ABSTRACT

WU, SHENG-JHIH. Large Deviation Results for a Randomly Indexed Branching Process with Applications to Finance and Physics. (Under the direction of Min Kang.)

The large deviation behavior of a randomly indexed branching process is explored for the first time. We consider a branching process subordinated by a Poisson process. The large deviation behavior of the ratio of successive generation sizes deviating from the expected number of children of each individual is studied. Assume that at least one child in each birth, under various moment conditions on the offspring distribution, the rate of convergence is exponential. We also investigate the behavior under conditioning on non-extinction at the present generation as well as conditioning on non-extinction at the next generation. Conditioned on the limiting random variable of a sequence of normalized population sizes being positive, the large deviation probabilities decay super-exponentially. In addition, for the difference between the limiting random variable and the associated martingale sequence, the rate of convergence is super-exponential. Some limit theorems concerning the rate of convergence of the generating function as well as that of its inverse function are obtained. The results are then applied to a mean reversion in stock market and to a neutron fluctuation control problem.

Large Deviation Results for a Randomly Indexed Branching Process with Applications to Finance and Physics

by
Sheng-Jhih Wu

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Mathematics

Raleigh, North Carolina

2012

APPROVED BY:

Ronald Fulp	Robert Martin
Moody Chu	Min Kang Chair of Advisory Committee

DEDICATION

To my parents and brother

for their support and love

BIOGRAPHY

Sheng-Jhih Wu was born in a lovely town, Guanshi, in Hsinchu, Taiwan. He is pleased to be a part of an excellent family consisting of his parents and a younger brother. Perhaps because his parents are both teachers, he also desires to devote himself to education. He received his Bachelor of Business Administration degree in Finance from Ming Chuan University in 2000 and his Master of Business Administration degree in Finance from National Chung Cheng University in 2005. He went to the United States to pursue a master's degree in economics at North Carolina State University in 2007. Soon after arriving, he was quickly enthralled by mathematics, and began engaging himself in the graduate program in mathematics.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Min Kang for bringing me into the world of mathematics. Without taking her probability course, I would not study mathematics. Thank her for being so patient with me. It is not an easy work to teach a student starting from having hard time communication in mathematics to completing a doctoral study in mathematics. I especially want to thank her for being a role model of an excellent mathematician and a sincere person. I have learned a lot from her: not only mathematics but also sincereness. I hope I am able to be a role model for students as she is.

I also thank Dr. Moody Chu and his wife Mrs. Joyce Chu for their warm care, guidance, and prayers. I have seen how they take care of family and people around them. I hope I am able to contribute myself as they do.

I thank my committee members, Dr. Robert Martin and Dr. Ronald Fulp, for their wonderful instruction in their analysis courses from undergraduate to graduate level. I have enjoyed their courses very much. Also thank them for their suggestions on my work.

I appreciate Dr. Kazufumi Ito for the kind opportunity of attaining some valuable teaching experience.

I would also like to thank my advisor Dr. An-Sing Chen from National Chung Cheng University for his advising and supporting. Without sharing me with his story of study in the United State, I would probably not come here to pursue my doctoral study.

I thank my parents and brother. Without their understanding and encouragement, I would not have made my way so far. I also thank every member of my big family. Hearing they calling my name “Sheng-Jhih” is the sweetest encouragement for me.

TABLE OF CONTENTS

Chapter 1

Large Deviation Theory

1.1 Background

Roughly speaking, large deviation is a theory about rare events. More precisely, large deviation theory studies a family of probabilities of rare events that decay exponentially fast. It is one of the most active areas in probability theory nowadays. The theory has been widely used in various fields such as probability theory and stochastic processes, and other disciplines such as physics, engineering, and finance.

Let us start with a classical problem in probability theory, namely, the asymptotic behavior of the empirical mean of independent, identically distributed (i.i.d.) random variables. Consider a sequence of i.i.d. real-valued random variables $X_1, X_2, \cdots$ on a probability space $(\Omega, \mathcal{F}, P)$. Let $E(X_1) = \mu$ and $Var(X_1) = \sigma^2$. Consider sum of n i.i.d. random variables $S_n = X_1 + \cdots + X_n$ and empirical mean of n i.i.d. random variables $\bar{S}_n = \frac{X_1 + \cdots + X_n}{n}$. Two standard theorems dealing with the asymptotic behavior of the empirical mean $\bar{S}_n$ are the Weak Law of Large Numbers (WLLN) and the Central Limit Theorem (CLT).

Weak Law of Large Numbers:

$\bar{S}_n$ converges to μ in probability as n goes to infinity.

The WLLN shows that the empirical mean $\bar{S}_n$ converges to the mean μ as $n \to \infty$.

Central Limit Theorem:

$$\frac{\bar{S}_n - \mu}{\sigma/\sqrt{n}} \text{ converges to Z in distribution as n goes to infinity,}$$

where Z is a standard Gaussian random variable.

Consequently, for any $b > 0$, the CLT gives $P(|\bar{S}_n - \mu| \geq b(\frac{\sigma}{\sqrt{n}})) \to 2\Phi(-b)$, where $\Phi(\cdot)$ is the cumulative distribution function of the standard Gaussian random variable. Thus, the CLT asserts that the probability that $\bar{S}_n$ deviates from μ by an amount of order $1/\sqrt{n}$ is asymptotically approximated by a probability from a standard normal distribution. The order $1/\sqrt{n}$ is the meaningful order of the fluctuations that remains in the limit.

Notice that since the WLLN gives the convergence of the empirical mean to μ, the mean of one random variable, in probability, it grants the convergence in distribution as well. However, it does not provide information about the rate of the convergence in distribution. On the other hand, the CLT gives the rate of the convergence in distribution being of the order of $1/\sqrt{n}$.

The theory of large deviations is concerned with the asymptotic behavior of the probability $P(|\bar{S}_n - \mu| \geq b\sigma)$. Thus, the theory of large deviations deals with the events where $\bar{S}_n$ deviates from μ by an amount of order 1. In contrast to the typical fluctuation of the order of $1/\sqrt{n}$, the fluctuation which is of the order of 1 is much bigger. This kind of fluctuation is a large deviation because the differences between the empirical mean and μ stay larger than a constant as the number of samples n grows — there has to be a larger and larger conspiracy going on among the samples to remain the empirical mean deviating from μ in the same way. This is why the theory of large deviations is called “large”.

Roughly speaking, under a certain moment condition, a basic result of large deviation theory indicates

$$\lim_{n\to\infty} \frac{1}{n} \log P(\bar{S}_n \geq \mu + b) = -I(b) < 0, \text{ where I is a “rate”.} \tag{1.1}$$

$$\text{That is, } P(\bar{S}_n \geq \mu + b) = e^{-nI(b)+o(n)}, \tag{1.2}$$

$$\text{i.e.,} \quad P(\bar{S}_n \geq \mu + b) \approx e^{-nI(b)} \text{ when n is large enough.} \tag{1.3}$$

Notice that while the WLLN gives the convergence of the empirical mean to the mean in probability, it does not provide information about the rate of the convergence in probability. Large deviation theory gives the rate of the convergence in probability being of the order of n on the logarithmic scale (from the point of view of (1.1)). In other words, the rate of decay of the probability of this large deviation event is exponential (from the point of view of (1.3)). Large deviation theory is concerned mainly with quantifying the "rate" I. This result is the Cramer's Theorem, which will be formally stated below.

On a proper domain of this function, define

$$\Lambda(\lambda) := \log M(\lambda) := \log E(e^{\lambda X_1}),$$

where $M(\lambda)$ is the moment generating function of X_1 evaluated at λ. Let $\mathcal{D}_\Lambda = \{\lambda : \Lambda(\lambda) < \infty\}$. Define the Fenchel-Legendre transformation of Λ to be

$$\Lambda^*(x) := \sup_{\lambda \in \mathbb{R}} [\lambda x - \Lambda(\lambda)].$$

Λ^* turns out to be the rate function for the i.i.d. random variables $\{X_i\}$. Let $\mathcal{D}_{\Lambda^*} = \{x : \Lambda^*(x) < \infty\}$.

Theorem 1.1 ***(Cramer's Theorem)***
Let $X_1, X_2, \cdots$ be i.i.d. $\mathbb{R}$-valued random variables satisfying $0 \in \mathcal{D}_\Lambda^\circ$, the interior of $\mathcal{D}_\Lambda$. Then for any $a > E(X_1)$,

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq an) = -\Lambda^*(a). \tag{1.4}$$

Remark 1.1 *For any $a < E(X_1) = \mu$, the theorem says $\lim_{n\to\infty} \frac{1}{n} \log P(S_n \leq an) = -\Lambda^*(a)$.*

Notice that by the properties of the rate function Λ^*, if $a > \mu$, then $\Lambda^*(x) \geq \Lambda^*(a)$ for all $x \geq a$. Hence (1.4) can be rewritten as

$$\lim_{n\to\infty} \frac{1}{n} \log P(S_n \geq an) = -\inf_{x \geq a} \Lambda^*(x). \tag{1.5}$$

A quotation from [20] is that (1.5) exhibits a key principle in large deviation theory: "any large deviation result is done in the least unlikely of all the unlikely ways!" That is, if an unlikely event $\{S_n \geq an\}$ happens, it is very likely that it happens in the most likely way since a is the minimizer of $\Lambda^*(x)$ on $[a, \infty)$ and this event is realized at this cheapest cost (highest probabilty).

It is worth mentioning that in this section we give the motivation for the large deviation theory through a sequence of i.i.d. random variables, but, in fact, the general theory of large deviations deals with a family of random objects that are not necessarily independent nor identically distributed. We will see more general theory as we move to the latter sections.

1.2 The Large Deviation Principle

Notice that the context in the remaining of chapter 1 serves as an introduction to the important results in the theory of large deviations. The reader is referred to [12] and [20] for details. In this section, we introduce the large deviation principle, which characterizes the asymptotic behavior of a family of probability measures according to a rate function. The characterization is through asymptotic lower and upper exponential bounds on the values which a measure in the family assigns to measurable subsets of a topological space.

Let $\{\mu_n\}$ be a family of probability measures on $(\mathcal{X}, \mathcal{B})$ where $\mathcal{X}$ is a topological space and $\mathcal{B}$ is the Borel σ-algebra on $\mathcal{X}$.

Definition 1.1 ***(Rate Function)***
An extended real-valued function $I : \mathcal{X} \to [0, \infty]$ defined on a topological space $\mathcal{X}$ is said to be a rate function if it is not identically ∞ and is lower semi-continuous, i.e., for all $\alpha \in [0, \infty)$, the level set $\{x : I(x) \leq \alpha\}$ is a closed subset of $\mathcal{X}$. A good rate function is a rate function where all the level sets associated with it are compact subsets of $\mathcal{X}$. The effective domain of a rate function I, denoted by $\mathcal{D}_I$, is the set of points x in $\mathcal{X}$ such that $I(x)$ are finite, i.e., $\mathcal{D}_I = \{x \in \mathcal{X} : I(x) < \infty\}$.

A rate function is used to formulate a large deviation principle for a family of probability measures as can be seen from the following definition.

Definition 1.2 ***(Large Deviation Principle)***
$\{\mu_n\}$ is said to satisfy the large deviation principle with a rate function I, if for each $\Gamma \in \mathcal{B}$,

$$-\inf_{x \in \Gamma^0} I(x) \leq \liminf_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq \limsup_{n \to \infty} \frac{1}{n} \log \mu_n(\Gamma) \leq -\inf_{x \in \bar{\Gamma}} I(x),$$

where Γ° is the interior of Γ and $\bar{\Gamma}$ is the closure of Γ.

Remark 1.2 *The definition of the large deviation principle given above is equivalent to the following:*

(a) for any closed subset F of $\mathcal{X}$,

$$\limsup_{n \to \infty} \frac{1}{n} \log \mu_n(F) \leq -\inf_{x \in F} I(x), \text{ and}$$

(b) for any open subset G of $\mathcal{X}$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \mu_n(G) \geq -\inf_{x \in G} I(x).$$

Remark 1.3 *Recall that a family of probability measures* $\{\mu_n\}$ *is said to converge weakly to a probability measure* μ*, if either of the following conditions holds:*

(a) for any closed subset F *of* $\mathcal{X}$*,*

$$\limsup_{n\to\infty} \mu_n(F) \le \mu(F), \text{ and}$$

(b) for any open subset G *of* $\mathcal{X}$*,*

$$\liminf_{n\to\infty} \mu_n(G) \ge \mu(G).$$

The roles of open and closed sets in the definition of the large deviation principle above are similar to those in that of the weak convergence of probability measures. Therefore, we can view the two bounds in the definition of the large deviation principle above as analogues of weak convergence on an exponential scale.

In proving the large deviation principle, it is quite often to prove the upper bound for compact sets first and then to extend it to closed sets. Thus, this motivates the following definition of the weak large deviation principle.

Definition 1.3 ***(Weak Large Deviation Principle)***
$\{\mu_n\}$ *is said to satisfy the weak large deviation principle with a rate function I, if*

(a) for any compact subset K *of* $\mathcal{X}$*,*

$$\limsup_{n\to\infty} \frac{1}{n}\log \mu_n(K) \le -\inf_{x\in K} I(x), \text{ and}$$

(b) for any open subset G *of* $\mathcal{X}$*,*

$$\liminf_{n\to\infty} \frac{1}{n}\log \mu_n(G) \ge -\inf_{x\in G} I(x).$$

Definition 1.4 ***(Exponential Tightness)***
$\{\mu_n\}$ *is said to be exponentially tight, if for any* $\alpha < \infty$*, there exists a compact subset* K_α *of* $\mathcal{X}$ *such that*

$$\limsup_{n\to\infty} \frac{1}{n}\log \mu_n(K_\alpha^c) < -\alpha, \text{ where } K_\alpha^c \text{ is the complement of } K_\alpha.$$

The exponential tightness says that for this family of probability measures, most of the probability mass on an exponential scale is concentrated on compact sets asymptotically. The exponentially tightness strengthens the weak large deviation principle to a full large deviation principle as the following remark indicates.

Remark 1.4 *If $\{\mu_n\}$ is exponentially tight and satisfies the weak large deviation principle with a rate function I, then $\{\mu_n\}$ satisfies the large deviation principle with I being a good rate function.*

1.3 Some Important Results in the Theory of Large Deviations

In this section, we investigate some essential results in large deviation theory. First of all, it would be convenient if we could move a large deviation principle from one space to another. The so-called contraction principle enables us to generate one large deviation principle from another through contraction. Therefore, we could prove a large deviation principle in a simpler space and then transfer it into a more sophisticated target space.

Theorem 1.2 ***(Contraction Principle)***
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Hausdorff spaces and $f : \mathcal{X} \to \mathcal{Y}$ a continuous function. Suppose $\{\mu_n\}$ is a family of probability measures on $\mathcal{X}$ that satisfies the large deviation principle with a good rate function $I_{\mathcal{X}} : \mathcal{X} \to [0, \infty]$. Then $\{\mu_n \circ f^{-1}\}$ is a family of probability measures on $\mathcal{Y}$ which satisfies the large deviation principle with a good rate function $I_{\mathcal{Y}} : \mathcal{Y} \to [0, \infty]$ defined as $I_{\mathcal{Y}}(y) = inf\{I_{\mathcal{X}}(x) : x \in \mathcal{X}, y = f(x)\}$.

The next theorem indicates that if the family of probability measures is exponentially tight, then the above contraction principle could work in the opposite direction.

Theorem 1.3 ***(Inverse Contraction Principle)***
Let $\mathcal{X}$ and $\mathcal{Y}$ be two Hausdorff spaces and $f : \mathcal{Y} \to \mathcal{X}$ a continuous bijective function. Suppose $\{\mu_n\}$ is a family of probability measures on $\mathcal{Y}$ that is exponentially tight. If $\{\mu_n \circ f^{-1}\}$ satisfies the large deviation principle with a rate function $I_{\mathcal{X}} : \mathcal{X} \to [0, \infty]$, then $\{\mu_n\}$ satisfies the large deviation principle with the good rate function $I_{\mathcal{Y}} = I_{\mathcal{X}}(f)$.

Let us notice that the above theorem allows us to strengthen the large deviation principle from a coarser topology to a finer one. This result is stated in the following theorem.

Theorem 1.4 *Let $(\mathcal{X}, \tau_1)$ and $(\mathcal{X}, \tau_2)$ be two Hausdorff spaces with topology τ_1 coarser than topology τ_2. Let $\{\mu_n\}$ be a family of probability measures on $(\mathcal{X}, \tau_2)$ that is exponentially tight. If $\{\mu_n\}$ satisfies the large deviation principle on $(\mathcal{X}, \tau_1)$, then the same large deviation principle also holds for $(\mathcal{X}, \tau_2)$.*

Recall that in the preceding section, we mention that the definition of the large deviation principle is similar to that of the weak convergence of a family of probability measures and we give two equivalent definitions of the weak convergence. In fact, those two definitions of the weak convergence are equivalent to

$$\lim_{n\to\infty}\int_{\mathcal{X}} f(x)\mu_n(dx) = \int_{\mathcal{X}} f(x)\mu(dx) \quad \text{for all f} \in C_b(\mathcal{X}),$$

where $C_b(\mathcal{X})$ is the space of bounded continuous functions on $\mathcal{X}$.

This suggests that the large deviation principle is suited for handling the convergence of integrals of exponential functionals. This is formulated as the Varadhan's Integral Lemma, which could be viewed as a starting point for the theory of large deviations. It is a far-reaching generalization of the Laplace's method to abstract spaces and is a very handy tool for applications of the large deviation theory.

Theorem 1.5 ***(Varadhan's Integral Lemma)***
Let $\{\mu_n\}$ satisfy the large deviation principle on a topological space $\mathcal{X}$ with a good rate function $I : \mathcal{X} \to [0, \infty]$. Suppose $f : \mathcal{X} \to \mathbb{R}$ is a continuous function. Assume that either the following tail condition (i) or the moment condition (ii) are satisfied, where

(i)

$$\lim_{M\to\infty}\limsup_{n\to\infty}\frac{1}{n}\log E[e^{f(Z_n)n}\mathbf{1}_{\{f(Z_n)\geq M\}}] = -\infty, \textit{ and}$$

(ii)

$$\limsup_{n\to\infty}\frac{1}{n}\log E[e^{\gamma f(Z_n)n}] < \infty, \textit{ for some } \gamma > 1.$$

Then

$$\lim_{n\to\infty}\frac{1}{n}\log E[e^{f(Z_n)n}] = \sup_{x\in\mathcal{X}}\{f(x) - I(x)\}.$$

There is a Varadhan's Integral Lemma in the inverse direction, which is obtained by Bryc. It is useful for it serves as a way to establish the large deviation principle for an exponentially tight family of probability measures.

Theorem 1.6 ***(Inverse Varadhan's Integral Lemma)***
Define $\Lambda_n(f) = \frac{1}{n}\log\int_{\mathcal{X}} e^{f(x)n}\mu_n(dx)$ for any $f \in C_b(\mathcal{X})$. If $\{\mu_n\}$ is exponentially tight and the limit $\lim_{n\to\infty}\Lambda(f) \in \mathbb{R}$ exists for any $f \in C_b(\mathcal{X})$, then $\{\mu_n\}$ satisfies the large deviation principle on $\mathcal{X}$ with a good rate function I given by

$$I(x) = \sup_{f\in C_b(\mathcal{X})}\{f(x) - \Lambda(f)\}.$$

Moreover, for each $f \in C_b(\mathcal{X})$,

$$\Lambda(f) = \sup_{x\in\mathcal{X}}\{f(x) - I(x)\}.$$

There is an alternative version of the Varahdan's Integral Lemma, which allows us to generate a new large deviation principle from an old one by "tilting". It is called the tilted large deviation principle.

Theorem 1.7 ***(Tilted Large Deviation Principle)***
Let $\{\mu_n\}$ satisfy the large deviation principle on $\mathcal{X}$ with a good rate function I. Let $f : \mathcal{X} \to \mathbb{R}$ be a continuous function that is bounded above. For each Borel subset S of $\mathcal{X}$, define

$$J_n(S) := \int_S e^{f(x)n} \mu_n(dx).$$

Further, define a new family $\{\mu_n^f\}$ of probability measures by

$$\mu_n^f(S) := \frac{J_n(S)}{J_n(\mathcal{X})}.$$

Then $\{\mu_n^f\}$ satisfies the large deviation principle on $\mathcal{X}$ with a good rate function $I_f : \mathcal{X} \to [0, \infty]$ given by

$$I_f(x) = \sup_{y \in \mathcal{X}} \{f(y) - I(y)\} - \{f(x) - I(x)\}.$$

1.4 Three Levels of Large Deviations

In the literature, there are three levels of large deviations. They describe the large deviation behavior at different levels. These levels will be made precise in this section. Only the i.i.d. case of a sequence of random variables for these three levels will be discussed for simplicity. We now define three levels of large deviation principle for a sequence of i.i.d. random variables, $X_1, X_2, X_3, \cdots$ on a probability space $(\Omega, \mathcal{F}, P)$ with a common probability law ρ.

The level-1 large deviation is about the empirical mean of the random variable sequence. The concept of the level-1 large deviation has been mentioned in the preceding section and hence we give a brief description here. We focus on the case of a sequence of i.i.d. real-valued random variables, $X_1, X_2, X_3, \cdots$ for the level-1. Let S_n be the n-th partial sum of $X_1, X_2, X_3, \cdots$, i.e., $S_n = \sum_{i=1}^n X_i$. Then $\bar{S}_n = \frac{S_n}{n}$ is called the empirical mean of the n-th partial sum. Assume that the common mean $E(X_1)$ is finite, then by the WLLN the sequence $\{\bar{S}_n\}_{n=1}^\infty$ converges to $E(X_1)$ in probability. The level-1 large deviation investigates the asymptotic behavior of $\{\bar{S}_n\}_{n=1}^\infty$ deviating away from $E(X_1)$. Let $\mu_n^{(1)}$ be the probability law of $\bar{S}_n$. Then $\{\mu_n^{(1)}\}_{n=1}^\infty$ converges weakly to the unit point measure $\delta_{E(X_1)}$. Let A be an arbitrary Borel subset of $\mathbb{R}$ where A does not contain $E(X_1)$, then the level-1 large deviation studies the exponential decay of the sequence of probabilities $\{\mu_n^{(1)}(A)\}_{n=1}^\infty$ to zero as n goes to infinity with an exponential rate depending on A. The most classical theorem for i.i.d. case level-1 large deviation is Cramer's Theorem, which is briefly mentioned in the previous section. For completeness, we state it here in a slightly different form.

Recall that $\Lambda(\lambda) = \log M(\lambda) = \log E(e^{\lambda X_1})$, where $M(\cdot)$ is the moment generating function of the random variable X_1, is the logarithmic moment generating function. Then the Fenchel-Legendre transformation of Λ is

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}} [\lambda x - \Lambda(\lambda)].$$

Below we state Cramer's Theorem, which is a classical level-1 large deviation theorem.

Theorem 1.8 ***(Level-1 Large Deviation Principle)***
$\{\mu_n^{(1)}\}$ satisfies the large deviation principle with a convex rate function Λ^*, namely,

(a) for any closed subset F of $\mathbb{R}$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n^{(1)}(F) \leq -\inf_{x\in F} \Lambda^*(x), \text{ and}$$

(b) for any open subset G of $\mathbb{R}$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n^{(1)}(G) \geq -\inf_{x\in G} \Lambda^*(x).$$

Now let us move onto the level-2 large deviation and $\mathcal{X}$ be a Polish space. Let $X_1, X_2, X_3, \cdots$ be a sequence of $\mathcal{X}$-valued random variables with common probability law $\rho \in \mathcal{M}_1(\mathcal{X})$, where $\mathcal{M}_1(\mathcal{X})$ denotes the space of probability measures on $\mathcal{X}$. Suppose ρ is unknown, then given the first n samples, one may try to estimate the true probability law ρ by

$$L_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{X_i}.$$

L_n is called the empirical measure of $X_1, X_2, \cdots, X_n$. Notice that $L_n : \Omega \to \mathcal{M}_1(\mathcal{X})$ is a random probability measure, i.e., $L_n(\omega) \in \mathcal{M}_1(\mathcal{X})$ for all $\omega \in \Omega$. Let $\mu_n^{(2)}$ be the probability law of L_n. By ergodic theorem, the sequence $\{L_n\}_{n=1}^{\infty}$ converges weakly to ρ almost surely and this implies that $\{\mu_n^{(2)}\}_{n=1}^{\infty}$ converges weakly to the unit point measure δ_ρ.

Define the weak topology on $\mathcal{M}_1(\mathcal{X})$ as the topology generated by the sets of the form

$$U_{f,x,\epsilon} = \{\nu \in \mathcal{M}_1(\mathcal{X}) : |\int_{\mathcal{X}} f d\nu - x| < \epsilon\},$$

where $f \in C_b(\mathcal{X})$, $x \in \mathbb{R}$, and $\epsilon > 0$. Then $\mathcal{M}_1(\mathcal{X})$ is a Polish space since $\mathcal{X}$ is and since $M_1(\mathcal{X})$ is equipped with the weak topology. Let A be any Borel subset of $\mathcal{M}_1(\mathcal{X})$ where A does not contain δ_ρ, then the level-2 large deviation studies the exponential decay of the sequence of probabilities $\{\mu_n^{(2)}(A)\}_{n=1}^{\infty}$ to zero as n goes to infinity with an exponential rate depending on A. The most well-known theorem for i.i.d. case level-2 large deviation is the Sanov's Theorem, which is established by Sanov in 1957. This is an important extension of Cramer's theorem to empirical measures of real-valued i.i.d. random variables. Here we provide a more general Sanov's Theorem in a Polish space set-up, which is a classical level-2 large deviation theorem.

