical branching processes with immigration, Frontiers of Mathematics in China 12(2017) 1201–1220.
[16] S. J. Wu, Large deviation results for a randomly indexed branching process with applications to finance and physics, Doctoral Thesis, Graduate Faculty of North Carolina State University, (2012).