Theorem 1.9 ***(Level-2 Large Deviation Principle)***
$\{\mu_n^{(2)}\}$ satisfies the large deviation principle with a convex rate function Λ^*, namely,

(a) for any closed subset F of $\mathcal{M}_1(\mathcal{X})$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \mu_n^{(2)}(F) \leq -\inf_{\nu\in F} \Lambda^*(\nu), \textit{ and}$$

(b) for any open subset G of $\mathcal{M}_1(\mathcal{X})$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \mu_n^{(2)}(G) \geq -\inf_{\nu\in G} \Lambda^*(\nu),$$

where $\Lambda^*(\nu) = \sup_{f\in C_b(\mathcal{X})}[\int_{\mathcal{X}} f d\nu - \Lambda(f)]$ for any $\nu \in \mathcal{M}_1(\mathcal{X})$ and $\Lambda(f) = \log E[e^{\int_{\mathcal{X}} f d\delta_{X_1}}] = \log E[e^{f(X_1)}] = \log \int_{\mathcal{X}} e^f d\rho$ for all $f \in C_b(\mathcal{X})$.
Furthermore, define the relative entropy of the probability measure ν with respect to ρ as

$$H(\nu|\rho) = \begin{cases} \int_{\mathcal{X}} \log \frac{d\nu}{d\rho}\, d\nu & \textit{if} \quad \nu \textit{ is absolutely continous with respect to } \rho \\ \infty & \textit{otherwise.} \end{cases}$$

Then $\Lambda^*(\nu) = H(\nu|\rho)$.

It is worth mentioning a difference between the set-up for Cramer's theorem and that of Sanov's theorem. The former concerns large deviations away from a deterministic number (common mean of i.i.d. random variables) and the rate function is defined on $\mathbb{R}$, whereas the latter deals with deviations away from a probability measure (common probability law of i.i.d. random variables) and the rate function is defined on the space of probability measures. And notice that $L_n \in \mathcal{M}_1(\mathcal{X})$ is, in fact, the empirical mean of $\delta_{X_1}, \delta_{X_2}, \cdots, \delta_{X_n}$. Therefore, from this point of view, Sanov's theorem is actually a generalization of Cramer's theorem to a space of probability measures.

The level-3 large deviation extends the idea of empirical measure in the level-2 large deviation as shall be made clear in the following context. Let $X_1, X_2, X_3, \cdots$ be a sequence of i.i.d. $\mathcal{X}$-valued random variables on a probability space $(\Omega, \mathcal{F}, P)$, where $\mathcal{X}$ is a Polish space. Assume that $X_1, X_2, X_3, \cdots$ have a common probability law $\rho \in \mathcal{M}_1(\mathcal{X})$, where $\mathcal{M}_1(\mathcal{X})$ denotes the space of probability measures on $\mathcal{X}$. Given a positive integer n, repeat the sequence $X_1, X_2, \cdots, X_n$ periodically into an infinite sequence by

$$X^{(n)}(\omega) = (X_1(\omega), X_2(\omega), \cdots, X_n(\omega), X_1(\omega), X_2(\omega), \cdots, X_n(\omega), \cdots).$$

Then $X^{(n)} : \Omega \to \mathcal{X}^{\mathbb{N}}$ is a random variable. Let θ be the left-shifting mapping on $\mathcal{X}^{\mathbb{N}}$ defined by $(\theta x)_j = x_{j+1}$ for $j \in \mathbb{N}$, $x \in \mathcal{X}^{\mathbb{N}}$ and $\theta^i = \theta(\theta^{i-1})$ for $i = 1, 2, \cdots$. Define the empirical process corresponding to $X_1, X_2, \cdots, X_n$ by

$$R_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\theta^i(X^{(n)})}.$$

Notice that $R_n : \Omega \to \mathcal{M}_1(\mathcal{X}^{\mathbb{N}})$ is a random probability measure, where $\mathcal{M}_1(\mathcal{X}^{\mathbb{N}})$ denotes the space of probability measures on $\mathcal{X}^{\mathbb{N}}$. Since R_n may be identified with probability measures on processes, the large deviation principle associated with R_n is referred to as process-level large deviation principle. In fact, it is easy to see that $R_n(\omega)$ is a θ-invariant probability measure on $\mathcal{X}^{\mathbb{N}}$ for all $\omega \in \Omega$ since $X^{(n)}$ is constructed by a repetition with a period n. Therefore, $R_n(\omega) \in \mathcal{M}_1^{\theta}(\mathcal{X}^{\mathbb{N}})$ for all $\omega \in \Omega$, where

$$\mathcal{M}_1^{\theta}(\mathcal{X}^{\mathbb{N}}) = \{\nu \in \mathcal{M}_1(\mathcal{X}^{\mathbb{N}}) : \nu \circ \theta^{-1} = \nu\}$$

denotes the space of all θ-invariant probability measures on $\mathcal{X}^{\mathbb{N}}$.

Let $\mu_n^{(3)}$ be the probability law of R_n. By Birkhoff's Ergodic Theorem, the sequence $\{R_n\}_{n=1}^{\infty}$ converges weakly to $\rho^{\mathbb{N}}$ almost surely and this implies that $\{\mu_n^{(3)}\}_{n=1}^{\infty}$ converges weakly to the unit point measure $\delta_{\rho^{\mathbb{N}}}$. Let A be an arbitrary Borel subset of $\mathcal{M}_1^{\theta}(\mathcal{X}^{\mathbb{N}})$ where A does not contain $\delta_{\rho^{\mathbb{N}}}$, then the level-3 large deviation studies the exponential decay of the sequence of probabilities $\{\mu_n^{(3)}(A)\}_{n=1}^{\infty}$ to zero as n goes to infinity with an exponential rate depending on A.

For k $\in \mathbb{N}$, let d_k be the total variation distance on $\mathcal{M}_1(\mathcal{X}^{\mathbb{N}})$. Define

$$d(\mu, \nu) := \sum_{k=1}^{\infty} \frac{1}{2^k} d_k(p_k\mu, p_k\nu),$$

where $p_k : \mathcal{X}^{\mathbb{N}} \to \mathcal{X}^k$ is the projection of $\mathcal{X}^{\mathbb{N}}$ onto $\mathcal{X}^k$ defined by $p_k(x) = (x_1, x_2, \cdots, x_k)$ for $x = (x_1, x_2, \cdots, x_k, \cdots)$ and $p_k\mu = \mu \circ p_k^{-1}$. Then d makes $\mathcal{M}_1^{\theta}(\mathcal{X}^{\mathbb{N}})$ into a Polish space. The following theorem concerns the large deviation behavior of $\{\mu_n^{(3)}\}_{n=1}^{\infty}$ away from $\delta_{\rho^{\mathbb{N}}}$, which is a level-3 large deviation result.

Theorem 1.10 ***(Level-3 Large Deviation Principle)***
Let $B_a(\rho^{\mathbb{N}}) = \{\nu \in \mathcal{M}_1^{\theta}(\mathcal{X}^{\mathbb{N}}) : d(\nu, \rho^{\mathbb{N}}) \leq a\}$ for $a > 0$. Let

$$I_{\rho}^{\infty}(\nu) = \begin{cases} sup_{k\geq 2} H(p_k\nu | p_{k-1}\nu \otimes \rho) & \text{if} \quad \nu \in \mathcal{M}_1^{\theta}(\mathcal{X}^{\mathbb{N}}) \\ \infty & \text{otherwise,} \end{cases}$$

where $H(\cdot|\cdot)$ is the relative entropy. Define

$$J(a) := \inf_{\nu \in B_a^c(\rho^{\mathbb{N}})} I_{\rho}^{\infty}(\nu).$$

Then $\{\mu_n^{(3)}\}$ satisfies the large deviation principle with a good rate function I_{ρ}^{∞}, namely, for each $a > 0$,

(a) $\limsup_{n\to\infty} \frac{1}{n} \log \mu_n^{(3)}((B_a(\rho^{\mathbb{N}}))^c) \leq -J(a-)$, *where* $J(a-) = \lim_{\epsilon \downarrow 0} J(a - \epsilon)$.

(b) $\liminf_{n\to\infty} \frac{1}{n} \log \mu_n^{(3)}((B_a(\rho^{\mathbb{N}}))^c) \geq -J(a)$.

1.5 Large Deviations for Dependent Random Variables

In the preceding section, we discussed the large deviation behavior of i.i.d. random variables. In this section, we investigate the large deviation behavior of non-i.i.d. cases. We first introduce the Gartner-Ellis theorem, which is an important theorem in the non-i.i.d. scenario. By using the Gartner-Ellis theorem, one can acquire a level-1 large deviation result for a discrete-time, irreducible, finite state space Markov chains. For level-2 and level-3 large deviations, we provide corresponding results for a discrete-time Markov chain that satisfies some strong uniformity assumption.

Let $Z_1, Z_2, Z_3, \cdots$ be a sequence of $\mathbb{R}^d$-valued random variables, where Z_n possesses the probability law μ_n with logarithmic moment generating function

$$\Lambda_n(\lambda) = \log E[e^{\langle \lambda, Z_n \rangle}],$$

where $\langle \lambda, Z_n \rangle = \sum_{i=1}^{d} \lambda_i (Z_n)_i$ is the inner product of λ and Z_n in $\mathbb{R}^d$.

Assumption (*) For each $\lambda \in \mathbb{R}^d$, the logarithmic moment generating function, defined as the limit

(1) $\Lambda(\lambda) = \lim_{n\to\infty} \frac{1}{n}\Lambda_n(n\lambda) \in [-\infty, \infty]$ exists.

(2) $\overrightarrow{0} \in \mathcal{D}_\Lambda^\circ$, the interior of $\mathcal{D}_\Lambda$, where $\mathcal{D}_\Lambda = \{\lambda \in \mathbb{R}^d : \Lambda(\lambda) < \infty\}$.

Let Λ^* denote the Fenchel-Legendre transformation of Λ, i.e.,

$$\Lambda^*(x) = \sup_{\lambda \in \mathbb{R}^d} [\langle x, \lambda \rangle - \Lambda(\lambda)] \quad \text{for } x \in \mathbb{R}^d.$$

In order to discuss Gartner-Ellis theorem, we need some definitions, and they are given below.

Definition 1.5 ***(Exposed Point)***
A point $x \in \mathbb{R}^d$ is called an exposed point of Λ^*, if there exists a point $\lambda \in \mathbb{R}^d$ such that

$$\langle \lambda, x \rangle - \Lambda^*(x) > \langle \lambda, y \rangle - \Lambda^*(y) \quad \textit{for all } y \neq x.$$

Such λ is called an exposing hyperplane for x.

Definition 1.6 ***(Essential Smoothness)***
A convex function $\Lambda : \mathbb{R}^d \to (-\infty, \infty]$ is essentially smooth, if

(a) $\mathcal{D}_\Lambda^\circ$ is non-empty,

(b) Λ is differentiable on $\mathcal{D}_\Lambda^\circ$, and

(c) Λ is steep, namely, $lim_{n\to\infty}|\nabla\Lambda(\lambda_n)| = \infty$, whenever $\{\lambda_n\}$ is a sequence in $\mathcal{D}_\Lambda^\circ$ converging to a boundary point of $\mathcal{D}_\Lambda^\circ$.

Now let us state Gartner-Ellis theorem.

Theorem 1.11 ***(Gartner-Ellis Thorem)***
Let the assumption (*) hold.

(a) For any closed subset F of $\mathbb{R}^d$,

$$\limsup_{n\to\infty} \frac{1}{n}\log \mu_n(F) \leq -\inf_{x\in F} \Lambda^*(x).$$

(b) For any open subset G of $\mathbb{R}^d$,

$$\liminf_{n\to\infty} \frac{1}{n}\log \mu_n(G) \geq -\inf_{x\in G\cap\mathcal{F}} \Lambda^*(x),$$

where F is the set of exposed points of Λ^* whose exposing hyperplane belongs to $\mathcal{D}_\Lambda^\circ$.

(c) If Λ is an lower semi-continuous, essentially smooth function, then $G \cap \mathcal{F}$ in (b) may be replaced by G. Consequently, $\{\mu_n\}$ satisfies the large deviation principle with a good rate function Λ^*.

Gartner-Ellis theorem is now applied to a level-1 large deviation for a discrete-time, irreducible, finite state space Markov chain. Let $Y_1, Y_2, Y_3, \cdots$ be a discrete-time, irreducible Markov chain taking values in a finite state space Σ. Assume that $|\Sigma| = N$. Let $\mathit{\Pi} = \{\pi(i,j)\}_{i,j=1}^{|\Sigma|}$ be the stochastic matrix for the Markov chain. Let $P_{y_0}^\pi$ be the Markov probability measure associated with the stochastic matrix $\mathit{\Pi}$ and the initial state $y_0 \in \Sigma$, i.e.,

$$P_{y_0}^\pi(Y_1 = y_1, Y_2 = y_2, \cdots, Y_n = y_n) = \pi(y_0, y_1)\prod_{i=1}^{n-1}\pi(y_i, y_{i+1}).$$

We now establish the level-1 large deviation principle for additive functionals of Markov Chain $Y_1, Y_2, Y_3, \cdots$. Let $g : \Sigma \to \mathbb{R}^d$ be a deterministic function. Let $X_k = g(Y_k)$ for all $k \in \mathbb{Z}_+$. Define the empirical mean of $X_1, X_2, \cdots, X_n$ to be

$$Z_n := \frac{1}{n}\sum_{k=1}^{n} X_k.$$

We would like to establish the large deviation principle for the empirical mean, $\{Z_n\}$. For any $\lambda \in \mathbb{R}^d$, consider a non-negative matrix $\mathit{\Pi}_\lambda$ whose elements are

$$\pi_\lambda(i,j) = \pi(i,j)e^{\langle\lambda, g(j)\rangle} \quad \text{for } i,j \in \Sigma.$$

Let $\rho(\mathit{\Pi}_\lambda)$ denote the Perron-Frobenius eigenvalue of the matrix $\mathit{\Pi}_\lambda$. Define

$$I(z) = \sup_{\lambda\in\mathbb{R}^d} [\langle\lambda, z\rangle - \log\rho(\mathit{\Pi}_\lambda)] \quad \text{for each } z \in \mathbb{R}^d. \tag{1.6}$$

Theorem 1.12 *$\{Z_n\}$ satisfies the large deviation principle with a convex, good rate function I*

given in (1.6). More precisely, for any subset Γ *of* $\mathbb{R}^d$ *and any initial state* $y_0 \in \Sigma$*,*

(a) Upper Bound:

$$\limsup_{n\to\infty} \frac{1}{n} \log P_{y_0}^{\pi}(Z_n \in \Gamma) \leq -\inf_{z\in\bar{\Gamma}} I(z).$$

(b) Lower Bound:

$$\liminf_{n\to\infty} \frac{1}{n} \log P_{y_0}^{\pi}(Z_n \in \Gamma) \geq -\inf_{z\in\Gamma^\circ} I(z).$$

In the same setting, we now establish the level-2 large deviation principle for this discrete-time, irreducible, finite state space Markov chain $Y_1, Y_2, Y_3, \cdots$. Define the empirical measure associated with $Y_1, Y_2, \cdots, Y_n$ as

$$L_n^Y = (L_n^Y(1), L_n^Y(2), \cdots, L_n^Y(|\Sigma|)) : \Omega \to \mathcal{M}_1(\Sigma),$$

where $L_n^Y(i) = \frac{1}{n}\sum_{k=1}^n \delta_{Y_k}(i)$, $i = 1, \cdots, |\Sigma|$.
Suppose that $\mathit{\Pi}$ is the stochastic matrix for the Markov chain $Y_1, Y_2, Y_3, \cdots$ that is irreducible. Suppose that $\mu = (\mu_1, \mu_2, \cdots, \mu_{|\Sigma|})$ is the stationary distribution of the Markov chains, i.e. the unique left eigenvector associated with the eigenvalue 1 of $\mathit{\Pi}$, whose entries are non-negative and sum to 1, that satisfies the equation

$$\mu = \mu\mathit{\Pi}.$$

By ergodic theorem, the sequence $\{L_n^Y\}_{n=1}^\infty$ converges to μ in probability as $n \to \infty$. Therefore, $\{L_n^Y\}$ is a good candidate for a large deviation principle.

By identifying $g : \Sigma \to [0,1]^{|\Sigma|}$ defined by $g(Y_k) = X_k = (\delta_{Y_k}(1), \delta_{Y_k}(2), \cdots, \delta_{Y_k}(|\Sigma|))$, we can apply the previous theorem to obtain the large deviation principle for $\{L_n^Y\}$ as the following theorem indicates.

Theorem 1.13 $\{L_n^Y\}$ *satisfies the large deviation principle with a rate function* I *defined as*

$$I(\nu) = \sup_{\lambda\in\mathbb{R}^d} [\langle\lambda, \nu\rangle - \log\rho(\mathit{\Pi}_\lambda)] \quad \textit{for each } \nu \in \mathcal{M}_1(\Sigma),$$

where $\mathit{\Pi}_\lambda$ *is a non-negative matrix whose elements are*

$$\pi_\lambda(i,j) = \pi(i,j)e^{\langle\lambda, g(j)\rangle} = \pi(i,j)e^{\lambda_j} \quad \textit{for } i, j \in \Sigma.$$

Explicitly, for any subset Γ *of* $\mathcal{M}_1(\Sigma)$ *and any initial state* $y_0 \in \Sigma$*,* *(a) upper bound:*

$$\limsup_{n\to\infty} \frac{1}{n} \log P^{\pi}_{y_0}(L^Y_n \in \Gamma) \leq - \inf_{\nu \in \bar{\Gamma}} I(\nu), \text{ and}$$

(b) lower bound:

$$\liminf_{n\to\infty} \frac{1}{n} \log P^{\pi}_{y_0}(L^Y_n \in \Gamma) \geq - \inf_{\nu \in \Gamma^\circ} I(\nu).$$

Now we consider an extension from a finite state space to a general Polish state space. We first establish a level-2 large deviation principle in this setting and then move into a level-3 large deviation principle in the same setting.

Let Σ be a Polish space. Let $Y_1, Y_2, Y_3, \cdots$ be a discrete-time Markov chain with a state space Σ. Let $\mathcal{M}_1(\Sigma)$ denote the space of probability measures on Σ equipped with the Levy metric, which makes it into a Polish space with convergence compatible with the weak convergence. Let $\pi(y_0, \cdot)$ be a transition probability measure, i.e., $\pi(y_0, \cdot) \in \mathcal{M}_1(\Sigma)$ for any $y_0 \in \Sigma$. Let $P_{n,y_0} \in \mathcal{M}_1(\Sigma^n)$ denote the probability measure which assigns to any Borel measurable set $\Gamma \subseteq \Sigma^n$ the value

$$P_{n,y_0}(\Gamma) = \int_\Gamma \prod_{i=1}^{n-1} \pi(y_i, dy_{i+1}) \pi(y_0, dy_1).$$

Define the empirical measure of $Y_1, Y_2, \cdots, Y_n$ as

$$L^Y_n := \frac{1}{n} \sum_{k=1}^{n} \delta_{Y_k}.$$

Let μ_{n,y_0} denote the probability law of L^Y_n. We would like to establish the large deviation principle for $\{\mu_{n,y_0}\}$. In order to reach our goal for the discrete-time Markov chains $Y_1, Y_2, Y_3, \cdots$ with state space Σ, we need the following uniformity condition.

Assumption (**) There exists an integer m (between 0 and some integer N) and a constant $M \geq 1$ such that

$$\pi^m(y_0, \cdot) \leq \frac{M}{N} \sum_{k=1}^{N} \pi^k(\sigma, \cdot) \quad \text{for any } y_0, \sigma \in \Sigma,$$

where $\pi^k(\sigma, \cdot)$ is the k-step transition probability measure with the initial state σ, i.e.,

$$\pi^k(\sigma, \cdot) = \int_\Sigma \pi^{k-1}(\tau, \cdot) \pi(\sigma, d\tau).$$

The following theorem shows the large deviation principle for $\{\mu_{n,y_0}\}$.

Theorem 1.14 *Suppose the assumption (**) holds. Then for any $f \in C_b(\Sigma)$, the following* *limit exists:*

$$\Lambda(f) = \lim_{n\to\infty} \frac{1}{n}\log E_{y_0}[exp(\sum_{i=1}^{n} f(Y_i))],$$

where E_{y_0} denotes the expectation conditioned on the initial state y_0.
Furthermore, $\{\mu_{n,y_0}\}$ satisfies the large deviation principle with a convex good rate function Λ^*, where

$$\Lambda^*(\nu) = \sup_{f\in C_b(\Sigma)} [\langle f, \nu\rangle - \Lambda(f)]$$

is the Fenchel-Legendre transformation of Λ for $\nu \in \mathcal{M}_1(\Sigma)$.
Explicitly, for any subset Γ of $\mathcal{M}_1(\Sigma)$ and any initial state $y_0 \in \Sigma$,

(a) upper bound:

$$\limsup_{n\to\infty} \frac{1}{n}\log \mu_{n,y_0}(\Gamma) \le -\inf_{\nu\in\bar{\Gamma}} I(\nu), \text{ and}$$

(b) lower bound:

$$\liminf_{n\to\infty} \frac{1}{n}\log \mu_{n,y_0}(\Gamma) \ge -\inf_{\nu\in\Gamma^\circ} I(\nu).$$

We then study the level-3 large deviation behavior in the same setting. Let Σ be a Polish space. Let $Y_1, Y_2, Y_3, \cdots$ be a discrete-time Markov chain with a state space Σ. Let θ be the left-shifting mapping on $\Sigma^{\mathbb{Z}_+}$ defined by $(\theta y)_j = y_{j+1}$ for $j \in \mathbb{Z}_+$, $y \in \Sigma^{\mathbb{Z}_+}$ and $\theta^i = \theta(\theta^{i-1})$ for $i = 1, 2, \cdots$. Define the empirical process as

$$R_n^Y := \frac{1}{n}\sum_{i=1}^{n} \delta_{\theta^i(Y)},$$

where $Y = (Y_1, Y_2, Y_3, \cdots)$.

Notice that $R_n^Y : \Omega \to \mathcal{M}_1(\Sigma^{\mathbb{Z}_+})$ is a random probability measure, where $\mathcal{M}_1(\Sigma^{\mathbb{Z}_+})$ denotes the space of probability measures on $\Sigma^{\mathbb{Z}_+}$. Equipped with the weak topology, $\mathcal{M}_1(\Sigma^{\mathbb{Z}_+})$ is a Polish space. Let $p_k : \Sigma^{\mathbb{Z}_+} \to \Sigma^k$ denote the projection of $\Sigma^{\mathbb{Z}_+}$ onto Σ^k defined by $p_k(y) = (y_1, y_2, \cdots, y_k)$ for $y \in \Sigma^{\mathbb{Z}_+}$ and $p_k\mu = \mu \circ p_k^{-1}$ for $\mu \in \mathcal{M}_1(\Sigma^{\mathbb{Z}_+})$. We need the following definitions.

Definition 1.7 *A measure $\mu \in \mathcal{M}_1(\Sigma^k)$ is said to be shift invariant if, for any $\Gamma \in \mathcal{B}_{\Sigma^{k-1}}$,*

$$\mu(\{\tau \in \Sigma^k : (\tau_1, \cdots, \tau_{k-1}) \in \Gamma\}) = \mu(\{\tau \in \Sigma^k : (\tau_2, \cdots, \tau_k) \in \Gamma\}).$$

Moreover, for any $\nu \in \mathcal{M}_1(\Sigma^{k-1})$ and any transition kernel $\pi \in \mathcal{M}_1(\Sigma)$, define the probability measure $\nu \otimes \pi \in \mathcal{M}_1(\Sigma^k)$ by

$$(\nu \otimes \pi)(\Gamma) = \int_{\Sigma^{k-1}} \nu(dx) \int_{\Sigma} \pi(x_{k-1}, dy)\boldsymbol{1}_{\{(x,y)\in\Gamma\}} \quad \text{for any } \Gamma \in \mathcal{B}_{\Sigma^k}.$$

Definition 1.8 ***(Shift Invariance)***
A measure $\nu \in \mathcal{M}_1(\Sigma^{\mathbb{Z}_+})$ is said to be shift invariant if $p_k\nu$ is shift invariant in $\mathcal{M}_1(\Sigma^k)$ for all $k \in \mathbb{Z}_+$.

The next theorem is the large deviation result for the sequence $\{R_n^Y\}$.

Theorem 1.15 *Suppose the assumption (**) holds. Then $\{R_n^Y\}$ satisfies the large deviation principle with a good rate function*

$$I_\pi^\infty(\nu) = \begin{cases} sup_{k\geq 2} H(p_k\nu | p_{k-1}\nu \otimes \pi) & \text{if} \quad \nu \in \mathcal{M}_1(\Sigma^{\mathbb{Z}_+}) \text{ is shift invariant,} \\ \infty & \text{otherwise,} \end{cases}$$

where $H(\cdot|\cdot)$ is the relative entropy.

1.6 Sample Path Large Deviations

So far we have focused on the large deviation behavior of the collections of random elements at some fixed time points instead of the whole timeline, i.e., the realizations of those random elements are not considered a function of time. Given an $\omega \in \Omega$, a realization of an empirical mean S_n is a real number; that of an empirical measure L_n is a probability measure on some space; and that of an empirical process R_n is a probability measure on another space. None of these random elements take values in a space of functions evolving in time. It would be more interesting to investigate the large deviation behavior of the random elements taking values in the space of the functions of time. This is the purpose of this section.

In many situations, the matter of interest is some rare events that depend on time and it often touches the probability that the path of certain random process hits some special set. To introduce the sample path large deviations, we first provide the case of a random walk, which is the simplest case in the sample path large deviations. We then state the Brownian motion sample path large deviations as an application of the sample path large deviations of the random walk. Finally, we give an extension of the sample path large deviations of Brownian motion to that of the diffusion processes which are strong solutions of some stochastic differential equations.

Let $X_1, X_2, \cdots$ be a sequence of i.i.d. $\mathbb{R}^d$-valued random variables. Assume that the logarithmic moment generating function $\Lambda(\lambda) = \log E(e^{\langle\lambda, X_1\rangle})$ be finite for all $\lambda \in \mathbb{R}^d$. Let $\lfloor a \rfloor$ denote the integer part of a. Define

$$Z_n(t) := \frac{1}{n}\sum_{i=1}^{\lfloor nt \rfloor} X_i, \quad 0 \le t \le T,$$

where $Z_n : \Omega \to D([0,T])$, the space of functions continuous from right and having left limits that defined on $[0,T]$ with values in $\mathbb{R}^d$.

Let ν_n be the probability law of Z_n. Let $\Lambda^*(x) = sup_{\lambda \in \mathbb{R}^d}[\langle \lambda, x\rangle - \Lambda(\lambda)]$ denote the Fenchel-Legendre transformation of λ and $\mathcal{AC}_0$ denote the space of absolutely continuous functions that vanish at the origin. Notice that Z_n is a random walk in $\mathbb{R}^d$ with lifetime T for each $n \in \mathbb{N}$. The following theorem is the large deviation result for this sequence $\{\nu_n\}$ obtained by Mogulskii.

Theorem 1.16 $\{\nu_n\}$ *satisfies the large deviation principle with a good rate function*

$$I(f) = \begin{cases} \int_0^T \Lambda^*(\dot{f}(t))dt & \text{if} \quad f \in \mathcal{AC}_0, \\ \infty & \text{otherwise.} \end{cases}$$

More specifically,

(a) for any closed subset F of $D([0,T])$,

$$\limsup_{n\to\infty} \frac{1}{n} \log \nu_n(F) \le -\inf_{f\in F} I(f), \text{ and}$$

(b) for any open subset G of $D([0,T])$,

$$\liminf_{n\to\infty} \frac{1}{n} \log \nu_n(G) \ge -\inf_{f\in G} I(f).$$

It is obvious that for all large deviation results, we can replace the countable index n by a continuous index ϵ and treat $\epsilon = \frac{1}{n}$ as a special case. Therefore, the above theorem could be extended to the following setting.

Let μ_ϵ be the probability law of

$$Y_\epsilon(t) := \epsilon \sum_{i=1}^{\lfloor \frac{t}{\epsilon} \rfloor} X_i, \quad 0 \le t \le T.$$

Then the case of ν_n and $Z_n(t)$ becomes the special case of μ_ϵ and $Y_\epsilon(t)$ with $\epsilon = \frac{1}{n}$. Hence the above theorem could be extended to the following general theorem.

Theorem 1.17 *$\{\mu_\epsilon\}$ satisfies the large deviation principle with a good rate function*

$$I(f) = \begin{cases} \int_0^T \Lambda^*(\dot{f}(t))dt & \text{if} \quad f \in \mathcal{AC}_0 \\ \infty & \text{otherwise.} \end{cases}$$

More specifically,

(a) for any closed subset F of $D([0,T])$,

$$\limsup_{\epsilon \to 0} \epsilon \log \mu_\epsilon(F) \leq - \inf_{f \in F} I(f), \text{ and}$$

(b) for any open subset G of $D([0,T])$,

$$\liminf_{\epsilon \to 0} \epsilon \log \mu_\epsilon(G) \geq - \inf_{f \in G} I(f).$$

We now provide the large deviation result for a Brownian motion, which is obtained by Schilder. Let B_t be a $d-dimentional$ standard Brownian motion. Define the process

$$B_\epsilon(t) := \sqrt{\epsilon} B_t, \quad 0 \leq t \leq T,$$

where $B_\epsilon : \Omega \to \mathcal{C}_0([0,T] : \mathbb{R}^d)$, the space of $\mathbb{R}^d$-valued continuous functions defined on $[0,T]$ that vanish at the origin.

Recall that a standard Brownian motion could be constructed as a scaling limit of a symmetric random walk. More precisely,

$$B_t = \lim_{n \to \infty} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} X_i,$$

where X_i, $i = 1, 2, \cdots$, are i.i.d. random variables with common mean 0 and variance 1, and the limit of this convergence is in distribution sense.

Alternatively, in the ϵ setting,

$$B_t = \lim_{\epsilon \to 0} \sqrt{\epsilon} \sum_{i=1}^{\lfloor \frac{t}{\epsilon} \rfloor} X_i.$$

Define the process

$$\hat{B}_\epsilon(t) := B_\epsilon(\epsilon \lfloor \frac{t}{\epsilon} \rfloor).$$

Then

$$\begin{aligned}
\hat{B}_\epsilon(t) &= B_\epsilon(\epsilon\lfloor\frac{t}{\epsilon}\rfloor) \\
&= \sqrt{\epsilon}B_{\epsilon\lfloor\frac{t}{\epsilon}\rfloor} \\
&= \sqrt{\epsilon}(\sqrt{\epsilon}\sum_{i=1}^{\lfloor\frac{t}{\epsilon}\rfloor} X_i) \quad \text{in distribution} \\
&= \epsilon\sum_{i=1}^{\lfloor\frac{t}{\epsilon}\rfloor} X_i \\
&= Y_\epsilon(t),
\end{aligned}$$

where X_i, $i = 1, 2, \cdots$, are i.i.d. standard Gaussian random variables taking values in $\mathbb{R}^d$.

Therefore, $\hat{B}_\epsilon(\cdot)$ is simply a process $Y_\epsilon(\cdot)$ with the special choice of X_i, $i = 1, 2, \cdots$. Through this observation, we then can provide a large deviation result for a Brownian motion as an application of the above theorem concerning that for a random walk.

Let H_1 denote the space of absolutely continuous function with a square integrable derivative, i.e., $H_1 := \{\int_0^t f(s)ds : f \in L_2([0,T])\}$, where $L_2([0,T])$ is the space of square integrable functions defined on $[0,T]$ with values in $\mathbb{R}^d$. Let p_ϵ be the probability law of B_ϵ.

Theorem 1.18 *$\{p_\epsilon\}$ satisfies the large deviation principle with a good rate function*

$$I(g) = \begin{cases} \frac{1}{2}\int_0^T |\dot{g}(t)|^2 dt & \text{if} \quad g \in H_1 \\ \infty & \text{otherwise.} \end{cases}$$

More specifically,

(a) for any closed subset F of $\mathcal{C}_0([0,T] : \mathbb{R}^d)$,

$$\limsup_{\epsilon\to 0} \epsilon \log p_\epsilon(F) \leq -\inf_{g\in F} I(g), \text{ and}$$

(b) for any open subset G of $\mathcal{C}_0([0,T] : \mathbb{R}^d)$,

$$\liminf_{\epsilon\to 0} \epsilon \log p_\epsilon(G) \geq -\inf_{g\in G} I(g).$$

With the understanding of a large deviation result for a Brownian motion, we can now move onto that for diffusion processes. First, we consider the simplest case as an instructive example.

For each ϵ, let $\{x_\epsilon(t)\}$ be a diffusion process which is the unique solution of the following stochastic differential equation,

$$dx_\epsilon(t) = b(x_\epsilon(t))dt + \sqrt{\epsilon}dB_t, \quad 0 \le t \le T,\ x_\epsilon(0) = 0,$$

where $b : \mathbb{R} \to \mathbb{R}$ is a uniformly Lipschitz continuous function.

Let $\tilde{p}_\epsilon$ denote the probability law of x_ϵ. Notice that $x_\epsilon : \Omega \to \mathcal{C}_0([0,T] : \mathbb{R})$, the space of $\mathbb{R}$-valued continuous functions defined on $[0,T]$ that vanish at the origin. Let F be a function from $\mathcal{C}_0([0,T] : \mathbb{R})$ to $\mathcal{C}_0([0,T] : \mathbb{R})$ defined by $F(g) = h$, where h is the unique continuous solution of

$$h(t) = \int_0^t b(h(s))ds + g(t), \quad 0 \le t \le T.$$

It is clear that F is a continuous function. Notice that $\tilde{p}_\epsilon = p_\epsilon \circ F^{-1}$, where p_ϵ is the probability law of B_ϵ. Therefore, the large deviation principle for $\{x_\epsilon(t)\}$ here is a simple application of the contraction principle.

Theorem 1.19 *$\{\tilde{p}_\epsilon\}$ satisfies the large deviation principle with a good rate function*

$$I(h) = \begin{cases} \frac{1}{2}\int_0^T |\dot{h}(t) - b(h(t))|^2 dt & \text{if} \quad h \in H_1 \\ \infty & \text{otherwise.} \end{cases}$$

More specifically,

(a) for any closed subset F of $\mathcal{C}_0([0,T] : \mathbb{R})$,

$$\limsup_{\epsilon \to 0} \epsilon \log \tilde{p}_\epsilon(F) \le -\inf_{h \in F} I(h), \text{ and}$$

(b) for any open subset G of $\mathcal{C}_0([0,T] : \mathbb{R})$,

$$\liminf_{\epsilon \to 0} \epsilon \log \tilde{p}_\epsilon(G) \ge -\inf_{h \in G} I(h).$$

Our second example of a large deviation result for diffusion processes is described below. Let $\mathcal{C}([0,T] : \mathbb{R}^d)$ denote the space of $\mathbb{R}^d$-valued continuous functions defined on $[0,T]$. Let $\{x_\epsilon(t)\}$ be the diffusion process which is the unique solution of the following stochastic differential equation,

$$dx_\epsilon(t) = b(x_\epsilon(t))dt + \sqrt{\epsilon}\sigma(x_\epsilon(t))dB_t, \quad 0 \le t \le T,\, x_\epsilon(0) = x,$$

where all entries of $b : \mathbb{R}^d \to \mathbb{R}^d$ and all entries of $\sigma : \mathbb{R}^d \to \mathbb{R}^d \times \mathbb{R}^d$ are bounded, uniformly Lipschitz continuous functions.

Let μ_ϵ be the probability law of x_ϵ. Define

$$I^x(h) := \inf\{\frac{1}{2}\int_0^T |\dot{g}(t)|^2 dt \Big| g \in H_1 : h(t) = x + \int_0^t b(h(s))ds + \int_0^t \sigma(h(s))\dot{g}(s)ds\}.$$

The large deviation principle for $\{x_\epsilon(t)\}$ is an application of the contraction principle.

Theorem 1.20 *$\{\mu_\epsilon\}$ satisfies the large deviation principle with a good rate function I^x. More specifically,*

(a) for any closed subset F of $\mathcal{C}([0,T] : \mathbb{R}^d)$,

$$\limsup_{\epsilon\to 0} \epsilon \log \mu_\epsilon(F) \le - \inf_{h\in F} I^x(h), \text{ and}$$

(b) for any open subset G of $\mathcal{C}([0,T] : \mathbb{R}^d)$,

$$\liminf_{\epsilon\to 0} \epsilon \log \mu_\epsilon(G) \ge - \inf_{h\in G} I^x(h).$$

Chapter 2

Branching Processes

2.1 Branching Processes Overview

Branching processes is one of the classical areas in applied probability and is by now still an active area of research. Generally, they may be thought of as mathematical models for the evolution of stochastic systems whose dynamics consist of components reproducing new members according to some probability laws. There are two major differences between branching processes and deterministic population models. For one thing, the dynamics of the former are described by randomness through some probabilistic laws on the number of offspring and on the life spans, whereas the the evolution of the latter is non-random. For the other, the former are individual-based models for the growth of populations because the propagation mechanisms of the former are described by microscopic behavior of the systems. Besides the mathematical interest on these, it is not surprising that there are many fruitful applications of this stochastic process in physics, biology, chemistry and elsewhere. For example, neutron fluctuations and cosmic ray cascade in physics, polymerase chain reaction and cell kinetics in biology, spread of surnames in genealogy, chemical chain reactions in chemistry have been studied through various branching processes. Although the term “branching processes” was coined by Kolmogorov and Dmitriev [24] in 1974, the study of this subject has a much longer history. The first research in this process was about the probability of extinction of the surnames in the British peerage, which was conducted by Francis Galton [18] in 1873. Later on, he and Henry Watson successfully solved the problem (see [34] and [35]). The model they used is called the Galton-Watson branching process, which is the simplest and oldest branching process among all. Since the success of the investigation into the Galton-Watson branching process, the study of this area has been growing extensively.

In the following section, the Galton-Watson branching process will be mathematically defined and explained in more details. Here we give a rough idea of branching processes for example. Consider a population model that consists of human, animal species, or particles in general. Suppose that an ancestor of the population has a life span of length one and gives birth to its children at the end of that unit of time and then dies immediately. One can include the case that there is a positive probability that no birth is given. The number of children is non-negative integer-valued following to a law of certain offspring distribution. The ancestor is at the 0th generation and his children are of the first generation and so on. They live for a unit amount of time and give birth to some number of children according to the same probability law at the end of the unit time right before they die. The procedure continuous in this manner. For each of the individuals, the number of children is independent of that of the other individuals. This property is called the branching property in some literature. If the size of population reaches zero at some point, then the procedure stops. This could be regarded as extinction of the population. If the process does not become extinct in finite time, then the process will go on forever. The Galton-Watson branching process is the simplest branching process among all the others because it has a discrete-time framework and because the number of children of each individual is independent of the others. This simple branching process has three major features:

1. All individuals in the same generation have identical offspring distribution.

2. Individuals do not affect each other's number of children.

3. The offspring distribution remains the same in the dynamics.

The first characteristic says that there is no difference among different individuals with respect to their reproduction distribution. The second property says that their reproduction are independent. The third feature means there is no difference between reproduction distribution in different generations. It is these three basic features that make the Galton-Watson branching process mathematically easier to handle than other more sophisticated branching processes.

Nowadays there are many variants of the Galton-Watson branching process. One generalization is to consider several types of individuals in the process. The type of a particle is considered as a fixed attribute throughout its life span. The type, for instance, could be its genotype, mutant type, or any other characters of interest. The offspring distribution may be affected by the type. Particles of the same type have the same offspring distribution, however, individuals of different types may have different offspring distributions. The branching process that has several types of individuals is called a multi-type branching process. Other than having several types in the systems, a multi-type branching process has the same reproduction mechanism as a single-type Galton-Watson branching process. Particles live for one unit of time and give birth to their children according to a fixed probability law determined by their type and with no influence on each other. The first study of multi-type processes appears to be tackled by Kolmogrov and Dmitriev [24] in 1947. They considered a process in a continuous-time framework.

Another generalization is by relexing the assumption that particles reproduce independently from one another. In this direction, the most common dependence is population-size dependence in the process. In a population-size dependent branching process, the offspring distributions depend on the size of the whole population. It is often assumed that the reproduction reduces as the population increases to fit the real-world phenomenon due to the limitations of resources, although it is not mathematically motivated. The first research in a discrete-time, population-size dependent branching process was conducted by Klebaner [22] in 1983. Boiko [8] studied a population-size dependent branching process in a continuous-time framework. Another important dependence is considered as interaction between particles. Notice that population-size dependence is a global dependence because the population as a whole affects the reproduction behavior. On the other hand, interaction between particles is a local dependence because it is often assumed that the interactions occur between type-attractive particles, type-repulsive particles, sibling particles, or neighborhood particles under spatial settings. This kind of interactive dependence between particles makes the system complex and the study of this area is active and still developing.

One of the unrealistic assumptions in the Galton-Watson branching processes is the fixed offspring distribution throughout different generations. This assumption implies that the environment for reproduction is unchanged throughout the time. In the branching processes discussed above, the law governing the number of descendants is unchanged for all generations, however, in the real world, reproduction will be affected by many factors either inside or outside the system. Let us think of a population model of plants having a life cycle of one year. When the weather conditions are good for the growth of the plants, the population of the plants has a higher probability of growth. Since the weather conditions may differ from year to year, the probabilities of reproduction rate of the plants will be different each year. In this case, the weather conditions are the environment and they are not fixed throughout the lifetime of the plants. Therefore, a natural generalization of the Galton-Watson branching process may be by considering a process living in a changing environment. If the environment changes in a deterministic manner, then the branching process is said to be in a deterministically varying environment, while if the environments are chosen to be random, then the branching process is said to be in random environments. The latter is obviously more complicated to analyze than the former because in the latter case the changes in offspring distribution is not deterministic, and a variety of scenarios from different laws are possible. In fact, we can view branching process in a deterministically varying environment as a special case of branching process in random environments, for which a particular realization of offspring distributions under certain environment occurs with probability one. Furthermore, if the environment is based on i.i.d. random variables, then it is called a branching process in the i.i.d. random environment. This process was first studied by Smith and Wilkinson [32]. Athreya and Karlin generalized it to a non-i.i.d. case in [2] and [3]. Since then, the branching process in a random environment has been an extensively growing research subject due to the extra randomness from environment and increased flexibility for applied purposes.

In the previous branching processes, the systems are considered to be closed or isolated because the population only consists of particles generated inside the system and they cannot move in and out of the system. Therefore, a generalization may be done through allowing migration into the system. In a branching process with immigration, a random number of particles may immigrate into the system during each reproduction period. The number of immigrants follows a probabilistic law and it is often assumed that the immigrants have the same offspring distribution as that of those original particles in the system. It is worth mentioning that since immigration does not stop, the branching process cannot be assumed extinct although the process might have temporary periods of extinction. The other type of migration is emigration. In a branching process with emigration, the particles could choose to leave the system according to a probabilistic law right after their births. If their choice of leaving is independent of one another, then this does not introduce any added complicition mathematically because in this case the emigration could be viewed as an immediate death and hence we could ignore the emigrants as if they are not offspring and are not to be counted. If the choice of emigration is independent of one another but is dependent on the size of the population, then the process could be treated as a population-size dependent branching process. The more difficult, but somewhat more interesting, case may be that the decision of leaving is dependent on each other in the same generation.

Another direction for a variant is in the aspect of a time framework in which the process is defined. The preceding branching processes all live in a discrete-time framework. In the Galton-Watson branching process, each individual lives one unit of time, however, a natural extension is to allow the life span of the individuals to be a continuous random variable. The counterpart of the Galton-Watson branching process is the so-called continuous-time Markov branching process, which is the simplest branching process in a continuous-time framework. More specifically, individuals now live in a continuous-time frame and the life span is a random variable following identical and independent exponential distributions, which makes the process Markovian. If we relax the assumption of exponential lifetime, then we obtain a so-called age-dependent branching process, in which the lifetime is not necessarily exponentially distributed, and hence the resulting process may not be Markovian. Another more general continuous-time branching process is the Sevastyanov process, in which not only the life span is arbitrarily distributed but also the offspring distribution depends on the age of the individuals. Much of the early study on a continuous-time branching process was done by Kolmogrov and Dmitriev [24] and Sevastyanov [31].

There are some other variations of the Galton-Watson branching process. In a two-gender branching process, which was introduced by Daley [11], the reproduction only happens through mating of the male and female individuals. In this process, it is convenient to represent the size of population by the number of couples instead of the number of individuals. And the formation of couples is via a so-called mating function, which specifies the total number of couples deterministically. In a continuous-state branching process, the number of offsprings is a non-negative real number instead of a non-negative integer. This process was first invented by Jirina [21]. It is worth noting that the class of continuous-state branching processes with immigration coincides with the class of affine processes, which has some important applications in finance.

2.2 Preliminaries of the Galton-Watson Branching Processes

This section provides some well-known results on the Galton-Watson branching processes. Most of these results are needed in this dissertation and some others are listed for completeness.

A Galton-Watson branching process is a discrete-time Markov chain $\{Z_n\}_{n=0}^{\infty}$ on the space of non-negative integers. Let the offspring distribution be

$$\{p_k\}_{k=0}^{\infty}, \quad p_k \geq 0, \quad \sum_{k=0}^{\infty} p_k = 1.$$

The branching mechanism follows the offspring distribution that does not vary from individual to individual nor from generation to generation. More specifically, the process begins at time 0 with Z_0 ancestors, after one unit of time, each of them splits independently of each other into a random number of individuals according to the common probability law $\{p_k\}_{k=0}^{\infty}$, i.e., an individual gives no birth with probability p_0 and he has one child with probability p_1, and so on. The total number of children from all the ancestors is denoted as Z_1, which is the sum of Z_0 independent random variables, each with probability law $\{p_k\}_{k=0}^{\infty}$. The ancestors are said to be in the 0th generation and their children constitute the first generation. The particles in the first generation also have a unit life span. After one unit of time, each of them splits independently of each other and of the history into a random number of individuals following the same probability law $\{p_k\}_{k=0}^{\infty}$. The total number of their children is denoted as Z_2. And these children is said to be in the second generation. This way, the number of particles in the n-th generation is denoted as Z_n. The procedure continues unless $Z_n = 0$ for some positive integer n. If $Z_n = 0$ for some n, the branching process is said to be extinct. Apparently, $Z_{n'} = 0$ for all $n' \geq n$ if $Z_n = 0$. Thus, 0 is an absorbing state of the Markov chain $\{Z_n\}_{n=0}^{\infty}$.

In the literature, when the number of original ancestors needs to be emphasized, a Galton-Watson branching process with $Z_0 = j$ many original ancestors would be written as $\{Z_n^{(j)}\}_{n=0}^{\infty}$. And usually if $Z_0 = 1$, then it is customary to denote the process as $\{Z_n\}_{n=0}^{\infty}$ instead of $\{Z_n^{(1)}\}_{n=0}^{\infty}$. In most of this study, we assume that $Z_0 = 1$. Hence in the remaining part of this dissertation, when there is no confusion, a Galton-Watson branching process with only one ancestor will be written as $\{Z_n\}_{n=0}^{\infty}$.

An important feature of the Galton-Watson branching process is that the number of children of each individual at any given generation is independent of the other individuals existing in the same generation and of the whole history of the process. This fundamental feature of the Galton-Watson branching process leads to the following additive property. Suppose that there are $Z_0 = j$ many ancestors, then the process $\{Z_n^{(j)}\}_{n=0}^{\infty}$ is the sum of j independent copies of the Galton-Watson branching process $\{Z_n\}_{n=0}^{\infty}$.

The transition mechanism of the Galton-Watson branching process is given by the following transition probabilities

$$p(i,j) = P(Z_{n+1} = j | Z_n = i) = \begin{cases} p_j^{*i} & \text{if} \quad i \geq 1 \,\text{and}\, j \geq 0, \\ 1 & \text{if} \quad i = 0 \,\text{and}\, j = 0, \\ 0 & \text{if} \quad i = 0 \,\text{and}\, j > 0, \end{cases}$$

where $\{p_j^{*i}\}_{j=0}^{\infty}$ is the i-fold convolution of the offspring distribution $\{p_k\}_{k=0}^{\infty}$, i.e., it is the probability distribution of the sum of i many independent probability distributions $\{p_k\}_{k=0}^{\infty}$.

Another formulation of the Galton-Watson branching process may be done through the following recurrence equation

$$Z_{n+1} = \sum_{j=1}^{Z_n} X_{n,j}, \quad n = 0, 1, 2, \cdots \quad \text{and} \quad j = 1, 2, \cdots, Z_n,$$

where $X_{n,j}$ is the number of children of the j-th individual in the n-th generation and all $X_{n,j}$ are i.i.d. random variables over all n and j. Notice that since the offspring distribution does not vary over the different generations, in this dissertation, we use the ease notion X_j instead of $X_{n,j}$.

Let the probability generating function of the Galton-Watson branching process be

$$f(s) = E(s^{Z_1} \mid Z_0 = 1) = \sum_{k=0}^{\infty} p_k s^k, \quad |s| \leq 1.$$

In general, s can be negative and complex, however, in this dissertation, we focus on the case that $s \in [0, 1]$. It is a very useful tool in the study of branching processes. Also note that since in this study we assume $Z_0 = 1$ most of the time, we use the ease notation $E(s^{Z_1})$ to replace $E(s^{Z_1}|Z_0 = 1)$ sometimes when there is no confusion. Define the iterated compositions of the probability generating function by

$$f_{n+1}(s) = f[f_n(s)], \quad n = 1, 2, \cdots, \quad \text{and } f_0(s) = s.$$

It follows that $f_{n+m}(s) = f_n[f_m(s)]$, $n, m = 0, 1, \cdots$, and in particular,

$$f_{n+1}(s) = f_n[f(s)]. \tag{2.1}$$

Notice that by the branching property and the identical distribution, when the process starts with $Z_0 = j$ many ancestors, we have

$$\begin{aligned} \sum_{k=0}^{\infty} p(j,k)s^k &= E(s^{Z_1} \mid Z_0 = j) \\ &= E(s^{X_1+X_2+\cdots+X_j}) \\ &= [E(s^{Z_1} \mid Z_0 = 1)]^j \\ &= (f(s))^j. \end{aligned} \tag{2.2}$$

Let $p_n(i,k)$ denote the n-step transition probability from the state i to the state k. By Chapman-Kolmogorov equation, Fubini's Theorem, and (2.2),

$$\begin{aligned}
E(s^{Z_n}|Z_0=1) &= \sum_{k=0}^{\infty} p_n(1,k)s^k \\
&= \sum_{k=0}^{\infty}\sum_{l=0}^{\infty} p_{n-1}(1,l)p(l,k)s^k \\
&= \sum_{l=0}^{\infty} p_{n-1}(1,l)\sum_{k=0}^{\infty} p(l,k)s^k \\
&= \sum_{l=0}^{\infty} p_{n-1}(1,l)[f(s)]^l,
\end{aligned}$$

where Fubini's Theorem can be applied because of the fact that the terms $p_{n-1}(1,l)p(l,k)s^k$ are non-negative for all $l,k \in \mathbb{N}\cup\{0\}$.

Note that by applying the above procedure repeatedly and induction,

$$\begin{aligned}
\sum_{l=0}^{\infty} p_{n-1}(1,l)[f(s)]^l &= \sum_{l=0}^{\infty}\sum_{j=0}^{\infty} p_{n-2}(1,j)p(j,l)(f(s))^l \\
&= \sum_{j=0}^{\infty} p_{n-2}(1,j)\sum_{l=0}^{\infty} p(j,l)(f(s))^l \\
&= \sum_{j=0}^{\infty} p_{n-2}(1,j)[f(f(s))]^j \\
&= \sum_{i=0}^{\infty} p(1,i)[f_{n-1}(s)]^i \\
&= f[f_{n-1}(s)] \\
&= f_n(s).
\end{aligned}$$

Therefore, this shows that $f_n(s)$ is the probability generating function of Z_n, i.e.,

$$f_n(s) = E(s^{Z_n}|Z_0=1). \tag{2.3}$$

Combing (2.2) and (2.3), we have

$$[f_n(s)]^i = \sum_{k=0}^{\infty} p_n(i,k)s^k, \quad i \geq 1.$$

Since all the information for the transition probabilities is contained in $f_n(s)$, the probability generating function of Z_n, the study of this function is essential. In fact, the moments, if exist, of a Galton-Watson branching process could be calculated by taking left-sided derivatives of f_n at $s = 1$. For example, the mean of the offspring distribution $\{p_k\}_{k=0}^{\infty}$, i.e., the expected number of children of each particle, is denoted by m and could be calculated as $m = E(Z_1) = \sum_{k=0}^{\infty} p(1,k)k = \sum_{k=0}^{\infty} p(1,k)k1^{k-1} = \sum_{k=0}^{\infty} \lim_{s\to 1^-} p(1,k)ks^{k-1} = \lim_{s\to 1^-} \sum_{k=0}^{\infty} p(1,k)ks^{k-1}$ $= \lim_{s\to 1^-} \sum_{k=0}^{\infty} \frac{d}{ds}[p(1,k)s^k] = \lim_{s\to 1^-} \frac{d}{ds}[\sum_{k=0}^{\infty} p(1,k)s^k] = \lim_{s\to 1^-} f'(s) = f'(1^-)$, where the calculation above is justified by uniform convergence in the interval of convergence of the power series, and by dominated convergence theorem since $|p(1,k)ks^{k-1}| \leq p(1,k)k$ for $s \in [-1, 1]$ and $\sum_{k=0}^{\infty} p(1,k)k$ is finite by the assumption that the first moment exists. The Galton-Watson branching process is said to be sub-critical if $m < 1$, critical if $m = 1$, and super-critical if $m > 1$. The expected size of generation n is $E(Z_n) = m^n$, which can be computed by the calculation similar to the above one and differentiating 2.1 at $s = 1$.

Suppose that $p_0 + p_1 < 1$ and $p_k \neq 1$ for any $k \in \mathbb{N} \cup \{0\}$. Let q be the smallest root of the equation, $f(s) = s$ in $[0, 1]$. Clearly, $q = 0$ when $p_0 = 0$. The following gives some elementary properties of $f(s)$, the probability generating function of the offspring distribution.

(1) f is strictly convex and monotone increasing.

(2) $f(0) = \sum_{k=0}^{\infty} p(1,k)0^k = p(1,0) = p_0$ and $f(1) = \sum_{k=0}^{\infty} p(1,k)1^k = 1$.

(3) If $m \leq 1$, then $f(s) > s$ for all $0 \leq s < 1$.

(4) If $m > 1$, then $f(s) > s$ for all $0 \leq s < q$ and $f(s) = s$ for $s = q$ and $f(s) < s$ for all $q < s < 1$.

(5) If $m \leq 1$, then $q = 1$.

(6) If $m > 1$, then $q < 1$.

2.3 Asymptotic Behavior of the Galton-Watson Branching Processes

The study of the asymptotic behavior of the probability generating function of a Galton-Watson branching process is important since that can provide much information on the limit theorems about the process $\{Z_n\}_{n=0}^{\infty}$. Therefore, in this section, we provide some asymptotic results concerning $\{f_n\}_{n=0}^{\infty}$ and $\{Z_n\}_{n=0}^{\infty}$. Since these results will be used in our main theorems in the latter chapters, we provide the proofs here. Note that since in this study, the branching process is assumed to be super-critical, hence we mostly focus on the results regarding super-critical branching processes. The reader may find more results in [1] and [4].

The following result shows the convergence of f_n to q in an monotonic manner in two different regimes.

Proposition 2.1

(a) *For $s = q$ or $s = 1$, $f_n(s) = s$ for all n.*

(b) *For $0 \leq s < q$, $f_n(s) \uparrow q$ as $n \to \infty$.*

(c) *For $q < s < 1$, $f_n(s) \downarrow q$ as $n \to \infty$.*

Proof. Clearly, $f_n(q) = q$ for all n because $f(q) = q$. It is also obvious that $f_n(1) = 1$ for all n because $f(1) = 1$. Thus, (a) is shown.

For (b), let $0 < s < q$. Then by properties (3) and (4) in the section 2.2, $s < f(s)$. Since f is monotone increasing, $f(s) < f(q)$. Therefore, $s < f(s) < f(q)$, hence By monotonicity of f, we have

$$f(s) < f(f(s)) < f(f(q)), \text{ i.e., } f(s) < f_2(s) < f_2(q).$$

Similarly, we obtain the following relationship

$$s < f(s) < f_2(s) < f_3(s) < \cdots < f_n(s) < f_n(q) = q.$$

Hence, for $0 \leq s < q$, $f_n(s)$ converges increasingly to a finite number U, as $n \to \infty$, since it is an increasing sequence that is bounded above by q. Note that f is continuous at least on $[0, 1)$ since the power series f has the radius of convergence at least 1 and hence we have uniform convergence to f for $s \in [0, 1)$. Thus, for any fixed $s \in [0, q)$,

$$U = \lim_{n\to\infty} f_n(s) = \lim_{n\to\infty} f(f_{n-1}(s)) = f(\lim_{n\to\infty} f_{n-1}(s)) = f(U).$$

Therefore, $U = q$ since $f(U) = U \leq q$ and q is the smallest root of $f(s) = s$ in $[0, 1]$. This shows (b).

For (c), let $q < s < 1$. Then by the property (4) in section 2.2, $f(s) < s$. Since f is monotone increasing, $f(q) < f(s)$, hence $q < f(s) < s < 1$. By monotonicity of f, we have

$$f(q) < f(f(s)) < f(s) < f(1), \text{ i.e., } q < f_2(s) < f(s) < 1.$$

Similarly, we obtain the following relationship

$$q < f_n(s) < \cdots < f_2(s) < f(s) < 1.$$

Hence, for $q < s < 1$, $f_n(s)$ converges decreasingly to a non-negative number L, as $n \to \infty$, since it is an decreasing sequence that is bounded below by q. Note that $f_n(s) = f(f_{n-1}(s))$ and that f is a continuous function because it is a power series. Thus, for any fixed $s \in (q, 1]$,

$$L = \lim_{n\to\infty} f_n(s) = \lim_{n\to\infty} f(f_{n-1}(s)) = f(\lim_{n\to\infty} f_{n-1}(s)) = f(L).$$

Therefore, $L = q$ since $f(L) = L \leq q$ and q is the smallest root of $f(s) = s$ in $[0, 1]$ and there is no root in $(q, 1)$. This shows (c).

□

Recall that the original problem proposed by Galton was to find the probability of ultimate extinction, i.e., no particle exists after some finite number of generations. This is always a central question in the theory of branching processes. The probability of extinction of a Galton-Watson branching process is

$$P(Z_j = 0 \,\text{for some}\, j \geq 1) = \lim_{n\to\infty} P(Z_j = 0 \,\text{for some}\, j \leq n) = \lim_{n\to\infty} P(Z_n = 0) = \lim_{n\to\infty} f_n(0) = q.$$

Thus, the smallest root the equation, q of $f(s) = s$ in $[0, 1]$ is the probability of extinction for the process.

Note that if there are $Z_0 = l \geq 1$ ancestors, then by the branching property, the extinction probability of a Galton-Watson branching process is

$$P(Z_j^{(l)} = 0 \,\text{for some}\, j \geq 1) = q^l.$$

From proposition 2.1, we know that $\{f_n\}_{n=1}^{\infty}$ converges to q, however, it does not tell us the rate of convergence. The information about convergence rate is useful in the study of some finer limit theorems concerning $\{Z_n\}_{n=0}^{\infty}$. The following result gives the geometric rate of convergence of $f_n(s)$ for a super-critical process by analyzing $Q_n(s)$ defined as

$$Q_n(s) := \frac{f_n(s) - q}{(f'(q))^n} \quad \text{for } 0 \leq s < 1. \tag{2.4}$$

Theorem 2.1 *Assume that $m > 1$, then there exist real numbers $\{q_j\}_{j=0}^{\infty}$ such that*

$$\lim_{n\to\infty} \frac{f_n(s) - q}{(f'(q))^n} = \sum_{j=0}^{\infty} q_j s^j = Q(s) < \infty \quad \textit{for } 0 \le s < 1,$$

where $Q(s) = \lim_{n\to\infty} Q_n(s)$ for $0 \le s < 1$.
Moreover,

$$\lim_{n\to\infty} Q_n'(s) = Q'(s) < \infty \quad \textit{with } Q'(s) > 0 \quad \textit{for } 0 \le s < 1 \textit{ and } \lim_{s\to q} Q'(s) = 1.$$

Furthermore, $Q(s)$ is the unique solution of the functional equation

$$Q(f(s)) = f'(q)Q(s) \quad \textit{for } 0 \le s < 1, \quad \textit{with } Q(q) = 0.$$

Remark 2.1 *When $p_0 = 0$, $q = 0$ and $f'(q) = p_1$, and hence this theorem gives the following: Assume that $m > 1$ and $p_1 \neq 0$, then there exist real numbers $\{\hat{q}_j\}_{j=1}^{\infty}$ such that*

$$\lim_{n\to\infty} \frac{f_n(s)}{(p_1)^n} = \sum_{j=1}^{\infty} \hat{q}_j s^j = \hat{Q}(s) < \infty \quad \textit{for } 0 \le s < 1,$$

where $\hat{Q}(s) = \lim_{n\to\infty} \hat{Q}_n(s)$ for $0 \le s < 1$ and $\hat{Q}_n(s) = \frac{f_n(s)}{(p_1)^n}$ for $0 \le s < 1$.
Moreover,

$$\lim_{n\to\infty} \hat{Q}_n'(s) = \hat{Q}'(s) < \infty \quad \textit{with } \hat{Q}'(s) > 0 \quad \textit{for } 0 \le s < 1 \textit{ and } \lim_{s\to 0} \hat{Q}'(s) = 1.$$

Furthermore, $\hat{Q}(s)$ is the unique solution of the functional equation

$$\hat{Q}(f(s)) = p_1 \hat{Q}(s) \quad \textit{for } 0 \le s < 1, \quad \textit{with } \hat{Q}(0) = 0. \tag{2.5}$$

Proof. By the chain rule, the derivative of $Q_n(s)$ could be expressed as

$$\begin{aligned} Q_n'(s) &= \frac{f_n'(s)}{(f'(q))^n} \\ &= \frac{f'(f_{n-1}(s))f'(f_{n-2}(s))f'(f_{n-3}(s))\cdots f'(s)}{(f'(q))^n} \\ &= \prod_{j=0}^{n-1} \frac{f'(f_j(s))}{f'(q)} \\ &= \prod_{j=0}^{n-1} \{1 + [\frac{f'(f_j(s)) - f'(q)}{f'(q)}]\}. \end{aligned} \tag{2.6}$$

Recall that a result on infinite products says that if the series $\sum_{j=1}^{\infty} |a_j|$ is convergent, then the infinite products $\prod_{j=1}^{\infty}(1+a_j)$ is convergent. That is, $\prod_{j=1}^{\infty}(1+a_j)$ is finite and non-zero. Thus, now we want to show that $\sum_{j=0}^{\infty} |\frac{f'(f_j(s))-f'(q)}{f'(q)}| < \infty$, i.e., $\sum_{j=0}^{\infty} |f'(f_j(s)) - f'(q)| < \infty$, because this implies that $Q'(s) = \lim_{n\to\infty} Q'_n(s) = \prod_{j=0}^{\infty}\{1+[\frac{f'(f_j(s))-f'(q)}{f'(q)}]\}$ is finite and non-zero.

Fix s in $[0, 1)$. Let $\epsilon > 0$ be such that $q < q + \epsilon < 1$ and that $f'(q + \epsilon) < 1$. Recall that $f_n(s)$ converges to q, therefore, we can choose j_0 to be such that $f_j(s) < q + \epsilon$ for all $j \geq j_0$. By mean value theorem,

$$\frac{|f'(f_{j+j_0}(s)) - f'(q)|}{|f_{j+j_0}(s) - q|} \leq f''(q + \epsilon),$$

i.e., $|f'(f_{j+j_0}(s)) - f'(q)| \leq f''(q + \epsilon)|f_{j+j_0}(s) - q|$.

By using mean value theorem again,

$$\begin{aligned} \frac{|f_{j+j_0}(s) - q|}{|f_{j+j_0-1}(s) - q|} &= \frac{|f_{j+j_0}(s) - f_{j+j_0}(q)|}{|f_{j+j_0-1}(s) - f_{j+j_0-1}(q)|} \\ &= \frac{|f(f_{j+j_0-1}(s)) - f(f_{j+j_0-1}(q))|}{|f_{j+j_0-1}(s) - f_{j+j_0-1}(q)|} \\ &\leq f'(q + \epsilon), \end{aligned}$$

i.e., $|f_{j+j_0}(s) - q| \leq f'(q + \epsilon)|f_{j+j_0-1}(s) - q|$. Therefore, $|f'(f_{j+j_0}(s)) - f'(q)| \leq f''(q + \epsilon)f'(q + \epsilon)|f_{j+j_0-1}(s) - q|$.
Similarly,

$$\begin{aligned} |f'(f_{j+j_0}(s)) - f'(q)| &\leq f''(q + \epsilon)(f'(q + \epsilon))^j |f_{j_0}(s) - q| \\ &< f''(q + \epsilon)(f'(q + \epsilon))^j |q + \epsilon - q| \\ &= \epsilon f''(q + \epsilon)(f'(q + \epsilon))^j. \end{aligned}$$

Thus,

$$\begin{aligned} \sum_{j=0}^{\infty} |f'(f_{j+j_0}(s)) - f'(q)| &\leq \epsilon f''(q + \epsilon) \sum_{j=0}^{\infty} (f'(q + \epsilon))^j \\ &= \epsilon f''(q + \epsilon) \frac{1}{1 - f'(q + \epsilon)} \\ &< \infty. \end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{j=0}^{\infty}|f'(f_j(s))-f'(q)| &= \sum_{j=0}^{j_0-1}|f'(f_j(s))-f'(q)|+\sum_{j=j_0}^{\infty}|f'(f_j(s))-f'(q)| \\ &< \infty.\end{aligned}$$

This implies that

$$\lim_{n\to\infty} Q_n'(s) = Q'(s) \text{ is finite and non-zero for } 0 \le s < 1. \tag{2.7}$$

It is clear that $Q_n'(s) > 0$ for $0 \le s < 1$. We now show that $\lim_{s\to q} Q'(s) = 1$. By monotonicity of f_n and f', we have

$$\begin{aligned}-\log Q'(s) &= -\sum_{j=0}^{\infty}\log\frac{f'(f_j(s))}{f'(q)} \\ &= \sum_{j=0}^{\infty}\log\frac{f'(q)}{f'(f_j(s))} \\ &\le \sum_{j=0}^{\infty}\log\frac{f'(q)}{f'(f_j(0))} \\ &\le \sum_{j=0}^{\infty}[\frac{f'(q)}{f'(f_j(0))}-1] \\ &\le \sum_{j=0}^{\infty}\frac{f'(q)-f'(f_j(0))}{f'(f_0(0))} \\ &= \sum_{j=0}^{\infty}\frac{f'(q)-f'(f_j(0))}{p_1} \\ &= \frac{1}{p_1}\sum_{j=0}^{\infty}|f'(f_j(0))-f'(q)| < \infty.\end{aligned}$$

Notice that by the continuity of f', $\lim_{s\to q} -\log\frac{f'(f_j(s))}{f'(q)} = -\log\frac{f'(f_j(q))}{f'(q)} = 0$. Also note that $-\log\frac{f'(f_j(s))}{f'(q)} \le \log\frac{f'(q)}{f'(f_j(0))}$ and that $\sum_{j=0}^{\infty}\log\frac{f'(q)}{f'(f_j(0))} < \infty$. Hence by dominated convergence theorem,

$$\begin{aligned}
\lim_{s\to q} -\log Q'(s) &= \lim_{s\to q} \sum_{j=0}^{\infty} -\log \frac{f'(f_j(s))}{f'(q)} \\
&= \sum_{j=0}^{\infty} \lim_{s\to q} -\log \frac{f'(f_j(s))}{f'(q)} \\
&= 0.
\end{aligned}$$

This implies that $\lim_{s\to q} Q'(s) = 1$.

Note that

$$Q(s) = \int_q^s Q'(t)dt \quad \text{for } 0 \le s < 1.$$

From $Q_n(q) = 0$,

$$Q_n(s) = Q_n(s) - Q_n(q) = \int_q^s Q_n'(t)dt.$$

Note that

$$Q_n'(t) \le \begin{cases} Q'(t) & \text{if} \quad q < t < 1 \\ 1 & \text{if} \quad 0 \le t \le q \end{cases} \quad \text{for all n} \in \mathbb{N}$$

and that

$$\begin{cases} \int_q^s Q'(t)dt = Q(s) < \infty \\ \int_q^s 1dt = s - q < \infty. \end{cases}$$

Hence by dominated convergence theorem,

$$\begin{aligned}
\lim_{n\to\infty} Q_n(s) &= \lim_{n\to\infty} \int_q^s Q_n'(t)dt \\
&= \int_q^s \lim_{n\to\infty} Q_n'(t)dt \\
&= \int_q^s Q'(t)dt \\
&= Q(s) - Q(q) \\
&= Q(s).
\end{aligned}$$

Notice that

$$\begin{aligned}
Q_n(s) &= \frac{f_n(s)-q}{(f'(q))^n} \\
&= \frac{(\sum_{j=0}^{\infty} P(Z_n=j|Z_0=1)s^j)-q}{(f'(q))^n} \\
&= \frac{\sum_{j=0}^{\infty}[P(Z_n=j|Z_0=1)-b_j]s^j}{(f'(q))^n} \\
&= \sum_{j=0}^{\infty}\frac{[P(Z_n=j|Z_0=1)-b_j]}{(f'(q))^n}s^j,
\end{aligned}$$

where $b_j = q$ if $j=0$ and $b_j=0$ otherwise. Thus, $Q_n(s)$ is a power series for $0 \le s < 1$. Since $Q(s)$ is a limit of a sequence of power series, it is also a power series, say,

$$Q(s) = \sum_{j=0}^{\infty} q_j s^j < \infty \quad \text{for } 0 \le s < 1.$$

Now we want to show that $Q(f(s)) = f'(q)Q(s)$ for $0 \le s < 1$.

$$\begin{aligned}
Q_n(f(s)) &= \frac{f_n(f(s))-q}{(f'(q))^n} \\
&= \frac{f_{n+1}(s)-q}{(f'(q))^n} \\
&= f'(q)\frac{f_{n+1}(s)-q}{(f'(q))^{n+1}} \\
&= f'(q)Q_{n+1}(s).
\end{aligned}$$

Thus, $\lim_{n\to\infty} Q_n(f(s)) = \lim_{n\to\infty} f'(q)Q_{n+1}(s)$ and hence $Q(f(s)) = f'(q)Q(s)$.

To prove uniqueness, we assume that $Q^1(s)$ and $Q^2(s)$ are two of such solutions. Then by repeated using $Q^i(f(s)) = f'(q)Q^i(s)$, $i = 1, 2$, and by the triangle inequality,

$$\begin{aligned}
0 &\leq |Q^1(s) - Q^2(s)| \\
&= \frac{1}{f'(q)}|Q^1(f(s)) - Q^2(f(s))| \\
&= \frac{1}{(f'(q))^2}|Q^1(f_2(s)) - Q^2(f_2(s))| \\
&= \frac{1}{(f'(q))^n}|Q^1(f_n(s)) - Q^2(f_n(s))| \\
&= \frac{Q_n(s)}{f_n(s) - q}|Q^1(f_n(s)) - Q^2(f_n(s))| \\
&= |Q_n(s)|\Big|\frac{Q^1(f_n(s)) - Q^2(f_n(s))}{f_n(s) - q}\Big| \\
&\leq |Q_n(s)|\Big(\Big|\frac{Q^1(f_n(s))}{f_n(s) - q} - 1\Big| + \Big|1 - \frac{Q^2(f_n(s))}{f_n(s) - q}\Big|\Big).
\end{aligned}$$

Since $\lim_{n\to\infty} \frac{Q^i(f_n(s))}{f_n(s)-q} = \lim_{s\to q} \frac{Q^i(s)}{s-q} = \lim_{s\to q}(Q^i)'(s) = 1$ by L'Hopital's Rule rule, we have, by taking limits on both sides of the inequality above,

$$|Q^1(s) - Q^2(s)| = 0 \quad \text{for } 0 \leq s < 1.$$

Therefore, $Q(s)$ is the unique solution of the functional equation

$$Q(f(s)) = f'(q)Q(s) \quad \text{for } 0 \leq s < 1.$$

□

Since the sequence of probability generating functions $\{f_n\}_{n=1}^{\infty}$ of a super-critical Galton-Watson branching process decays geometrically and $f_n(s) = \sum_{j=0}^{\infty} p_n(i,j)s^j$, it is expected that the sequence of transition probabilities $\{p_n(i,j)\}_{n=1}^{\infty}$ also has a geometric rate of decay.

Theorem 2.2 *Assume that $m > 1$, then*

$$\textit{for all } i, j \geq 1, \ \lim_{n\to\infty} \frac{p_n(i,j)}{(f'(q))^n} = iq^{i-1}q_j, \quad \textit{and}$$

$$\textit{for all } i \geq 1 \textit{ and } j = 0, \ \lim_{n\to\infty} \frac{p_n(i,j) - q^i}{(f'(q))^n} = iq^{i-1}q_j.$$

Remark 2.2 *In particular, if $p_0 = 0$, then $q = 0$ and $f'(q) = p_1$. Hence this theorem translates to the following:*

Assume that $m > 1$ and $p_1 \neq 0$, then

$$\text{for } j \geq 0, \quad \lim_{n\to\infty} \frac{p_n(i,j)}{p_1^n} = \begin{cases} q_j & \text{if} \quad i = 1, \\ 0 & \text{if} \quad i > 1, \end{cases}$$

and moreover, for $i \geq 1$ and $j \geq 1$,

$$\lim_{n\to\infty} \frac{p_n(i,j)}{(p_1^n)^i} = q_{ij}, \text{ where } q_{ij} \text{ satisfies } \sum_{j=1}^{\infty} q_{ij} s^j = (\sum_{k=1}^{\infty} q_k s^k)^i \quad \text{for } 0 \leq s < 1.$$

Proof. From a series expansion, we have

$$\frac{(f_n(s))^i - q^i}{(f'(q))^n} = \frac{f_n(s) - q}{(f'(q))^n}[(f_n(s))^{i-1} + (f_n(s))^{i-2}q + (f_n(s))^{i-3}q^2 + \cdots + q^{i-1}].$$

If we take limit on both sides, we reach

$$\begin{aligned} \lim_{n\to\infty} \frac{(f_n(s))^i - q^i}{(f'(q))^n} &= \lim_{n\to\infty} \frac{f_n(s) - q}{(f'(q))^n}[(f_n(s))^{i-1} + (f_n(s))^{i-2}q + (f_n(s))^{i-3}q^2 + \cdots + q^{i-1}] \\ &= Q(s)[q^{i-1} + q^{i-2}q + q^{i-3}q^2 + \cdots + q^{i-1}] \\ &= Q(s)iq^{i-1} \\ &= (\sum_{j=0}^{\infty} q_j s^j) iq^{i-1} \\ &= \sum_{j=0}^{\infty} iq^{i-1} q_j s^j < \infty \quad \text{for } 0 \leq s < 1, \end{aligned}$$

since $\sum_{j=0}^{\infty} q_j s^j < \infty \quad 0 \leq s < 1$.

On the other hand, from the power series expansion of $f_n(s)$, we see that

$$\begin{aligned} \lim_{n\to\infty} \frac{(f_n(s))^i - q^i}{(f'(q))^n} &= \lim_{n\to\infty} \frac{\sum_{j=0}^{\infty} p_n(i,j)s^j - q^i}{(f'(q))^n} \\ &= \lim_{n\to\infty} \frac{\sum_{j=0}^{\infty} [p_n(i,j) - c_j]s^j}{(f'(q))^n} \\ &= \lim_{n\to\infty} \sum_{j=0}^{\infty} \frac{p_n(i,j) - c_j}{(f'(q))^n} s^j \\ &= \sum_{j=0}^{\infty} \lim_{n\to\infty} \frac{p_n(i,j) - c_j}{(f'(q))^n} s^j, \end{aligned}$$

$$\text{where} \quad c_j = \begin{cases} q^i & \text{if} \quad j = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that the interchange of summation and limit above is justified by the uniform convergence of power series in the interval of convergence. By comparing coefficients, we obtain that

$$\text{for } i \geq 1 \text{ and } j \geq 1, \ \lim_{n\to\infty} \frac{p_n(i,j)}{(f'(q))^n} = iq^{i-1}q_j, \quad \text{and}$$

$$\text{for } i \geq 1 \text{ and } j = 0, \ \lim_{n\to\infty} \frac{p_n(i,j) - q^i}{(f'(q))^n} = iq^{i-1}q_j.$$

□

The next result indicates that when p_0 and p_1 are both zero, the rate of convergence of $f_n(s)$ is super-geometric. Note that we use the notation "≈" to denote the approximation.

Theorem 2.3 *Let $p_0 = p_1 = 0$ and $l = inf\{j : j \geq 2, p_j \neq 0\}$. Then*

$$f_n(s) = s^{l^n} p_l^{\sum_{j=0}^{n-1} l^j} (R_n(s))^{l^n}, \tag{2.8}$$

where $R_n(s) = \prod_{j=0}^{n-1}(1 + \gamma g(f_j(s)))^{\frac{1}{l^{j+1}}}$ uniformly converges to $R(s)$ on $[0,1]$ with $R(0) = 1$ and $R(1) < \infty$, where $\gamma = \frac{1-p_l}{p_l}$ and $g(s) = \sum_{j=l+1}^{\infty} \frac{p_j}{1-p_l} s^{j-l}$.
Moreover,

$$\lim_{n\to\infty} \Big(\frac{R_n(s)}{R(s)}\Big)^{l^n} = 1 \quad \textit{for } 0 \leq s < 1 \tag{2.9}$$

and hence

$$f_n(s) \approx p_l^{\frac{-1}{l-1}} \big(p_l^{\frac{1}{l-1}} sR(s)\big)^{l^n} \quad \textit{for large } n.$$

Furthermore, $R(s)$ is the unique solution of the functional equation

$$f(s)R(f(s)) = p_l(sR(s))^l,$$

subject to the conditions $R(0) > 0$ and R is continuous on $[0,1]$.

Proof. Define $\gamma := \frac{1-p_l}{p_l}$ and $g(s) := \sum_{j=l+1}^{\infty} \frac{p_j}{1-p_l} s^{j-l}$. Then probability generating function $f(s)$ could be written as

$$f(s) = \sum_{j=0}^{\infty} p_j s^j = p_l s^l (1 + \gamma g(s)).$$

Apparently, g is a probability generating function and $0 < \gamma < \infty$. Thus,

$$f_n(s) = f(f_{n-1}(s)) = p_l(f_{n-1}(s))^l (1 + \gamma g(f_{n-1}(s))). \tag{2.10}$$

Let $h_n(s) = (f_n(s))^{\frac{1}{l^n}}$, then by iterating the identity,

$$\begin{aligned}
h_n(s) &= (f_n(s))^{\frac{1}{l^n}} \\
&= [p_l(f_{n-1}(s))^l(1+\gamma g(f_{n-1}(s)))]^{\frac{1}{l^n}} \\
&= [p_l(1+\gamma g(f_{n-1}(s)))]^{\frac{1}{l^n}} h_{n-1}(s) \\
&= [p_l(1+\gamma g(f_{n-1}(s)))]^{\frac{1}{l^n}} [p_l(1+\gamma g(f_{n-2}(s)))]^{\frac{1}{l^{n-1}}} h_{n-2}(s) \\
&= s\prod_{j=0}^{n-1}[p_l(1+\gamma g(f_j(s)))]^{\frac{1}{l^{j+1}}}.
\end{aligned}$$

Thus,

$$f_n(s) = (h_n(s))^{l^n} = s^{l^n} p_l^{\sum_{j=0}^{n-1} l^j} (R_n(s))^{l^n},$$

where $R_n(s) = \prod_{j=0}^{n-1}(1+\gamma g(f_j(s)))^{\frac{1}{l^{j+1}}}$. Clearly, $R_n(s)$ is increasing in n for $0 \leq s \leq 1$ and $g(1) = 1$. Let

$$R(s) = \prod_{j=0}^{\infty}(1+\gamma g(f_j(s)))^{\frac{1}{l^{j+1}}} \quad \text{for } 0 \leq s \leq 1.$$

Note that the infinite product $R(s)$ exists since

$$\begin{aligned}
\log R(s) &= \sum_{j=0}^{\infty}\frac{1}{l^{j+1}}\log(1+\gamma g(f_j(s))) \\
&\leq \log(1+\gamma)\sum_{j=0}^{\infty}\frac{1}{l^{j+1}} \\
&< \infty.
\end{aligned}$$

We want to show that $\lim_{n\to\infty}\sup_{0\leq s\leq 1}|R_n(s)-R(s)| = 0$ because this implies $R_n(s)$ converges uniformly to $R(s)$ for $0 \leq s \leq 1$. This is equivalent to show that $\lim_{n\to\infty}\sup_{0\leq s\leq 1}\log\frac{R(s)}{R_n(s)} = 0$. Notice that

$$\begin{aligned}
0 &\leq \log\frac{R(s)}{R_n(s)} \\
&= \sum_{j=n}^{\infty}\frac{1}{l^{j+1}}\log(1+\gamma g(f_j(s))) \\
&\leq \log(1+\gamma)\sum_{j=n}^{\infty}\frac{1}{l^{j+1}}
\end{aligned}$$

and hence

$$\lim_{n\to\infty}\sup_{0\leq s\leq 1}\log\frac{R(s)}{R_n(s)} = 0,$$

since $l \geq 2$ by definition. Therefore, $R_n(s)$ converges uniformly to $R(s)$ for $0 \leq s \leq 1$.

To show that $\lim_{n\to\infty} \left(\frac{R_n(s)}{R(s)}\right)^{l^n} = 1$ for $0 \leq s < 1$, we first realize that

$$\begin{aligned} 0 &\leq l^n \log \frac{R(s)}{R_n(s)} \\ &= l^n \sum_{j=n}^{\infty} \frac{1}{l^{j+1}} \log(1 + \gamma g(f_j(s))) \\ &= \sum_{r=0}^{\infty} \frac{1}{l^{r+1}} \log(1 + \gamma g(f_{n+r}(s))). \end{aligned}$$

Secondly, $g(0) = 0$ and g is continuous since it is a power series. Observe that, for $0 \leq s < 1$, $g(s) < 1$ and $1 \leq 1 + \gamma g(f_{n+r}(s)) < 1 + \gamma$. Then, for $0 \leq s < 1$, $|\frac{1}{l^{r+1}} \log(1 + \gamma g(f_{n+r}(s)))| < \frac{1}{l^{r+1}} \log(1+\gamma)$ for all n and $\sum_{r=0}^{\infty} \frac{1}{l^{r+1}} \log(1+\gamma) = \frac{1}{l-1} \log(1+\gamma) < \infty$. Since g and f are continuous and $g(0) = 0$, $lim_{n\to\infty} \frac{1}{l^{r+1}} \log(1 + \gamma g(f_{n+r}(s))) = 0$. Hence by dominated convergence theorem,

$$\begin{aligned} \lim_{n\to\infty} l^n \log \frac{R(s)}{R_n(s)} &= \lim_{n\to\infty} \sum_{r=0}^{\infty} \frac{1}{l^{r+1}} \log(1 + \gamma g(f_{n+r}(s))) \\ &= \sum_{r=0}^{\infty} \lim_{n\to\infty} \frac{1}{l^{r+1}} \log(1 + \gamma g(f_{n+r}(s))) \\ &= 0. \end{aligned}$$

Therefore, $\lim_{n\to\infty} \left(\frac{R_n(s)}{R(s)}\right)^{l^n} = 1$ for $0 \leq s < 1$. Hence, $\lim_{n\to\infty} \frac{f_n(s) s^{-l^n} p_l^{-(1+l+\cdots+l^{n-1})}}{(R(s))^{l^n}} = 1$. Therefore,

$$f_n(s) \approx p_l^{\frac{-1}{l-1}} \left(p_l^{\frac{1}{l-1}} s R(s)\right)^{l^n} \quad \text{for large n.}$$

Clearly, $R(0) > 0$. R is continuous on $[0, 1]$ since it is a uniform limit of a sequence of continuous functions $\{R_n\}_{n=1}^{\infty}$ on $[0, 1]$. Below we see that $R(s)$ satisfies the functional equation

$f(s)R(f(s)) = p_l(sR(s))^l$, since

$$\begin{aligned} f(s)R(f(s)) &= f(s)\prod_{j=0}^{\infty}(1+\gamma g(f_{j+1}(s)))^{\frac{1}{l^{j+1}}} \\ &= f(s)\prod_{i=1}^{\infty}(1+\gamma g(f_i(s)))^{\frac{1}{l^i}} \\ &= s^l p_l(1+\gamma g(s))\prod_{i=1}^{\infty}(1+\gamma g(f_i(s)))^{\frac{1}{l^i}} \\ &= p_l s^l \prod_{j=0}^{\infty}(1+\gamma g(f_j(s)))^{\frac{1}{l^j}}. \end{aligned}$$

On the other hand,

$$\begin{aligned} p_l(sR(s))^l &= p_l s^l[\prod_{j=0}^{\infty}(1+\gamma g(f_j(s)))^{\frac{1}{l^{j+1}}}]^l \\ &= p_l s^l \prod_{j=0}^{\infty}(1+\gamma g(f_j(s)))^{\frac{1}{l^j}}. \end{aligned}$$

We now show that solution to the functional equation is unique. Let R^1 and R^2 be two such solutions. Then they are continuous on $[0,1]$ and are both positive at 0. Define $r(s) = \frac{R^2(s)}{R^1(s)}$ for $0 \le s \le 1$. Then r is continuous on $[0,1]$ and by plugging it into the functional equation, we have $r(f(s)) = (r(s))^l$ and hence by iterating, for $0 \le s \le 1$,

$$\begin{aligned} r(s) &= (r(f(s)))^{\frac{1}{l}} \\ &= ((r(f(f(s))))^{\frac{1}{l}})^{\frac{1}{l}} \\ &= (r(f_2(s)))^{\frac{1}{l^2}} \\ &= (r(f_n(s)))^{\frac{1}{l^n}}. \end{aligned}$$

Since r is continuous at 0, $\lim_{n\to\infty} r(f_n(s)) = r(0) > 0$ and hence $r(s) = \lim_{n\to\infty} r(s) = \lim_{n\to\infty}(r(f_n(s)))^{\frac{1}{l^n}} = 1$ for $0 \le s < 1$. Since r is continuous at 1 and $r(1^-) = 1$, we have $r(1) = 1$. Therefore, $R^1(s) = R^2(s)$ for $0 \le s \le 1$. This proves the uniqueness.

□

Note that in general we allow the probability generating function $f(s)$ to be defined on $[0,\infty)$. Besides the probability generating function $f(s)$, its inverse function $g(s)$ for $0 \le s < \infty$, is quite useful in the analysis of a Galton-Watson branching process. Notice that the n-th repeated self-composition $g_n(s)$ of $g(s)$ is the inverse function of $f_n(s)$ for all n. In particular, $g_0(s) = s$ because $f_0(s) = s$. Since $g_n(s)$ is the inverse function of $f_n(s)$ for all n, the properties of $g_n(s)$

could be inferred from $f_n(s)$. Let $s_0 > 1$ be such that $f(s_0) < \infty$. The following are some of the properties of $g_n(s)$:

(1) g is strictly concave and monotone increasing.

(2) $g(p_0) = 0$ and $g(1) = 1$.

(3) If $m \leq 1$, $f(s) > s$ for $0 \leq s < 1$.

(4) If $m > 1$, then $g(s) > s$ for $q < s < 1$ and $g(s) = s$ for $s = q$ or $s = 1$ and $g(s) < s$ for $0 < s < q$ and $1 < s \leq f(s_0)$ if $f(s_0) < \infty$.

(5) $g_n(s) \uparrow 1$ as $n \to \infty$ for $q < s < 1$ and $g_n(s) \downarrow 1$ as $n \to \infty$ for $1 < s \leq f(s_0)$ if $f(s_0) < \infty$.

Assume that $p_0 = 0$ and that $f(s_0) < \infty$ for some $s_0 > 1$. Define

$$\tilde{Q}_n(s) := m^n(g_n(s) - 1) \quad \text{for } 1 \leq s \leq f(s_0). \tag{2.11}$$

The next theorem indicates that the rate of convergence of $\{g_n\}_{n=0}^{\infty}$ is geometric, which makes sense since that of $\{f_n\}_{n=0}^{\infty}$ is also geometric.

Theorem 2.4 *Assume that $p_0 = 0$ and that $f(s_0) < \infty$ for some $s_0 > 1$. Then*

$$\tilde{Q}_n(s) \downarrow \tilde{Q}(s) \quad \text{as } n \to \infty \quad \text{for } 1 \leq s \leq f(s_0),$$

where $\tilde{Q}(s)$ is the unique solution of the functional equation

$$\tilde{Q}(f(s)) = m\tilde{Q}(s) \quad \text{for } 1 \leq s \leq f(s_0)$$

with $0 < \tilde{Q}(s) < \infty$ for $1 < s \leq f(s_0)$, $\tilde{Q}(1) = 0$, and $\tilde{Q}'(1) = 1$.

Proof. Fix s in $[1, f(s_0)]$. Since $g'(1) = \frac{1}{f'(g(1))} = \frac{1}{m}$ and by the chain rule, the derivative of $\tilde{Q}_n(s)$ could be expressed as

$$\begin{aligned}
\tilde{Q}_n'(s) &= m^n g_n'(s) \\
&= m^n g'(g_{n-1}(s))g'(g_{n-2}(s))g'(g_{n-3}(s)) \cdots g'(s) \\
&= \prod_{j=0}^{n-1} mg'(g_j(s)) \qquad (2.12) \\
&= \prod_{j=0}^{n-1} \{1 + m[g'(g_j(s)) - g'(1)]\}.
\end{aligned}$$

Recall that in the proof of proposition 2.2, we mention that a result on infinite products says that if the series $\sum_{j=1}^{\infty} |a_j|$ is convergent, then the infinite products $\prod_{j=1}^{\infty}(1+a_j)$ is convergent, i.e., $\prod_{j=1}^{\infty}(1+a_j) \neq 0 < \infty$. Thus, now we want to show that $\sum_{j=0}^{\infty} |m[g'(g_j(s)) - g'(1)]| < \infty$ because this implies that $\tilde{Q}'(s) = \lim_{n\to\infty} \tilde{Q}'_n(s) = \prod_{j=0}^{\infty}\{1 + m[g'(g_j(s)) - g'(1)]\}$ is finite and non-zero.

Let $\epsilon > 0$ be such that $1 < 1 + \epsilon \leq f(s_0)$ and that $\beta = g'(1+\epsilon) < 1$. Choose j_0 to be such that $g_j(s) < 1 + \epsilon$ for all $j \geq j_0$. By mean value theorem,

$$\frac{|g'(g_{j+j_0}(s)) - g'(1)|}{|g_{j+j_0}(s) - 1|} \leq |g''(1+\epsilon)|,$$

i.e., $|g'(g_{j+j_0}(s)) - g'(1)| \leq |g''(1+\epsilon)||g_{j+j_0}(s) - 1|$.

Using mean value theorem again,

$$\begin{aligned}
\frac{|g_{j+j_0}(s) - 1|}{|g_{j+j_0-1}(s) - 1|} &= \frac{|g_{j+j_0}(s) - g_{j+j_0}(1)|}{|g_{j+j_0-1}(s) - g_{j+j_0-1}(1)|} \\
&= \frac{|g(g_{j+j_0-1}(s)) - g(g_{j+j_0-1}(1))|}{|g_{j+j_0-1}(s) - g_{j+j_0-1}(1)|} \\
&\leq |g'(1+\epsilon)| \\
&= \beta,
\end{aligned}$$

i.e., $|g_{j+j_0}(s) - 1| \leq \beta |g_{j+j_0-1}(s) - 1|$. Thus, $|g'(g_{j+j_0}(s)) - g'(1)| \leq |g''(1+\epsilon)|\beta|g_{j+j_0-1}(s) - 1|$. Similarly,

$$\begin{aligned}
|g'(g_{j+j_0}(s)) - g'(1)| &\leq |g''(1+\epsilon)|\beta^j|1 + \epsilon - 1| \\
&= \epsilon|g''(1+\epsilon)|\beta^j.
\end{aligned}$$

Thus,

$$\begin{aligned}
\sum_{j=0}^{\infty} |g'(g_{j+j_0}(s)) - g'(1)| &< \epsilon|g''(1+\epsilon)|\sum_{j=0}^{\infty}\beta^j \\
&= \epsilon|g''(1+\epsilon)|\frac{1}{1-\beta} \\
&< \infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{j=0}^{\infty}|m[g'(g_j(s))-g'(1)]| &= m\{\sum_{j=0}^{j_0-1}|g'(g_j(s))-g'(1)|+\sum_{j=j_0}^{\infty}|g'(g_j(s))-g'(1)|\}\\ &< \infty.\end{aligned}$$

This implies that

$$\lim_{n\to\infty}\tilde{Q}_n'(s)=\tilde{Q}'(s)\neq 0<\infty \quad \text{for } 1\leq s\leq f(s_0). \tag{2.13}$$

Note that

$$\tilde{Q}(s)=\int_1^s \tilde{Q}'(t)dt \quad \text{for } 1\leq s\leq f(s_0).$$

Since $\tilde{Q}_n(1)=0$,

$$\tilde{Q}_n(s)=\tilde{Q}_n(s)-\tilde{Q}_n(1)=\int_1^s \tilde{Q}_n'(t)dt.$$

Note that since $\frac{g'(g_j(s))}{g'(1)}\leq 1$ for $1\leq s\leq f(s_0)$,

$$\tilde{Q}_n'(t)=\prod_{j=0}^{n-1}\frac{g'(g_j(t))}{g'(1)}\leq 1 \quad \text{for } 1\leq t\leq f(s_0) \text{ and for all n.}$$

Hence by dominated convergence theorem,

$$\begin{aligned}\lim_{n\to\infty}\tilde{Q}_n(s) &= \lim_{n\to\infty}\int_1^s \tilde{Q}_n'(t)dt\\ &= \int_1^s \lim_{n\to\infty}\tilde{Q}_n'(t)dt\\ &= \int_1^s \tilde{Q}'(t)dt\\ &= \tilde{Q}(s)-\tilde{Q}(1)\\ &= \tilde{Q}(s).\end{aligned}$$

In fact, $\tilde{Q}_n(s)\downarrow\tilde{Q}(s)$ as $n\to\infty$ for $1\leq s\leq f(s_0)$. This can be seen from the fact that $\frac{g'(g_j(s))}{g'(1)}\leq 1$ for $1\leq s\leq f(s_0)$. Hence, $\tilde{Q}_n'(s)=\prod_{j=0}^{n-1}\frac{g'(g_j(s))}{g'(1)}$ is non-increasing in n for $1\leq s\leq f(s_0)$. Consequently, $\tilde{Q}_n(s)=\int_1^s\tilde{Q}_n'(t)dt$ is non-increasing in n for $1\leq s\leq f(s_0)$.

Since $\tilde{Q}_n(s)\downarrow\tilde{Q}(s)$ as $n\to\infty$ for $1\leq s\leq f(s_0)$ and $\tilde{Q}_n(s)<\infty$ for all n, $\tilde{Q}(s)<\infty$ for $1\leq s\leq f(s_0)$. Notice that $\tilde{Q}(s)>0$ for $1<s\leq f(s_0)$ because $\tilde{Q}(s)=\int_1^s\tilde{Q}'(t)dt$ and $\tilde{Q}'(t)>0$ for $1\leq t\leq f(s_0)$. Since $\tilde{Q}'(s)=\lim_{n\to\infty}\tilde{Q}_n'(s)=\prod_{j=0}^{\infty}\{1+m[g'(g_j(s))-g'(1)]\}$, it is clear that $\tilde{Q}'(1)=1$.

Note that we have

$$\begin{aligned}
\tilde{Q}_n(f(s)) &= m^n(g_n(f(s)) - 1) \\
&= m^n(g_{n-1}(s) - 1) \\
&= m\tilde{Q}_{n-1}(s).
\end{aligned}$$

Thus, $\lim_{n\to\infty} \tilde{Q}_n(f(s)) = \lim_{n\to\infty} m\tilde{Q}_{n-1}(s)$ and hence $\tilde{Q}(f(s)) = m\tilde{Q}(s)$. To prove uniqueness, we assume that $\tilde{Q}^1(s)$ and $\tilde{Q}^2(s)$ are two of such solutions. Then by iterating and by the triangle inequality,

$$\begin{aligned}
0 &\leq |\tilde{Q}^1(f(s)) - \tilde{Q}^2(f(s))| \\
&= m|\tilde{Q}^1(s) - \tilde{Q}^2(s)| \\
&= m^2|\tilde{Q}^1(g(s)) - \tilde{Q}^2(g(s))| \\
&= m^{n+1}|\tilde{Q}^1(g_n(s)) - \tilde{Q}^2(g_n(s))| \\
&= \frac{\tilde{Q}_{n+1}(s)}{g_{n+1}(s) - 1}|\tilde{Q}^1(g_n(s)) - \tilde{Q}^2(g_n(s))| \\
&= \tilde{Q}_{n+1}(s)|\frac{\tilde{Q}^1(g_n(s)) - \tilde{Q}^2(g_n(s))}{g_{n+1}(s) - 1}| \\
&\leq \tilde{Q}_{n+1}(s)\{|\frac{\tilde{Q}^1(g_n(s))}{g_{n+1}(s) - 1} - 1| + |1 - \frac{\tilde{Q}^2(g_n(s))}{g_{n+1}(s) - 1}|\}.
\end{aligned}$$

Since $\lim_{n\to\infty} \frac{\tilde{Q}^i(g_n(s))}{g_{n+1}(s)-1} = \lim_{s\to 1} \frac{\tilde{Q}^i(s)}{s-1} = \lim_{s\to 1}(\tilde{Q}^i)'(s) = 1$ by L'Hopital's Rule rule, we have, by taking limits on both sides of the inequality above,

$$|\tilde{Q}^1(f(s)) - \tilde{Q}^2(f(s))| = 0 \quad \text{for } 1 \leq s \leq f(s_0).$$

Hence, $\tilde{Q}^1(f(s)) = \tilde{Q}^2(f(s))$ for $1 \leq s \leq f(s_0)$. Consequently, $\tilde{Q}^1(s) = \tilde{Q}^2(s)$ for $1 \leq s \leq f(s_0)$. Therefore, $\tilde{Q}(s)$ is the unique solution of the functional equation

$$\tilde{Q}(f(s)) = m\tilde{Q}(s) \quad \text{for } 1 \leq s \leq f(s_0).$$

□

The following two theorems regarding limiting behavior of $\{Z_n\}_{n=0}^{\infty}$ are well-known and essential in our study. The first is about the ratio of successive population sizes of generations.

Theorem 2.5 *Assume that $p_0 = 0$ and $m < \infty$. Then $\lim_{n\to\infty} \frac{Z_{n+1}}{Z_n} = m$ almost surely.*

Proof. Let $\{X_i\}_{i=1}^k$ denote i.i.d. copies of Z_1 with a finite mean $m < \infty$. Then by the branching property

$$\frac{Z_{n+1}}{Z_n} = \frac{1}{Z_n} \sum_{i=1}^{Z_n} X_i.$$

Notice that since $p_0 = 0$, $Z_n \neq 0$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty} Z_n = \infty$ almost surely. By law of large numbers,

$$\lim_{n\to\infty} \frac{1}{Z_n} \sum_{i=1}^{Z_n} X_i = m \quad \text{almost surely.}$$

□

Define $W_n = \frac{Z_n}{m^n}, n = 0, 1, 2, \cdots$. Then the following theorem is a result on convergence of $\{W_n\}_{n=0}^{\infty}$.

Theorem 2.6 *Assume that $m < \infty$. Let $\{\mathcal{F}_n\}_{n=0}^{\infty}$ be the filtration where $\mathcal{F}_n = \sigma(Z_0, Z_1, \cdots, Z_n)$ for all $n \in \mathbb{N}$. Then $\{W_n\}_{n=0}^{\infty}$ is a non-negative martingale with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$. Consequently, by the martingale convergence theorem, there exists some random variable W such that $\lim_{n\to\infty} W_n = W$ almost surely.*

Proof. It is clear that by branching property

$$\begin{aligned} E(Z_{n+1}|Z_n) &= E(\sum_{i=1}^{Z_n} X_i|Z_n) \\ &= Z_n E(X_1) \\ &= mZ_n \quad \text{almost surely.} \end{aligned}$$

Since $\{Z_n\}_{n=0}^{\infty}$ is a Markov chain, so is $\{W_n\}_{n=0}^{\infty}$. Therefore, by conditioning and induction, we have

$$\begin{aligned} E(Z_{n+k}|Z_n, Z_{n-1}, \cdots, Z_0) &= E(Z_{n+k})|Z_n) \\ &= E(E(Z_{n+k}|Z_{n+k-1}, Z_{n+k-2}, \cdots, Z_n)|Z_n) \\ &= E(E(Z_{n+k}|Z_{n+k-1}|Z_n) \\ &= E(mZ_{n+k-1}|Z_n) \\ &= mE(Z_{n+k-1}|Z_n) \\ &= m^k Z_n \quad \text{for all } n, k = 0, 1, 2, \cdots. \end{aligned}$$

Dividing both sides by m^{n+k} we get

$$E(\frac{Z_{n+k}}{m^{n+k}}|Z_n, Z_{n-1}, \cdots, Z_0) = \frac{m^k Z_n}{m^{n+k}},$$

i.e., $E(W_{n+k}|Z_n, Z_{n-1}, \cdots, Z_0) = W_n$. We obtain $E(W_{n+k}|W_n, W_{n-1}, \cdots, W_0) = W_n$ since $E(W_{n+k}|Z_n, Z_{n-1}, \cdots, Z_0) = E(W_{n+k}|W_n, W_{n-1}$. Clearly, $E(|W_n|) < \infty$ for all n and W_n is $\mathcal{F}_n$-measurable for all n since Z_n is. Therefore, $\{W_n\}_{n=0}^{\infty}$ is a martingale with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$. It is obvious that $W_n \geq 0$ for all n. Since $\{W_n\}_{n=0}^{\infty}$ is a non-negative martingale, by the martingale convergence theorem, it converges to some random variable W almost surely.

□

2.4 Motivations

In the present study, we consider a randomly indexed branching process, an extension of the classical Galton-Watson branching process in which the subordinator is a Poisson process, proposed by Epps in [14]. In his work, the model with four particular mixture offspring distributions was applied to study the evolution of stock prices. The statistical investigation on various estimates and some parameters of the process were done in [13], which indicated this process is a particular case of the branching process in a random environment of i.i.d. type. Recently, this randomly indexed branching process has been brought to attention in both theoretical and applied sense. On theoretical side, [27] and [26] considered a critical branching process subordinated by a general renewal process. They investigated the probability of non-extinction, the asymptotic behavior of the moments, and also limiting distributions under normalization. In a more applied direction, [25] derived a formula for the fair price of an European call option based on modeling the underlying stock price by this process with a Poisson subordinator and with a geometric offspring distribution. Later on, a formula for the fair price of an up-and-out call option, a particular form of a barrier option, was derived in [28].

In this study, we investigate a large deviation rate concerning an interesting quantity - the ratio of successive generation sizes Z_{n+1}/Z_n. As pointed out in [1], viewed from a statistical inference pint of view, this ratio is a reasonable estimate of the expected number of children of each individual. Hence, it is meaningful to study the asymptotic behavior of its deviation from m. Athreya has intensely studied this aspect for various Galton-Watson branching processes (See [1], [5], [6], and [7]). In [1], he considered a super-critical Galton-Watson branching process with the probability of giving no birth being zero. Under some finite moment hypothesis, it was shown that the rate of decay of large deviation probabilities is geometric if the probability of having one child is positive and is super-geometric, if the probability of having one child is zero. Other than the ratio of successive generation sizes, two related quantities were studied. One is the ratio conditioned on the limiting random variable W of a sequence of normalized population sizes $\frac{Z_n}{m^n}$ being non-degenerate at zero. The other is the difference between the associated martingale sequence $\frac{Z_n}{m^n}$ and this limiting random variable W to which the sequence converges. The results showed that the rates of decay of the large deviation probabilities are super-geometric for these two quantities. In [6], Athreya and Vidyashankar generalized these results to a multi-type case. In [5], they extended these results to the critical scenario.

A natural question may be: would these large deviation results still hold for continuous-time Markov branching processes? A useful tool for studying continuous-time Markov branching processes is Kingman's theorem in [23]. All results for a Galton-Watson branching process carry over to continuous-time Markov branching processes by Kingman's theorem. Since these large deviation results were obtained in a Galton-Watson branching process, they should hold for a continuous-time Markov branching process.

Then what is the point to study these large deviation results for a randomly indexed branching process? For one thing, in a randomly indexed branching process, all individuals existing in a given generation have the same life span and simultaneously give births to their children. That is, for individuals in a given generation the numbers of their children are independent of each other while the lifetimes are not. However, in a continuous-time Markov branching process, both the lifetime and the number of children of each individual are independent. Thus, a randomly indexed branching process lives in continuous time and evolves synchronously. It is this characteristic which makes this process a bridge of a Galton-Watson branching process and a continuous-time Markov branching process and allows us to study synchronic evolution in a continuous-time set-up. What is more, there is no such handy device as Kingman's theorem carrying over the results on a Galton-Watson branching process to a randomly indexed branching process. Thus, they are not done by using Kingman's theorem. One final point is that there is no study that has ever tried to investigate a large deviation dynamic on a randomly indexed branching process. Therefore, this study contributes to the literature of randomly indexed branching processes in that the large deviation behavior is explored for the first time.

Our main target is the ratio of the successive generation sizes of a randomly indexed branching process subordinated by a Poisson process. The branching process in this study is assumed to be super-critical with either trivial or non-trivial probability of giving no birth. This setting generalizes the model considered in [1]. We mainly deal with the ratio of successive generation sizes. Under a moment condition, we examine this behavior from three perspectives. First, for the unconditional case (Theorem 4.1), the rate of decay is exponential. Second, for the case conditioned on the present generation being non-extinct (Theorem 4.2), the rate of decay is also exponential with the limit different from that of the unconditional case. Third, for the case conditioned on the next generation being non-extinct (Theorem 4.3), the rate of decay is still exponential with the limit different from the limit in Theorem 4.2. It is worth mentioning that the limit in the third case is less than or equal to that of the second case. The result of this comparison is in agreement with our intuition since conditioned on the future generation being non-extinct implies the current generation being non-extinct and hence the process survives longer. Consequently, by law of large numbers, the deviation from the mean should be smaller when conditioned on the next generation being non-extinct. Furthermore, under certain conditions weaker than the moment condition assumed in the first three theorems, the result of Theorem 4.1 is obtained in Theorem 4.4 and Corollary 4.1. In order to prove Theorem 4.6, we build a finite uniform exponential moment result for a martingale in Theorem 4.5. In Theorem 4.6, we obtain a super-exponential decay rate for the martingale used in Theorem 4.5. Under the same moment condition used in the first three theorems, Theorem 4.7 shows that the rate of decay of the ratio conditioned on the limiting random variable of a sequence of normalized population sizes being non-degenerate at zero is super-exponential no matter whether the probability of having one child is zero or not. In addition to these main theorems, we also obtain some limit results on a randomly indexed branching process. Limit theorems about the probability generating function and other related function are established in proposition 3.1 to proposition 3.3; limit results related to two martingales, $\{e^{-u'(1)t} Z_{N(t)}\}_{t \geq 0}$ and $\{Z_{N(t)} m^{-N(t)}\}_{t \geq 0}$, are established in proposition 3.4 and proposition 3.5.

We present a summary of the present study. Chapter 3 investigates some limit results on a Poisson randomly indexed branching process and provides preliminary lemmas as well as propositions needed to establish the theorems in the next chapter. Chapter 4 is devoted to a large deviation behavior of the process. The main result of this study is Theorem 4.2, which is included in chapter 4. Applications of our results to finance and physics are explored in chapter 5. Chapter 6 gives the conclusion of this study and outlines the directions of future research.

Chapter 3

Limit Theorems about Poisson Randomly Indexed Branching Processes

In this chapter, we first define a Poisson randomly indexed branching process and then prove some propositions concerning the asymptotic behavior of a Poisson randomly indexed branching process. These results are the building blocks for our theorems in the next chapter.

Let $\{Z_n\}_{n=0}^{\infty}$ and $\{N(t)\}_{t\geq 0}$ be two independent stochastic processes on the same probability space $(\Omega, \mathcal{F}, P)$ with the following characters (1) and (2).

(1) $\{Z_n\}_{n=0}^{\infty}$ is a Galton-Watson branching process with an offspring distribution $\{p_i\}_{i=0}^{\infty}$ and hence for each $n \in \mathbb{N}$, the probability generating function (p.g.f.) of Z_n is $f_n(s) = E(s^{Z_n})$ for $|s| \leq 1$, which is the n-fold iteration of $f(s) = E(s^{Z_1}) = \sum_{i=0}^{\infty} p_i s^i, |s| \leq 1$, the p.g.f. of Z_1. Throughout the first part of the study, we assume that our branching process starts from one ancestor, i.e., $Z_0 = 1$ a.s., and that it is super-critical with a finite mean, i.e., $1 < m = \sum_{i=0}^{\infty} p_i i < \infty$.

(2) $\{N(t)\}_{t\geq 0}$ is a Poisson process with intensity λ. Hence, $N(0) = 0$ and $\lim_{t\to\infty} N(t) = \infty$ a.s..

Definition 3.1 *The continuous time process $\{Z_{N(t)}\}_{t\geq 0}$ is called a Poisson randomly indexed branching process (PRIBP).*

Remark 3.1 *PRIBP is a continuous time Markov chain. Let $F_N(s,t) := E(s^{Z_{N(t)}})$ be the p.g.f. of $Z_{N(t)}$. Define $u(s) := \lambda(f(s) - s)$. Taking partial derivative of $F_N(s,t)$ with respect to s at*

$s = 1$ gives the expectation of $Z_{N(t)}$. In other words,

$$\begin{aligned}
E(Z_{N(t)}) &= \frac{\partial}{\partial s}[F_N(s,t)]\big|_{s=1} \\
&= \frac{\partial}{\partial s}[E(s^{Z_{N(t)}})]\big|_{s=1} \\
&= \frac{\partial}{\partial s}[\sum_{n=0}^{\infty} f_n(s)P(N(t)=n)]\big|_{s=1} \\
&= \sum_{n=0}^{\infty} \frac{d}{ds}[f_n(s)]\big|_{s=1} P(N(t)=n) \\
&= \sum_{n=0}^{\infty} m^n P(N(t)=n) \\
&= E(m^{N(t)}) \\
&= e^{\lambda(m-1)t} \\
&= e^{u'(1)t},
\end{aligned}$$

where the interchange of summation and differentiation is justified by the uniform convergence of the series.

Lemma 3.1 below provides a useful device for bridging some asymptotic results from ordinarily indexed stochastic processes to those for randomly indexed stochastic processes. Note that the subordinators are commonly taken to be counting processes but not necessarily. Thus, if we want to investigate the limiting behavior for a randomly indexed stochastic process, then we can study the limiting behavior for an ordinarily indexed stochastic process. Lemma 3.2 indicates that if an ordinarily indexed stochastic process is a martingale, then the corresponding randomly indexed process is also martingales.

Lemma 3.1 *Let $\{Y_n\}_{n=0}^{\infty}$ be a sequence of random variables and $\{M(t)\}_{t\geq 0}$ be a non-negative integer-valued process that is independent of $\{Y_n\}_{n=0}^{\infty}$ with $\lim_{t\to\infty} M(t) = \infty$ a.s.. Assume that $\lim_{n\to\infty} Y_n = Y$ a.s. for some random variable Y. Then $\lim_{t\to\infty} Y_{M(t)} = Y$ a.s..*

Proof. Let $A = \{\omega : \lim_{t\to\infty} M(t,\omega) = \infty\}$, $B = \{\omega : \lim_{n\to\infty} Y_n(\omega) = Y(\omega)\}$, and $C = \{\omega : \lim_{t\to\infty} Y_{M(t,\omega)}(\omega) = Y(\omega)\}$. By the assumptions, $P(A) = 1 = P(B)$. Since $A \cap B \subseteq C$, $P(C) \geq P(A \cap B) = 1$. Therefore, $P(C) = 1$, i.e., $\lim_{t\to\infty} Y_{M(t)} = Y$ a.s..

□

Lemma 3.2 *Let $\{Y_n\}_{n=0}^{\infty}$ be a martingale with respect to the filtration $\{\mathcal{F}_n\}_{n=0}^{\infty}$ where $\mathcal{F}_n = \sigma(Y_m; m \leq n)$. Let $\{M(t)\}_{t\geq 0}$ be an increasing, non-negative integer-valued process independent* *of* $\{Y_n\}_{n=0}^{\infty}$*. Let* $\mathcal{G}_t = \sigma(M(s), Y_{M(s)}; s \leq t)$*, i.e., the filtration generated by* $M(\cdot)$ *and* $Y_{M(\cdot)}$ *up to time* t*. Assume that* $E|Y_{M(t)}| < \infty$ *for all* t*. Then* $\{Y_{M(t)}\}_{t\geq 0}$ *is a martingale with respect to* $\{\mathcal{G}_t\}_{t\geq 0}$*.*

Proof. Let $A = \{M_{t_0} = m_0,\ M_{t_1} = m_1,\ \dots,\ M_{t_k} = m_k\}$ and $B = \{Y_{m_0} \in B_0,\ Y_{m_1} \in B_1,\ \dots,\ Y_{m_k} \in B_k\}$, where B_i is a Borel set for i and $t_k = t$. We need to show that $E\big(Y_{M(t+s)}|\mathcal{G}_t\big) = Y_{M(t)}$ for any $s \geq 0$. This is equivalent to show that $\int_{A\cap B} Y_{M(t+s)}dP = \int_{A\cap B} Y_{M(t)}dP$. First, let us see that

$$\begin{aligned}
\int_{A\cap B} Y_{M(t+s)}dP &= E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{B\}}Y_{M(t+s)}\big) \\
&= \sum_{j=0}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{B\}}\mathbf{1}_{\{M(t+s)=j\}}Y_j\big) \\
&= \sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{B\}}\mathbf{1}_{\{M(t+s)=j\}}Y_j\big) \\
&= \sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}}\big)E\big(\mathbf{1}_{\{B\}}Y_j\big)
\end{aligned}$$

since $\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}} \perp \mathbf{1}_{\{B\}}Y_j$. In addition,

$$\begin{aligned}
&\sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}}\big)E\big(\mathbf{1}_{\{B\}}Y_j\big) \\
=& \sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}}\big)E\big(\mathbf{1}_{\{B\}}Y_{m_k}\big) \\
=& \sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}}\mathbf{1}_{\{B\}}Y_{m_k}\big)
\end{aligned}$$

since $\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}} \perp \mathbf{1}_{\{B\}}Y_{m_k}$. Moreover,

$$\begin{aligned}
&\sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}}\mathbf{1}_{\{B\}}Y_{m_k}\big) \\
=& \sum_{j\geq m_k}^{\infty} E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{M(t+s)=j\}}\mathbf{1}_{\{B\}}Y_{M(t)}\big) \\
=& E\big(\mathbf{1}_{\{A\}}\mathbf{1}_{\{B\}}Y_{M(t)}\big) \\
=& \int_{A\cap B} Y_{M(t)}dP.
\end{aligned}$$

Therefore, $\{Y_{M(t)}\}_{t\geq 0}$ is a martingale with respect to $\{\mathcal{G}_t\}_{t\geq 0}$.

$\square$

As the key for studying asymptotic behavior of a Galton-Watson branching process is the p.g.f. $f_n(s) = E(s^{Z_n})$, the key for that of a randomly indexed branching process (RIBP) is $F_N(s,t) = E(s^{Z_{N(t)}})$. Proposition 3.1 shows that the convergence of $F_N(s,t)$ to q ,the probability of extinction, as $t \to \infty$, for $0 \leq s < 1$. Proposition 3.2 indicates an exponential rate of convergence to a power series satisfying a functional equation.

Proposition 3.1 $\lim_{t\to\infty} F_N(s,t) = q$ *for all* $0 \leq s < 1$.

Remark 3.2 *In particular, if* $p_0 = 0$*, then* $q = 0$*, and hence* $\lim_{t\to\infty} F_N(s,t) = 0$ *for all* $0 \leq s < 1$.

Proof. First, let us note that

$$\begin{aligned}
F_N(s,t) &= E(s^{Z_{N(t)}}) \\
&= E\big[E[s^{Z_{N(t)}}|N(t)]\big] \\
&= \sum_{n=0}^{\infty} E[s^{Z_{N(t)}}|N(t)=n]P(N(t)=n) \\
&= \sum_{n=0}^{\infty} E(s^{Z_n})P(N(t)=n) \\
&= \sum_{n=0}^{\infty} f_n(s)P(N(t)=n) \\
&= E(f_{N(t)}(s)).
\end{aligned}$$

Since $\lim_{n\to\infty} f_n(s) = q$ for all $0 \leq s < 1$, it is clear that $\lim_{t\to\infty} f_{N(t)}(s) = q$ a.s. for all $0 \leq s < 1$. Since $|f_{N(t)}(s)| \leq 1$ for all $0 \leq s < 1$ for all t and $\lim_{t\to\infty} N(t) = \infty$ a.s., by dominated convergence theorem, we have

$$\begin{aligned}
\lim_{t\to\infty} F_N(s,t) &= \lim_{t\to\infty} E(f_{N(t)}(s)) \\
&= E(q) \\
&= q.
\end{aligned}$$

$\square$

Proposition 3.2 *Assume that* $m \neq 1$*. If* $p_0 \neq 0$*, then*

$$\lim_{t\to\infty} e^{-u'(q)t}[F_N(s,t) - q] = Q(s) = \sum_{k=0}^{\infty} q_k s^k < \infty \quad \text{for all } 0 \leq s < 1.$$

Moreover, $Q(s)$ is the unique solution of the functional equation,

$$Q(f(s)) = f'(q)Q(s) \quad \textit{for all } 0 \le s < 1.$$

Remark 3.3 *In particular, if $p_0 = 0$ and $p_1 \neq 0$, then $\lim_{t\to\infty} e^{-\lambda(p_1-1)t}F_N(s,t) = \hat{Q}(s) = \sum_{k=1}^{\infty} \hat{q}_k s^k < \infty$ for all $0 \le s < 1$. Moreover, $\hat{Q}(s)$ is the unique solution of $\hat{Q}(f(s)) = p_1\hat{Q}(s)$ for all $0 \le s < 1$.*

Proof. Define $A(s,t) := e^{-u'(q)t}[F_N(s,t) - q] = e^{-u'(q)t}[E(f_{N(t)}(s)) - q]$. Then

$$\begin{aligned}
\frac{\partial}{\partial s}A(s,t) &= \frac{\partial}{\partial s}\{e^{-u'(q)t}[E(f_{N(t)}(s)) - q]\} \\
&= e^{-u'(q)t}\frac{\partial}{\partial s}\{[\sum_{n=0}^{\infty} f_n(s)P(N(t)=n)] - q\} \\
&= e^{-u'(q)t}\frac{\partial}{\partial s}[\sum_{n=0}^{\infty} f_n(s)P(N(t)=n)] \\
&= e^{-u'(q)t}\sum_{n=0}^{\infty}\frac{\partial}{\partial s}[f_n(s)P(N(t)=n)],
\end{aligned}$$

where the interchange of derivative and summation is justified by the uniform convergence of the series as the following argument shows.

We want to show that $\lim_{k\to\infty}\sup_{s\in[0,1)}|\sum_{n=0}^{\infty} f_n(s)P(N(t)=n) - \sum_{n=0}^{k} f_n(s)P(N(t)=n)| = 0$. Note that $0 \le \sup_{s\in[0,1)}|\sum_{n=0}^{\infty} f_n(s)P(N(t)=n) - \sum_{n=0}^{k} f_n(s)P(N(t)=n)| = \sup_{s\in[0,1)}\sum_{n=k+1}^{\infty} f_n(s)P(N(t)=n) \le \sum_{n=k+1}^{\infty} P(N(t)=n)$ and $\lim_{k\to\infty}\sum_{n=k+1}^{\infty} P(N(t)=n) = 0$ since it is the tail of the convergent series $\sum_{n=0}^{\infty} P(N(t)=n) = 1$.

Now let us see that

$$\begin{aligned}
e^{-u'(q)t}\sum_{n=0}^{\infty}\frac{\partial}{\partial s}[f_n(s)P(N(t)=n)] &= e^{-u'(q)t}\sum_{n=0}^{\infty}P(N(t)=n)\prod_{j=0}^{n-1}f'(f_j(s)) \\
&= e^{-\lambda[f'(q)-1]t}\sum_{n=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{n!}\prod_{j=0}^{n-1}f'(f_j(s)) \\
&= \sum_{n=0}^{\infty}\frac{e^{-\lambda f'(q)t}(\lambda f'(q)t)^n}{n!}\cdot\frac{1}{(f'(q))^n}\prod_{j=0}^{n-1}f'(f_j(s)) \\
&= \sum_{n=0}^{\infty}P(N(f'(q)t)=n)\prod_{j=0}^{n-1}\frac{f'(f_j(s))}{f'(q)},
\end{aligned}$$

where $N(f'(q)t))$ is a Poisson distributed random variable with parameter $\lambda f'(q)t$. Recall the definition of $Q_n(s)$ in (2.4). Define $V_n(s) := Q_n(s) + C$ for some constant $C \in \mathbb{R}$ for $s \in [0, 1)$. Then $V_n'(s) = Q_n'(s)$ which is expressed in (2.6). We note that

$$\begin{aligned} \sum_{n=0}^{\infty} P(N(f'(q)t) = n) \prod_{j=0}^{n-1} \frac{f'(f_j(s))}{f'(q)} &= \sum_{n=0}^{\infty} P(N(f'(q)t) = n) V_n'(s) \\ &= E[V'_{N(f'(q)t))}(s)]. \end{aligned}$$

This yields $\frac{\partial}{\partial s} A(s,t) = E[V'_{N(f'(q)t))}(s)]$ for $0 \leq s < 1$. Recall that we have the result (2.7). Hence this gives $\lim_{t\to\infty} Q'_{N(f'(q)t))}(s) = Q'(s)$ a.s. for $0 \leq s < 1$. Therefore, for $0 \leq s < 1$,

$$\begin{aligned} \lim_{t\to\infty} \frac{\partial}{\partial s} A(s,t) &= \lim_{t\to\infty} E[V'_{N(f'(q)t))}(s)] \\ &= E[\lim_{t\to\infty} V'_{N(f'(q)t))}(s)] \\ &= E[V'(s)] \\ &= V'(s). \end{aligned}$$

Note that the interchange of the limit and the expectation is legitimate due to dominated convergence theorem. Since f' is increasing in s, $f'(f_j(s)) \leq f'(q)$ for $0 \leq s \leq q$ and $f'(f_j(s)) > f'(q)$ for $q < s < 1$. Consequently, $V_n'(s) = Q_n'(s) \leq 1$ for $0 \leq s \leq q$ and $V_n'(s) = Q_n'(s) < Q'(s)$ for $q < s < 1$. Let

$$\psi(s) = \begin{cases} 1 & \text{if } 0 \leq s \leq q, \\ Q'(s) & \text{if } q < s < 1. \end{cases}$$

Then $V'_{N(f'(q)t))}(s) \leq \psi(s)$ a.s. for all $t \geq 0$. Also note that $E[\psi(s)] = E(1) = 1 < \infty$ for $0 \leq s \leq q$ and $E[\psi(s)] = E[Q'(s)] = Q'(s) < \infty$ for $q < s < 1$ and thus $E[\psi(s)] < \infty$ for all $0 \leq s < 1$.

Since $\lim_{n\to\infty} Q_n(s) = Q(s) = \sum_{k=0}^{\infty} q_k s^k \not\equiv 0$ for $0 \leq s < 1$, $Q(q) = 0$. Notice that $\lim_{n\to\infty} V_n(s) = \lim_{n\to\infty}[Q_n(s) + C] = Q(s) + C = V(s)$. Since $A(s,t) = A(s,t) - A(q,t) = \int_q^s \frac{\partial}{\partial v} A(v,t)\, dv$, by an argument of dominated convergence similar to the above argument, we can see

$$
\begin{aligned}
\lim_{t\to\infty} A(s,t) &= \lim_{t\to\infty} \int_q^s \frac{\partial}{\partial v} A(v,t)\,dv \\
&= \int_q^s \lim_{t\to\infty} \frac{\partial}{\partial v} A(v,t)\,dv \\
&= \int_q^s V'(v)\,dv \\
&= V(s) - V(q) \\
&= [Q(s)+C] - [Q(q)+C] \\
&= Q(s).
\end{aligned}
$$

Therefore,

$$
\lim_{t\to\infty} e^{-u'(q)t}[F_N(s,t)-q] = Q(s) = \sum_{k=0}^{\infty} q_k s^k < \infty \quad \text{for} \quad 0 \le s < 1.
$$

The proof of $Q(s)$ being the unique solution of $Q(f(s)) = f'(q)Q(s)$ for $0 \le s < 1$ is proved in Theorem 2.1.

Notice that when $p_0 = 0$, we need the condition that $p_1 \neq 0$. Since if $p_0 = 0$, then $q = 0$, and hence $f'(q) = f'(0) = p_1$. Thus, if $p_0 = 0$ and $p_1 = 0$, then $f'(q) = f'(0) = p_1 = 0$, and hence Q_n and Q'_n are undefined. In this case, the proof follows the similar lines as that in the case of $p_0 \neq 0$.

□

Let $g(s)$ be the inverse function of $f(s)$. Thus, the n-fold iteration $g_n(s)$ of $g(s)$ is $[f_n(s)]^{-1}$. Proposition 3.3 shows that $E(g_{N(t)}(s))$ converges at an exponential rate as $t \to \infty$.

Proposition 3.3 *Assume that $f(s_1) < \infty$ for some $s_1 > 1$. Let*

$$
\tilde{Q}_{N(t)}(s) = e^{\lambda(1-g'(1))t}[E(g_{N(t)}(s)) - 1].
$$

Then for $1 \le s \le f(s_1)$,

$$
E(g_{N(t)}(s)) \downarrow 1 \quad \textit{as } t \to \infty.
$$

Moreover,

$$
\lim_{t\to\infty} \tilde{Q}_{N(t)}(s) = \tilde{Q}(s),
$$

where $\tilde{Q}(s)$ is the unique solution of the functional equation,

$$\tilde{Q}(f(s)) = m\tilde{Q}(s) \quad \textit{for } 1 \leq s \leq f(s_1)$$

with $0 < \tilde{Q}(s) < \infty$ for $1 < s \leq f(s_1)$ with $\tilde{Q}(1) = 0$ and $\tilde{Q}'(1) = 1$.

Proof. Since $g_n(s) \downarrow 1$ as $n \to \infty$ for $1 \leq s \leq f(s_1)$, $g_{N(t)}(s) \downarrow 1$ as $t \to \infty$ a.s. for $1 \leq s \leq f(s_1)$. Hence, $E\big[g_{N(t)}(s)\big] \downarrow 1$ as $t \to \infty$ for $1 \leq s \leq f(s_1)$ by dominated convergence theorem, because $|g_{N(t)}(s)| \leq s$ for $1 \leq s \leq f(s_1)$ for all $t \geq 0$ and $E(s) = s < \infty$ and $g_{N(t)}(s) \downarrow 1$ as $t \to \infty$ a.s. for $1 \leq s \leq f(s_1)$.

Define $B(s,t) := e^{\lambda(1-g'(1))t}[E(g_{N(t)}(s)) - 1] = e^{\lambda(1-\frac{1}{m})t}[E(g_{N(t)}(s)) - 1]$. Then it follows that

$$\begin{aligned}
\frac{\partial}{\partial s}B(s,t) &= e^{\lambda(1-\frac{1}{m})t}\frac{\partial}{\partial s}\{[\sum_{n=0}^{\infty} g_n(s)P(N(t)=n)] - 1\} \\
&= e^{\lambda(1-\frac{1}{m})t}\frac{\partial}{\partial s}[\sum_{n=0}^{\infty} g_n(s)P(N(t)=n)] \\
&= e^{\lambda(1-\frac{1}{m})t}\sum_{n=0}^{\infty}\frac{\partial}{\partial s}[g_n(s)P(N(t)=n)],
\end{aligned}$$

where the interchange of derivative and summation is justified by the uniform convergence of the series as the following argument shows.

We want to show that $\lim_{k\to\infty}\sup_{s\in[0,1)}|\sum_{n=0}^{\infty} f_n(s)P(N(t) = n) - \sum_{n=0}^{k} f_n(s)P(N(t) = n)| = 0$. Note that $0 \leq \sup_{s\in[1,f(s_1)]}|\sum_{n=0}^{\infty} g_n(s)P(N(t) = n) - \sum_{n=0}^{k} g_n(s)P(N(t) = n)| = \sup_{s\in[0,1)}\sum_{n=k+1}^{\infty} g_n(s)P(N(t)=n) = \sum_{n=k+1}^{\infty} g_n(f(s_1))P(N(t)=n)$ since g_n is an increasing function of s. It is clear that $\sum_{n=k+1}^{\infty} g_n(f(s_1))P(N(t)=n) < \sum_{n=k+1}^{\infty} g_{k+1}(f(s_1))P(N(t)=n)$ since $g_n(s)$ is decreasing in n for $s > 1$. Hence

$$\begin{aligned}
& \lim_{k\to\infty}\sup_{s\in[0,1)}|\sum_{n=0}^{\infty} f_n(s)P(N(t)=n) - \sum_{n=0}^{k} f_n(s)P(N(t)=n)| \\
\leq\ & \lim_{k\to\infty}\sum_{n=k+1}^{\infty} g_{k+1}(f(s_1))P(N(t)=n) \\
=\ & \lim_{k\to\infty} g_{k+1}(f(s_1))\lim_{k\to\infty}\sum_{n=k+1}^{\infty} P(N(t)=n) \\
=\ & 0,
\end{aligned}$$

because $\lim_{k\to\infty} g_{k+1}(f(s_1)) = 1$ and $\lim_{k\to\infty} \sum_{n=k+1}^{\infty} P(N(t) = n) = 0$. Therefore, we have the uniform convergence. Now we see that

$$\begin{aligned}
e^{\lambda(1-\frac{1}{m})t} \sum_{n=0}^{\infty} \frac{\partial}{\partial s}[g_n(s)P(N(t) = n)] &= e^{\lambda(1-\frac{1}{m})t} \sum_{n=0}^{\infty} P(N(t) = n) \prod_{j=0}^{n-1} g'(g_j(s)) \\
&= e^{\lambda(1-\frac{1}{m})t} \sum_{n=0}^{\infty} \frac{e^{-\lambda t}(\lambda t)^n}{n!} \prod_{j=0}^{n-1} g'(g_j(s)) \\
&= \sum_{n=0}^{\infty} \frac{e^{-\frac{1}{m}\lambda t}(\frac{1}{m}\lambda t)^n}{n!} m^n \prod_{j=0}^{n-1} g'(g_j(s)) \\
&= \sum_{n=0}^{\infty} P(N(\frac{1}{m}t) = n) \prod_{j=0}^{n-1} mg'(g_j(s)),
\end{aligned}$$

where $N(\frac{1}{m}t)$ is a Poisson random variable with parameter $\frac{1}{m}\lambda t$. Recall the definition of $\tilde{Q}_n(s)$ in (2.11). Define $\tilde{V}_n(s) := \tilde{Q}_n(s) + C$ for some constant $C \in \mathbb{R}$ for $s \in [1, f(s_1)]$. Then $\tilde{V}_n'(s) = \tilde{Q}_n'(s)$ which is expressed in (2.12). We note that

$$\begin{aligned}
\sum_{n=0}^{\infty} P(N(\frac{1}{m}t) = n) \prod_{j=0}^{n-1} mg'(g_j(s)) &= \sum_{n=0}^{\infty} P(N(\frac{1}{m}t) = n)\tilde{V}'_n(s) \\
&= E[\tilde{V}'_{N(\frac{1}{m}t)}(s)].
\end{aligned}$$

This yields $\frac{\partial}{\partial s}B(s,t) = E[\tilde{V}'_{N(\frac{1}{m}t)}(s)]$ for $1 \geq s \leq f(s_1)$. Recall that we have the result (2.13). Hence this yields $\lim_{t\to\infty} \tilde{Q}'_{N(\frac{1}{m}t)}(s) = \tilde{Q}'(s)$ a.s. for $1 \leq s \leq f(s_1)$. Therefore, for $1 \leq s \leq f(s_1)$,

$$\begin{aligned}
\lim_{t\to\infty} \frac{\partial}{\partial s} B(s,t) &= \lim_{t\to\infty} E[\tilde{V}'_{N(\frac{1}{m}t)}(s)] \\
&= E[\lim_{t\to\infty} \tilde{V}'_{N(\frac{1}{m}t)}(s)] \\
&= E[\tilde{V}'(s)] \\
&= \tilde{V}'(s).
\end{aligned}$$

Note that the interchange of the limit and the expectation is justified by dominated convergence theorem. Since $g_j(s) \geq 1$ for $s \geq 1$ and g_j increasing in s and g' is decreasing in s, $g'(g_j(s)) \leq g'(1) = \frac{1}{m}$ and hence $\prod_{j=0}^{n-1} mg'(g_j(s)) \leq \prod_{j=0}^{n-1} m\frac{1}{m} = 1$. Also note that $E(1) = 1 < \infty$. Therefore, $|\tilde{V}'_{N(\frac{1}{m}t)}(s)| \leq 1$ a.s. for $s \geq 1$ for all $t \geq 0$.

From Theorem 2.4, we know that $\tilde{Q}(1) = 0$, $\tilde{Q}'(1) = 1$, and $\tilde{Q}(s)$ is the unique solution of the functional equation $\tilde{Q}(f(s)) = m\tilde{Q}(s)$ for $1 \le s \le f(s_1)$ with $0 < \tilde{Q}(s) < \infty$ for $1 < s \le f(s_1)$.

Notice that $\lim_{n\to\infty} \tilde{V}_n(s) = \lim_{n\to\infty}[\tilde{Q}_n(s) + C] = \tilde{Q}(s) + C = \tilde{V}(s)$. Since $B(s,t) = B(s,t) - B(1,t) = \int_1^s \frac{\partial}{\partial v} B(v,t)\, dv$, by an argument of dominated convergence similar to the above argument, we get

$$\begin{aligned}
\lim_{t\to\infty} B(s,t) &= \lim_{t\to\infty} \int_1^s \frac{\partial}{\partial v} B(v,t)\, dv \\
&= \int_1^s \lim_{t\to\infty} \frac{\partial}{\partial v} B(v,t)\, dv \\
&= \int_1^s \tilde{V}'(v)\, dv \\
&= \tilde{V}(s) - \tilde{V}(1) \\
&= [\tilde{Q}(s) + C] - [\tilde{Q}(1) + C] \\
&= \tilde{Q}(s).
\end{aligned}$$

Therefore,

$$\lim_{t\to\infty} e^{\lambda(1-g'(1))t}[E(g_{N(t)}(s)) - 1] = \tilde{Q}(s) \quad \text{for} \quad 1 \le s \le f(s_1).$$

$\square$

Let $W_n := \frac{Z_n}{m^n}$. Recall that $\{W_n\}_{n=0}^{\infty}$ is a non-negative martingale and hence converges a.s. to a random variable W. When we deal with the RIBP, we can consider tow analogues results. First of all, let $W_{N(t)} := e^{-u'(1)t} Z_{N(t)}$. It is worth mentioning that this quantity is essential when applying RIBP to option pricing because it enables us to identify the equivalent martingale measure. Proposition 3.4 shows that it is a martingale and a proof can be found in [28]. On the other hand, let $W^*_{N(t)} := Z_{N(t)} m^{-N(t)}$. Proposition 3.5 indicates that it is a martingale and a proof can be found in [36]. We provide these two martingale results for completeness.

Proposition 3.4 *Let $\mathcal{G}_t = \sigma(N(s), W_{N(s)}; s \le t)$. Then $\{W_{N(t)}\}_{t\ge 0}$ where $W_{N(t)} := e^{-u'(1)t} Z_{N(t)}$ is a non-negative martingale with respect to the filtration $\{\mathcal{G}_t\}_{t\ge 0}$ and*

$$\lim_{t\to\infty} W_{N(t)} = W' \quad a.s..$$

Proof. Let $t \ge 0$ and $v \ge 0$. We need to show that

$$E\big(e^{-u'(1)(t+v)} Z_{N(t+v)} \,|\, e^{-u'(1)t} Z_{N(t)}\big) = e^{-u'(1)t} Z_{N(t)}.$$

This is equivalent to show that

$$E\big(Z_{N(t+v)} \,|\, Z_{N(t)}\big) = e^{u'(1)v} Z_{N(t)}$$

because we know that

$$E\big(e^{-u'(1)(t+v)}Z_{N(t+v)}\,|\,e^{-u'(1)t}Z_{N(t)}\big) = E\big(e^{-u'(1)(t+v)}Z_{N(t+vs)}\,|\,Z_{N(t)}\big)$$

and because

$$E\big(Z_{N(t+v)}\,|\,Z_{N(t)}\big) = e^{u'(1)v}Z_{N(t)}$$

is equivalent to

$$E\big(e^{-u'(1)(t+v)}Z_{N(t+v)}\,|\,Z_{N(t)}\big) = e^{-u'(1)t}Z_{N(t)}.$$

Recall that for a Poisson process $\{N(t)\}_{t\geq 0}$, $N(t+v)-N(t)$ and $N(v)$ have the same distribution. Let $Z^{(j)}_{N(t)-N(v)}$ denote the number of offspring of the j-th of the $Z_{N(t)}$ particles existing at time t which are alive at time $t+v$. We can see that

$$\begin{aligned}
E\big(Z_{N(t+v)}|Z_{N(t)}\big) &= E\Big(\sum_{j=1}^{Z_{N(t)}} Z^{(j)}_{N(t+v)-N(t)}\,|\,Z_{N(t)}\Big) \\
&= E\Big(\sum_{j=1}^{Z_{N(t)}} Z^{(j)}_{N(v)}\,|\,Z_{N(t)}\Big) \\
&= \sum_{j=1}^{Z_{N(t)}} E\big(Z^{(j)}_{N(v)}\big) \\
&= Z_{N(t)}E(Z_{N(v)}) \\
&= Z_{N(t)}e^{u'(1)v}.
\end{aligned}$$

Therefore, $\{W_{N(t)}\}_{t\geq 0}$ is a martingale with respect to $\{\mathcal{G}_t\}_{t\geq 0}$. Clearly, $\{W_{N(t)}\}_{t\geq 0}$ is non-negative for all t. Since $\{W_{N(t)}\}_{t\geq 0}$ is a non-negative martingale, $\lim_{t\to\infty} W_{N(t)} = W'$ a.s. exists by martingale convergence theorem.

□

Proposition 3.5 *Let $W^*_{N(t)} := Z_{N(t)}m^{-N(t)}$. Then $\{W^*_{N(t)}\}_{t\geq 0}$ is a non-negative martingale with respect to $\{\mathcal{G}_t\}_{t\geq 0}$ and*

$$\lim_{t\to\infty} W^*_{N(t)} = W^* \quad \textit{a.s. for some random variable } W^*.$$

Furthermore, W^* has the same distribution as W.

Proof. Fix $t \geq 0$. It is clear that

$$\begin{aligned} E\big(W^*_{N(t)}\big) &= E\big(Z_{N(t)} m^{-N(t)}\big) \\ &= E\Big[E\big(Z_{N(t)} m^{-N(t)} \big| N(t)\big)\Big] \\ &= E\Big[m^{-N(t)} E\big(Z_{N(t)} \big| N(t)\big)\Big] \\ &= E\Big(m^{-N(t)} m^{N(t)}\Big) \\ &= 1 < \infty. \end{aligned}$$

Since $E\big|W^*_{N(t)}\big| < \infty$ for all $t \geq 0$ and $\{W^*_n\}_{n=0}^{\infty}$ is a martingale with respect to $\{\mathcal{F}_n\}_{n=0}^{\infty}$, $\{W^*_{N(t)}\}_{t\geq 0}$ is a martingale with respect to $\{\mathcal{G}_t\}_{t\geq 0}$ by Lemma 3.2. Clearly, $W^*_{N(t)}$ is non-negative for all $t \geq 0$. Since $\{W^*_{N(t)}\}_{t\geq 0}$ is a non-negative martingale, $\lim_{t\to\infty} W^*_{N(t)} = W^*$ a.s. and the random variable W^* exists by martingale convergence theorem. Since $\lim_{n\to\infty} W_n = W$ a.s., $\lim_{t\to\infty} W^*_{N(t)} = W$ a.s. by lemma 3.1. Therefore, W^* and W have the same distribution.

$\square$

Chapter 4

Large Deviation Results on Poisson Randomly Indexed Branching Processes

We mainly consider the large deviation behavior for the ratio of successive generation sizes $Z_{N(t)+1}/Z_{N(t)}$ deviating from the expected number of offspring m. Under a finite exponential moment condition on the offspring distribution, when $p_0 = 0$ and $p_1 > 0$, Theorem 4.1 shows that the rate of decay is exponential.

Theorem 4.1 *Assume that $p_0 = 0$ and $p_1 > 0$. Assume that $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$. Then for any $\varepsilon > 0$,*

$$\lim_{t\to\infty} e^{-\lambda(p_1-1)t} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon\big) = \sum_{k=1}^{\infty} \phi(k,\varepsilon)\hat{q}_k < \infty,$$

$$\textit{where} \quad \phi(k,\varepsilon) = P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon\big) \textit{ and } \{X_i\}_{i=1}^{k} \textit{ are i.i.d. copies of } Z_1.$$

Proof. Let us see that

$$\begin{aligned}\phi(k,\varepsilon) &\leq P\big(\sum_{i=1}^{k} X_i \geq k(m+\varepsilon)\big) + P\big(\sum_{i=1}^{k} X_i \leq k(m-\varepsilon)\big)\\ &= P\big(\gamma^{\sum_{i=1}^{k} X_i} \geq \gamma^{k(m+\varepsilon)}\big) + P\big(\beta^{\sum_{i=1}^{k} X_i} \geq \beta^{k(m-\varepsilon)}\big),\end{aligned}$$

for some $1 < \gamma < e^{\alpha_0}$ and $0 < \beta < 1$. In addition,

$$\begin{aligned} & P\big(\gamma^{\sum_{i=1}^k X_i} \geq \gamma^{k(m+\varepsilon)}\big) + P\big(\beta^{\sum_{i=1}^k X_i} \geq \beta^{k(m-\varepsilon)}\big) \\ \leq \;& E\big[\gamma^{\sum_{i=1}^k X_i}\big]\gamma^{-k(m+\varepsilon)} + E\big[\beta^{\sum_{i=1}^k X_i}\big]\beta^{-k(m-\varepsilon)}, \end{aligned}$$

by Markov's inequality. Moreover,

$$\begin{aligned} & E\big[\gamma^{\sum_{i=1}^k X_i}\big]\gamma^{-k(m+\varepsilon)} + E\big[\beta^{\sum_{i=1}^k X_i}\big]\beta^{-k(m-\varepsilon)} \\ = \;& \big[E(\gamma^{X_i})\big]^k\gamma^{-k(m+\varepsilon)} + \big[E(\beta^{X_i})\big]^k\beta^{-k(m-\varepsilon)} \end{aligned}$$

because of $\{X_i\}_{i=1}^k$ are i.i.d.. Furthermore,

$$\begin{aligned} & \big[E(\gamma^{X_i})\big]^k\gamma^{-k(m+\varepsilon)} + \big[E(\beta^{X_i})\big]^k\beta^{-k(m-\varepsilon)} \\ = \;& [f(\gamma)\gamma^{-(m+\varepsilon)}]^k + [f(\beta)\beta^{-(m-\varepsilon)}]^k \\ \leq \;& 2\big[max\{f(\gamma_\varepsilon)\gamma_\varepsilon^{-(m+\varepsilon)}, f(\beta_\varepsilon)\beta_\varepsilon^{-(m-\varepsilon)}\}\big]^k. \end{aligned}$$

Notice that for each $\varepsilon > 0$, there exists $1 < \gamma_\varepsilon < e^{\alpha_0}$ and $0 < \beta_\varepsilon < 1$ such that $0 < f(\gamma_\varepsilon)\gamma_\varepsilon^{-(m+\varepsilon)} < 1$ and $0 < f(\beta_\varepsilon)\beta_\varepsilon^{-(m-\varepsilon)} < 1$. Define $\delta_\varepsilon := max\{f(\gamma_\varepsilon)\gamma_\varepsilon^{-(m+\varepsilon)}, f(\beta_\varepsilon)\beta_\varepsilon^{-(m-\varepsilon)}\}$, then $\phi(k,\varepsilon) \leq 2(\delta_\varepsilon)^k$. It is obvious that $0 < \delta_\varepsilon < 1$. This can been seen from the Taylor expansion of f about 1. Let us identify that

$$\begin{aligned} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon\big) &= \sum_{k=1}^\infty P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, Z_{N(t)} = k\big) \\ &= \sum_{k=1}^\infty P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k\big)P\big(Z_{N(t)} = k\big) \\ &= \sum_{k=1}^\infty P\big(\big|\frac{1}{k}\sum_{i=1}^k X_i - m\big| > \varepsilon\big)P\big(Z_{N(t)} = k\big) \\ &= \sum_{k=1}^\infty \phi(k,\varepsilon)P\big(Z_{N(t)} = k\big). \end{aligned}$$

We need the following generalization of Lebesgue's dominated convergence theorem **(***)** (See p.270 in [30]):

Let $\{\mu_n\}$ be a sequence of measures that converges to a measure μ on a measurable space. Let $\{v_n\}$ and $\{u_n\}$ be two sequences of measurable functions that converge pointwise to v and u, respectively. If $|v_n| \leq u_n$ for all $n \in \mathbb{N}$ and $\lim_{n\to\infty}\int u_n d\mu_n = \int u d\mu < \infty$, then $\lim_{n\to\infty}\int v_n d\mu_n = \int v d\mu < \infty$.

In order to apply **(***)**, let us define $h_t(k) := e^{-\lambda(p_1-1)t}\phi(k,\varepsilon)P\big(Z_{N(t)} = k\big)$ and $r_t(k) := 2(\delta_\varepsilon)^k e^{-\lambda(p_1-1)t}P\big(Z_{N(t)} = k\big)$. Then for each $k \geq 1, h_t(k) \leq r_t(k)$ for all $t \geq 0$. By remark 5.3, it follows that

$$\lim_{t\to\infty} h_t(k) = \phi(k,\varepsilon) \lim_{t\to\infty} e^{-\lambda(p_1-1)t}P\big(Z_{N(t)} = k\big) = \phi(k,\varepsilon)\hat{q}_k \quad \text{and}$$

$$\lim_{t\to\infty} r_t(k) = 2(\delta_\varepsilon)^k \lim_{t\to\infty} e^{-\lambda(p_1-1)t}P\big(Z_{N(t)} = k\big) = 2(\delta_\varepsilon)^k\hat{q}_k.$$

Since

$$\begin{aligned}\sum_{k=1}^{\infty} r_t(k) &= 2e^{-\lambda(p_1-1)t}\sum_{k=1}^{\infty}(\delta_\varepsilon)^k P\big(Z_{N(t)} = k\big) \\ &= 2e^{-\lambda(p_1-1)t}F_N(\delta_\varepsilon, t),\end{aligned}$$

it follows, by remark 3.3, that

$$\lim_{t\to\infty}\sum_{k=1}^{\infty} r_t(k) = 2\hat{Q}(\delta_\varepsilon) < \infty.$$

Therefore, we have

$$\begin{aligned}\sum_{k=1}^{\infty}\lim_{t\to\infty} r_t(k) &= 2\sum_{k=1}^{\infty}(\delta_\varepsilon)^k\hat{q}_k \\ &= 2\hat{Q}(\delta_\varepsilon) \\ &= \lim_{t\to\infty}\sum_{k=1}^{\infty} r_t(k) < \infty.\end{aligned}$$

Therefore, by **(***)**,

$$\lim_{t\to\infty}\sum_{k=1}^{\infty} h_t(k) = \sum_{k=1}^{\infty}\lim_{t\to\infty} h_t(k) = \sum_{k=1}^{\infty}\phi(k,\varepsilon)\hat{q}_k < \infty.$$

It is clear that

$$\begin{aligned}\lim_{t\to\infty}\sum_{k=1}^{\infty} h_t(k) &= \lim_{t\to\infty}\sum_{k=1}^{\infty} e^{-\lambda(p_1-1)t}\phi(k,\varepsilon)P\big(Z_{N(t)} = k\big) \\ &= \lim_{t\to\infty} e^{-\lambda(p_1-1)t}P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\Big| > \varepsilon\Big).\end{aligned}$$

Therefore, we reach

$$\lim_{t\to\infty} e^{-\lambda(p_1-1)t} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon\big) = \sum_{k=1}^{\infty} \phi(k,\varepsilon)\hat{q}_k < \infty.$$

$\square$

Under the same finite exponential moment condition on the offspring distribution, when $p_0 \neq 0$ and conditional on $Z_{N(t)} > 0$, Theorem 4.2 shows that the rate of decay is also exponential.

Theorem 4.2 *Assume that $p_0 \neq 0$. Assume that $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$. Then for any $\varepsilon > 0$,*

$$\lim_{t\to\infty} e^{-u'(q)t} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} > 0\big) = \frac{\sum_{k=1}^{\infty} \varphi(k,\varepsilon) q_k}{1-q} < \infty,$$

where $\varphi(k,\varepsilon) = P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon\big)$.

Remark 4.1 *Since X_i's in Theorem 4.2 have different offspring distributions as they have in Theorem 4.1, we use the notations $\varphi(k,\varepsilon)$ and $\phi(k,\varepsilon)$, respectively.*

Proof. By the same estimate in Theorem 4.1, $\varphi(k,\varepsilon) \leq 2(\delta_\varepsilon)^k$. It is clear that

$$\begin{aligned}
& P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} > 0\big) \\
= & \sum_{k=0}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k \,\big|\, Z_{N(t)} > 0\big) \\
= & \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k\big)}{P(Z_{N(t)} > 0)}
\end{aligned}$$

since $\{Z_{N(t)} = k, Z_{N(t)} > 0\,;\, k \geq 0\} = \{Z_{N(t)} = k\,;\, k \geq 1\}$. It is obvious that

$$\begin{aligned}
& \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k\big)}{P(Z_{N(t)} > 0)} \\
= & \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k\big) P(Z_{N(t)} = k)}{P(Z_{N(t)} > 0)} \\
= & \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon\big) P(Z_{N(t)} = k)}{P(Z_{N(t)} > 0)} \\
= & \sum_{k=1}^{\infty} \frac{\varphi(k,\varepsilon) P(Z_{N(t)} = k)}{1 - P(Z_{N(t)} = 0)}.
\end{aligned}$$

In order to apply **(***)**, let us define $h_t(k) := \frac{e^{-u'(q)t}\varphi(k,\varepsilon)P(Z_{N(t)}=k)}{1-P(Z_{N(t)}=0)}$ and $r_t(k) := \frac{e^{-u'(q)t}2(\delta_\varepsilon)^k P(Z_{N(t)}=k)}{1-P(Z_{N(t)}=0)}$. Then for each $k \geq 1, h_t(k) \leq r_t(k)$ for all $t \geq 0$. Note that

$$P(Z_{N(t)} = 0) = F_N(0,t), \tag{4.1}$$

a consequence of $F_N(\cdot,t)$ being the probability generating function of $Z_{N(t)}$. By proposition 5.3, proposition 3.1 and (4.1), we have

$$\lim_{t\to\infty} h_t(k) = \frac{\varphi(k,\varepsilon)\lim_{t\to\infty} e^{-u'(q)t}P(Z_{N(t)} = k)}{\lim_{t\to\infty}[1 - P(Z_{N(t)} = 0)]} = \frac{\varphi(k,\varepsilon)q_k}{1-q} \quad \text{and}$$

$$\lim_{t\to\infty} r_t(k) = \frac{2(\delta_\varepsilon)^k \lim_{t\to\infty} e^{-u'(q)t}P(Z_{N(t)} = k)}{\lim_{t\to\infty}[1 - P(Z_{N(t)} = 0)]} = \frac{2(\delta_\varepsilon)^k q_k}{1-q}.$$

Summing $r_t(k)$ over k,

$$\begin{aligned}
\sum_{k=1}^{\infty} r_t(k) &= \frac{2e^{-u'(q)t}}{1 - P(Z_{N(t)} = 0)} \sum_{k=1}^{\infty} (\delta_\varepsilon)^k P(Z_{N(t)} = k) \\
&= \frac{2e^{-u'(q)t}}{1 - P(Z_{N(t)} = 0)} \Big[\sum_{k=0}^{\infty} (\delta_\varepsilon)^k P(Z_{N(t)} = k) - (\delta_\varepsilon)^0 P(Z_{N(t)} = 0)\Big] \\
&= \frac{2e^{-u'(q)t}}{1 - P(Z_{N(t)} = 0)} \big[F_N(\delta_\varepsilon, t) - P(Z_{N(t)} = 0)\big] \\
&= \frac{2e^{-u'(q)t}}{1 - P(Z_{N(t)} = 0)} \big[F_N(\delta_\varepsilon, t) - F_N(0,t)\big].
\end{aligned}$$

Hence, in the limit, we have, by proposition 3.2, that

$$\begin{aligned}
\lim_{t\to\infty} \sum_{k=1}^{\infty} r_t(k) &= \lim_{t\to\infty} \frac{2e^{-u'(q)t}}{1 - P(Z_{N(t)} = 0)} \left\{[F_N(\delta_\varepsilon, t) - q] - [F_N(0,t) - q]\right\} \\
&= \frac{2\left\{\lim_{t\to\infty} e^{-u'(q)t}[F_N(\delta_\varepsilon,t) - q] - \lim_{t\to\infty} e^{-u'(q)t}[F_N(0,t) - q]\right\}}{\lim_{t\to\infty}[1 - P(Z_{N(t)} = 0)]} \\
&= \frac{2[Q(\delta_\varepsilon) - Q(0)]}{1-q} \\
&= \frac{2[Q(\delta_\varepsilon) - q_0]}{1-q} < \infty.
\end{aligned}$$

Therefore,

$$\begin{aligned}\sum_{k=1}^{\infty}\lim_{t\to\infty} r_t(k) &= \frac{2}{1-q}\sum_{k=1}^{\infty}(\delta_\varepsilon)^k q_k \\ &= \frac{2}{1-q}\Big[\sum_{k=0}^{\infty}(\delta_\varepsilon)^k q_k - (\delta_\varepsilon)^0 q_0\Big] \\ &= \frac{2}{1-q}\big[Q(\delta_\varepsilon) - q_0\big] \\ &= \lim_{t\to\infty}\sum_{k=1}^{\infty} r_t(k) < \infty.\end{aligned}$$

By **(***)**, we have

$$\lim_{t\to\infty}\sum_{k=1}^{\infty} h_t(k) = \sum_{k=1}^{\infty}\lim_{t\to\infty} h_t(k) = \frac{\sum_{k=1}^{\infty}\varphi(k,\varepsilon)q_k}{1-q} < \infty.$$

It is clear that

$$\begin{aligned}\lim_{t\to\infty}\sum_{k=1}^{\infty} h_t(k) &= \lim_{t\to\infty}\sum_{k=1}^{\infty}\frac{e^{-u'(q)t}\varphi(k,\varepsilon)P(Z_{N(t)}=k)}{1-P(Z_{N(t)}=0)} \\ &= \lim_{t\to\infty} e^{-u'(q)t}P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\Big| > \varepsilon \,\Big|\, Z_{N(t)} > 0\Big).\end{aligned}$$

Therefore,

$$\lim_{t\to\infty} e^{-u'(q)t}P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\Big| > \varepsilon \,\Big|\, Z_{N(t)} > 0\Big) = \frac{\sum_{k=1}^{\infty}\varphi(k,\varepsilon)q_k}{1-q} < \infty.$$

$\square$

The motivation of Theorem 4.3 is to see the behavior of the consecutive generation ratio conditioned on $Z_{N(t)+1} > 0$ instead of conditioned on $Z_{N(t)}$ investigated in Theorem 4.2. The rate of decay in Theorem 4.3 is still exponential, however, the deviation from the mean is less than or equal to that of Theorem 4.2.

Theorem 4.3 *Assume that $p_0 \neq 0$ and that $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$. Then for any $\varepsilon > 0$,*

$$\lim_{t\to\infty} e^{-u'(q)t}P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\Big| > \varepsilon \,\Big|\, Z_{N(t)+1} > 0\Big)$$
$$= \begin{cases} \frac{\sum_{k=1}^{\infty}[\varphi(k,\varepsilon)-p_0^k]q_k}{1-q} < \infty, & \textit{if} \quad 0 < \varepsilon < m, \\ \frac{\sum_{k=1}^{\infty}\varphi(k,\varepsilon)q_k}{1-q} < \infty, & \textit{if} \quad \varepsilon \geq m. \end{cases}$$

Proof. We already know that $\varphi(k,\varepsilon) \le 2(\delta_\varepsilon)^k$ from Theorem 4.2. We can see that

$$\begin{aligned}
&P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)+1} > 0\big) \\
&= \sum_{k=0}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k \,\big|\, Z_{N(t)+1} > 0\big),
\end{aligned}$$

since $\{Z_{N(t)} = k, Z_{N(t)} > 0\,;\, k \ge 0\} = \{Z_{N(t)} = k\,;\, k \ge 1\}$. In addition,

$$\begin{aligned}
&\sum_{k=0}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k \,\big|\, Z_{N(t)+1} > 0\big) \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k,\, Z_{N(t)+1} > 0\big)}{P(Z_{N(t)+1} > 0)} \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k,\, Z_{N(t)+1} > 0\big) P\big(Z_{N(t)} = k,\, Z_{N(t)+1} > 0\big)}{P(Z_{N(t)+1} > 0)} \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k,\, Z_{N(t)+1} > 0\big) P\big(Z_{N(t)+1} > 0 \,\big|\, Z_{N(t)} = k\big) P\big(Z_{N(t)} = k\big)}{P(Z_{N(t)+1} > 0)} \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k,\, Z_{N(t)+1} > 0\big) P\big(Z_1 > 0 \,\big|\, Z_0 = k\big) P\big(Z_{N(t)} = k\big)}{P(Z_{N(t)+1} > 0)},
\end{aligned}$$

by the time homogeneity of the Markov chain $\{Z_n\}_{n=0}^{\infty}$. Moreover,

$$\begin{aligned}
&\sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k,\, Z_{N(t)+1} > 0\big) P\big(Z_1 > 0 \,\big|\, Z_0 = k\big) P\big(Z_{N(t)} = k\big)}{P(Z_{N(t)+1} > 0)} \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon \,\big|\, \sum_{i=1}^{k} X_i > 0\big) P\big(Z_1 > 0 \,\big|\, Z_0 = k\big) P\big(Z_{N(t)} = k\big)}{P(Z_{N(t)+1} > 0)} \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon,\, \sum_{i=1}^{k} X_i > 0\big) P\big(Z_{N(t)} = k\big) P\big(Z_1 > 0 \big| Z_0 = k\big)}{P(\sum_{i=1}^{k} X_i > 0) P(Z_{N(t)+1} > 0)} \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon,\, \sum_{i=1}^{k} X_i > 0\big) P\big(Z_{N(t)} = k\big)}{1 - P(Z_{N(t)+1} = 0)}.
\end{aligned}$$

Let $A := \Big\{|\frac{1}{k}\sum_{i=1}^{k} X_i - m| > \varepsilon\Big\}$ and $B := \Big\{\sum_{i=1}^{k} X_i > 0\Big\}$. Here we consider two different cases depending on the magnitude of ε relative to m.

Case 1: $0 < \varepsilon < m$.

In this case, $B^c \subset A$ and hence $A \cup B = \Omega$, therefore,

$$\begin{aligned}
&P\big(|\frac{1}{k}\sum_{i=1}^{k} X_i - m| > \varepsilon,\, \sum_{i=1}^{k} X_i > 0\big) \\
&= P(A) + P(B) - 1 \\
&= \varphi(k,\varepsilon) + P(\sum_{i=1}^{k} X_i > 0) - 1 \\
&= \varphi(k,\varepsilon) - P(\sum_{i=1}^{k} X_i = 0) \\
&= \varphi(k,\varepsilon) - p_0^k.
\end{aligned}$$

Case 2: $\varepsilon \geq m$.
In this case, $B^c \subset A^c$ and hence $A \subset B$ and thus $A \cap B = A$, therefore,

$$P\big(|\frac{1}{k}\sum_{i=1}^{k} X_i - m| > \varepsilon,\, \sum_{i=1}^{k} X_i > 0\big) = P(A) = \varphi(k,\varepsilon).$$

Thus, the conditional probability,

$$\begin{aligned}
&P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)+1} > 0\big) \\
&= \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon,\, \sum_{i=1}^{k} X_i > 0\big) P\big(Z_{N(t)} = k\big)}{1 - P(Z_{N(t)+1} = 0)} \\
&= \begin{cases} \sum_{k=1}^{\infty} \frac{[\varphi(k,\varepsilon) - p_0^k] P(Z_{N(t)} = k)}{1 - P(Z_{N(t)+1} = 0)} & \text{if} \quad 0 < \varepsilon < m, \\ \sum_{k=1}^{\infty} \frac{\varphi(k,\varepsilon) P(Z_{N(t)} = k)}{1 - P(Z_{N(t)+1} = 0)} & \text{if} \quad \varepsilon \geq m. \end{cases}
\end{aligned}$$

Now we apply **(***)** and the rest of the proof follows similar lines as that of Thereom 4.2 and hence we skip it here. Therefore, we reach

$$\begin{aligned}
&\lim_{t\to\infty} e^{-u'(q)t} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)+1} > 0\big) \\
&= \begin{cases} \frac{\sum_{k=1}^{\infty} [\varphi(k,\varepsilon) - p_0^k] q_k}{1-q} < \infty & \text{if} \quad 0 < \varepsilon < m, \\ \frac{\sum_{k=1}^{\infty} \varphi(k,\varepsilon) q_k}{1-q} < \infty & \text{if} \quad \varepsilon \geq m. \end{cases}
\end{aligned}$$

$\square$

In fact, the result in Theorem 4.1 can be attained by other conditions weaker than the finite exponential moment condition. Theorem 4.4 and Corollary 4.1 serve for this purpose.

Theorem 4.4 *Assume that $p_0 = 0$ and $p_1 > 0$. Assume that there exist constant $\eta > 0$ and $C_\varepsilon > 0$ such that $p_1 m^\eta > 1$ and $\phi(k, \varepsilon) \leq \frac{C_\varepsilon}{k^\eta}$ for each k. Then the result of Theorem 4.1 holds.*

Remark 4.2 *The condition in this theorem is actually weaker than the finite exponential moment condition. Recall that if we have the finite exponential moment condition, then we have a Chernoff type bound (an exponential upper bound), i.e., $\phi(k, \varepsilon) \leq De^{-kI(\varepsilon)}$ for some positive constant D and some positive function I. On the other hand, the condition in this theorem is merely a polynomial upper bound.*

Proof. Let $h_{N(t)}(k) := e^{-\lambda(p_1-1)t}\phi(k, \varepsilon)P(Z_{N(t)} = k)$ and $h'_{N(t)}(k) := e^{-\lambda(p_1-1)t}\frac{C(\varepsilon)}{k^\eta}P(Z_{N(t)} = k)$. Since $\phi(k, \varepsilon) \leq \frac{C_\varepsilon}{k^\eta}$ by the assumption, $h_{N(t)}(k) \leq h'_{N(t)}(k)$. By equation (5.1) in remark 5.3, we have $\lim_{t\to\infty} h_{N(t)}(k) = \hat{q}_k\phi(k, \varepsilon) = h(k)$ and $\lim_{t\to\infty} h'_{N(t)}(k) = \hat{q}_k\frac{C_\varepsilon}{k^\eta}$. Thus, $P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon\big) = \sum_{k=1}^\infty \phi(k, \varepsilon)P(Z_{N(t)} = k) = \sum_{k=1}^\infty h_{N(t)}(k)e^{\lambda(p_1-1)t}$. If we show that $\lim_{t\to\infty}\sum_{k=1}^\infty h'_{N(t)}(k) = \sum_{k=1}^\infty \hat{q}_k\frac{C_\varepsilon}{k^\eta} < \infty$, then, by **(***)**, we obtain

$$\begin{aligned}
\lim_{t\to\infty} e^{-\lambda(p_1-1)t}P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\Big| > \varepsilon\Big) &= \lim_{t\to\infty}\sum_{k=1}^\infty h_{N(t)}(k) \\
&= \sum_{k=1}^\infty \lim_{t\to\infty} h_{N(t)}(k) \\
&= \sum_{k=1}^\infty h(k) \\
&= \sum_{k=1}^\infty \phi(k, \varepsilon)\hat{q}_k < \infty.
\end{aligned}$$

Let us recall that for any non-negative random variable Y and $0 < \eta < \infty$,

$$\begin{aligned}
E(Y^{-\eta}) &= E\Big(\frac{1}{\Gamma(\eta)}\int_0^\infty e^{-tY}t^{\eta-1}\,dt\Big) \\
&= \frac{1}{\Gamma(\eta)}\int_0^\infty E(e^{-tY})t^{\eta-1}\,dt,
\end{aligned}$$

where Γ is the gamma function. We proceed to

$$\begin{aligned}
&\sum_{k=1}^\infty \frac{1}{k^\eta}e^{-\lambda(p_1-1)t}P(Z_{N(t)} = k) \\
= \;& e^{-\lambda(p_1-1)t}E(Z_{N(t)}^{-\eta}) \\
= \;& \frac{1}{\Gamma(\eta)}\int_0^\infty e^{-\lambda(p_1-1)t}F_N(e^{-v}, t)v^{\eta-1}\,dv
\end{aligned}$$

by the above recall. Moreover,

$$\begin{aligned}
& \frac{1}{\Gamma(\eta)}\int_0^\infty e^{-\lambda(p_1-1)t}F_N(e^{-v},t)v^{\eta-1}\,dv \\
= & \frac{1}{\Gamma(\eta)}\int_0^1 e^{-\lambda(p_1-1)t}F_N(s,t)\kappa(s)\,ds,
\end{aligned}$$

by letting $\kappa(s) := \frac{|\log s|^{\eta-1}}{s}$. Therefore, we obtain

$$\Gamma(\eta)e^{-\lambda(p_1-1)t}E(Z_{N(t)}^{-\eta}) = \int_0^1 e^{-\lambda(p_1-1)t}F_N(s,t)\kappa(s)\,ds. \tag{4.2}$$

Recall that $\lim_{t\to\infty} e^{-\lambda(p_1-1)t}F_N(s,t) = \hat{Q}(s) = \sum_{k=1}^\infty \hat{q}_k s^k$ by remark 3.3. First let us show that $e^{-\lambda(p_1-1)t}F_N(s,t)$ is increasing in t for fixed value of s. We can see that

$$\begin{aligned}
\frac{\partial}{\partial t}[e^{-\lambda(p_1-1)t}F_N(s,t)] &= \frac{\partial}{\partial t}[e^{-\lambda(p_1-1)t}\sum_{n=0}^\infty f_n(s)\frac{e^{-\lambda t}(\lambda t)^n}{n!}] \\
&= \frac{\partial}{\partial t}[e^{-\lambda p_1 t}]\sum_{n=0}^\infty f_n(s)\frac{(\lambda t)^n}{n!} + e^{-\lambda p_1 t}\sum_{n=0}^\infty f_n(s)\frac{\partial}{\partial t}[\frac{(\lambda t)^n}{n!}],
\end{aligned}$$

where the interchange of the derivative and the summation is justified by the uniform convergence in the interval of convergence of the series. Now we have

$$\begin{aligned}
& \frac{\partial}{\partial t}[e^{-\lambda p_1 t}]\sum_{n=0}^\infty f_n(s)\frac{(\lambda t)^n}{n!} + e^{-\lambda p_1 t}\sum_{n=0}^\infty f_n(s)\frac{\partial}{\partial t}[\frac{(\lambda t)^n}{n!}] \\
= & -\lambda p_1 e^{-\lambda p_1 t}\sum_{n=0}^\infty f_n(s)\frac{(\lambda t)^n}{n!} + e^{-\lambda p_1 t}\lambda\sum_{j=0}^\infty f_{j+1}(s)\frac{(\lambda t)^j}{j!} \\
> & -\lambda p_1 e^{-\lambda p_1 t}\sum_{n=0}^\infty f_n(s)\frac{(\lambda t)^n}{n!} + e^{-\lambda p_1 t}\lambda\sum_{j=0}^\infty p_1 f_j(s)\frac{(\lambda t)^j}{j!} \\
= & \ 0,
\end{aligned}$$

since $\frac{f_n(s)}{p_1^n}$ is strictly increasing in n and thus implies that $f_{n+1}(s) > p_1 f_n(s)$. Therefore, $e^{-\lambda(p_1-1)t}F_N(s,t)$ is increasing in t.

Since from remark 3.3 we know $\lim_{t\to\infty} e^{-\lambda(p_1-1)t}F_N(s,t) = \hat{Q}(s)$ for $0 \le s < 1$ and from above we know $\frac{\partial}{\partial t}[e^{-\lambda(p_1-1)t}F_N(s,t)] > 0$, we obtain

$$e^{-\lambda(p_1-1)t}F_N(s,t) \uparrow \hat{Q}(s) \quad \text{as } t\to\infty \quad \text{for } 0 \le s < 1. \tag{4.3}$$

Since we have (4.2) and (4.3), by Monotone Convergence Theorem, we reach

$$\Gamma(\eta)e^{-\lambda(p_1-1)t}E(Z_{N(t)}^{-\eta}) \uparrow \int_0^1 \hat{Q}(s)\kappa(s)\,ds \quad \text{as } t\to\infty.$$

Now the proof will be complete if we show that $\int_0^1 \hat{Q}(s)\kappa(s)\,ds < \infty$. Let g be the inverse function of f for $0 \le s \le 1$. Recall that $g_n(s) \uparrow 1$ as $n\to\infty$ for $0<s<1$. Let $0<s_0<1$, then $s_n = g_n(s_0) \uparrow 1$ as $n\to\infty$. By (2.5) in Remark 2.1, we have $\hat{Q}(f(s)) = p_1\hat{Q}(s)$ for $0 \le s < 1$. Define $J_n := \int_{s_n}^{s_{n+1}} \hat{Q}(s)\kappa(s)ds$. Then, by replacing $f(s)$ by u, in the second equality below, we see that

$$\begin{aligned} J_n &= \int_{s_n}^{s_{n+1}} \frac{\hat{Q}(f(s))}{p_1}\kappa(s)ds \\ &= \int_{s_{n-1}}^{s_n} \hat{Q}(u)\frac{\kappa(g(u))g'(u)}{p_1}du \\ &= \int_{s_{n-1}}^{s_n} \hat{Q}(u)\kappa(u)\frac{\kappa(g(u))g'(u)}{p_1\kappa(u)}du. \end{aligned}$$

Notice that $J_n < \infty$ for all $n \in \mathbb{N}$ and that

$$\lim_{u\uparrow 1} \frac{\kappa(g(u))g'(u)}{p_1\kappa(u)} = \frac{1}{p_1 m^\eta}.$$

Since $p_1 m^\eta > 1$ by assumption, for any $0 < \frac{1}{p_1 m^\eta} < \alpha < 1$, there exists an n_0 such that $\frac{\kappa(g(u))g'(u)}{p_1\kappa(u)} \le \alpha$ for $u \ge s_{n_0}$ since $\lim_{n\to\infty} s_n = 1$. Therefore, for all $n \ge n_0+2$,

$$\begin{aligned} J_n &= \int_{s_n}^{s_{n+1}} \hat{Q}(s)\kappa(s)ds \\ &= \int_{s_{n-1}}^{s_n} \hat{Q}(u)\kappa(u)\frac{\kappa(g(u))g'(u)}{p_1\kappa(u)}du \\ &\le \int_{s_{n-1}}^{s_n} \hat{Q}(u)\kappa(u)\alpha du \\ &= \alpha J_{n-1}. \end{aligned}$$

Note that since $J_n \le \alpha J_{n-1}$ foe all $n \ge n_0+2$, we have $J_n \le \alpha^{n-n_0-1} J_{n_0+1}$ foe all $n \ge n_0+2$. Thus,

$$\begin{aligned}
\int_{s_{n_0+2}}^{1} \hat{Q}(s)\kappa(s)ds &= \sum_{n=n_0+2}^{\infty} J_n \\
&\le \sum_{n=n_0+2}^{\infty} \alpha^{n-n_0-1} J_{n_0+1} \\
&= J_{n_0+1} \sum_{j=1}^{\infty} \alpha^j \\
&= J_{n_0+1} \frac{\alpha}{1-\alpha} < \infty
\end{aligned}$$

and hence we see that

$$\int_0^1 \hat{Q}(s)\kappa(s)ds = \int_0^{s_{n_0+2}} \hat{Q}(s)\kappa(s)ds + \int_{s_{n_0+2}}^{1} \hat{Q}(s)\kappa(s)ds < \infty.$$

□

Corollary 4.1 *Assume that $p_0 = 0$ and $p_1 > 0$. Assume that $E(Z_1^{2\eta+\delta}) < \infty$ for some $\delta > 0$ and $\eta \ge 1$ such that $p_1 m^\eta > 1$. Then the result of Theorem 4.1 holds.*

Proof. Since $E(Z_1^{2\eta+\delta}) < \infty$ for some $\delta > 0$ and $\eta \ge 1$, $\sup_k E\big|\frac{\sqrt{k}(\bar{X}_k - m)}{\sigma}\big|^{2\eta} < \infty$. Let us denote $\sup_k E\big|\frac{\sqrt{k}(\bar{X}_k - m)}{\sigma}\big|^{2\eta}$ by $C(\eta)$. By Markov's inequality,

$$\begin{aligned}
\phi(k,\varepsilon) &= P\big(\big|\bar{X}_k - m\big| > \varepsilon\big) \\
&\le P\Big(\Big|\frac{\sqrt{k}(\bar{X}_k - m)}{\sigma}\Big|^{2\eta} \ge \Big(\frac{\sqrt{k}\varepsilon}{\sigma}\Big)^{2\eta}\Big) \\
&\le E\Big|\frac{\sqrt{k}(\bar{X}_k - m)}{\sigma}\Big|^{2\eta} \Big/ \frac{k^\eta \varepsilon^{2\eta}}{\sigma^{2\eta}} \\
&\le \frac{C(\eta)\sigma^{2\eta}}{k^\eta \varepsilon^{2\eta}} \\
&= \frac{C_\varepsilon}{k^\eta},
\end{aligned}$$

by letting $C_\varepsilon := \frac{C(\eta)\sigma^{2\eta}}{\varepsilon^{2\eta}}$. By applying Theorem 4.4, the result is obtained.

□

Under a finite exponential moment assumption about the offspring distribution, Theorem 4.5 establishes a finite uniform exponential moment result for $W_{N(t)}$ which is needed in the proof of Theorem 4.6.

Theorem 4.5 *Assume that $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$. Then there exists $\alpha_1 > 0$ such that*

$$D_1 = \sup_{t\geq 0} E(exp(\alpha_1 W_{N(t)})) < \infty.$$

Proof. By the assumption, $f(e^{\alpha_0}) = E(exp(\alpha_0 Z_1)) < \infty$. If $f(s) \leq e^{\alpha_0}$, then $f_2(s) \leq f(e^{\alpha_0})$, i.e., if $s \leq g(e^{\alpha_0})$, then $f_2(s) \leq f(e^{\alpha_0})$ from the monotonicity of f. Further, if $f(s) \leq g(e^{\alpha_0})$, then $f_3(s) \leq f(e^{\alpha_0})$, i.e., if $s \leq g_2(e^{\alpha_0})$, then $f_3(s) \leq f(e^{\alpha_0})$. Thus, inductively, if $s \leq g_{n-1}(e^{\alpha_0})$, then $f_n(s) \leq f(e^{\alpha_0})$ for all $n \geq 1$.

For any fixed $t \geq 0$, observe that

$$\begin{aligned} E(exp(\alpha W_{N(t)})) &= E(exp(\alpha e^{-u'(1)t} Z_{N(t)})) \\ &= E(f_{N(t)}(e^{\theta e^{-u'(1)t}})) \\ &= \sum_{n=0}^{\infty} f_n(e^{\alpha e^{-u'(1)t}}) P(Z_{N(t)} = n) \\ &\leq f(e^{\alpha_0}), \end{aligned}$$

if $f_n(e^{\alpha e^{-u'(1)t}}) \leq f(e^{\alpha_0})$ for all $n \geq 0$. Notice that $f_n(e^{\alpha e^{-u'(1)t}}) \leq f(e^{\alpha_0})$ for all $n \geq 0$ if $e^{\alpha e^{-u'(1)t}} \leq g_{n-1}(e^{\alpha_0})$ for all $n \geq 0$. Moreover, $e^{\alpha e^{-u'(1)t}} \leq g_{n-1}(e^{\alpha_0})$ for all $n \geq 0$ if and only if $\alpha \leq e^{u'(1)t} \log g_{n-1}(e^{\alpha_0})$ for all $n \geq 0$. Furthermore, $\alpha \leq e^{u'(1)t} \log g_{n-1}(e^{\alpha_0})$ for all $n \geq 0$ is equivalent to $\alpha \leq \inf_n e^{u'(1)t} \log g_{n-1}(e^{\alpha_0})$. Let us observe that since $u'(1)$ is negative,

$$\begin{aligned} \inf_n e^{u'(1)t} \log g_{n-1}(e^{\alpha_0}) &= e^{u'(1)t} \log g_{-1}(e^{\alpha_0}) \\ &= e^{u'(1)t} \log f(e^{\alpha_0}) \\ &\geq e^{u'(1)0} \log f(e^{\alpha_0}) \\ &= \log f(e^{\alpha_0}) < \infty. \end{aligned}$$

Thus, for each fixed t, $E(exp(\alpha W_{N(t)})) \leq f(e^{\alpha_0})$ if $\alpha \leq \log f(e^{\alpha_0})$. Now choose $\alpha_1 = \log f(e^{\alpha_0})$, then for each fixed t, $E(exp(\alpha_1 W_{N(t)})) \leq f(e^{\alpha_0}) < \infty$. Therefore, $\sup_{t\geq 0} E(exp(\alpha_1 W_{N(t)})) \leq f(e^{\alpha_0}) < \infty$.

□

Theorem 4.6 gives the super-exponential rate of decay for $P(|W_{N(t)} - W'| > \varepsilon)$. This result is used in the proof of Theorem 4.7.

Theorem 4.6 *Assume that $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$. Then for any given $\varepsilon > 0$,* *there exist positive constants, $D_3 > 0$ and $\beta > 0$ such that*

$$P(|W_{N(t)} - W'| > \varepsilon) \leq D_3 exp(-\beta\varepsilon^{2/3} e^{\frac{1}{3}u'(1)t}).$$

Proof. Since $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$ by the assumption, there exists $\alpha_1 > 0$ such that $D_1 = \sup_{t\geq 0} E(exp(\alpha_1 W_{N(t)})) < \infty$ by Theorem 4.5. Also from the proof of Theorem 4.5, we know that we can choose $\alpha_1 = \log f(e^{\alpha_0})$. Let $\Psi(\alpha) = E(exp(\alpha W'))$ for $\alpha \leq \alpha_1$. Then $\Psi(\alpha) < \infty$ for $\alpha \leq \alpha_1$. This could be seen from the following by Fatou's Lemma,

$$\begin{aligned} E(exp(\alpha_1 W')) &= E(\lim_{t\to\infty} exp(\alpha_1 W_{N(t)})) \\ &\leq \liminf_{t\to\infty} E((exp(\alpha_1 W_{N(t)})) \\ &\leq \sup_{t\geq 0} E((exp(\alpha_1 W_{N(t)})) \\ &< \infty. \end{aligned}$$

Let $S_k = \sum_{i=1}^{k}(W^{(i)} - 1)$ where $\{W^{(i)}\}_{i=1}^{\infty}$ are i.i.d. copies of W'. Let $\alpha_2 := \min\{\alpha_1, 1\}$. We need to establish an upper bound for $E\big(exp(\alpha_2 S_k/\sqrt{k})\big)$.

from [1] First, note that since $Var(W) = \frac{2Var(X_1)}{m^2-m}$, by L'Hospital's Rule, $\lim_{u\to 0}\frac{\Psi(u)e^{-u}-1}{u^2} = \frac{Var(W)}{2} < \infty$, therefore, we have $\sup_{u\leq\alpha_1}|\frac{\Psi(u)e^{-u}-1}{u^2}| < \infty$. Let us denote this supremum by d. Now we observe that

$$\begin{aligned} \sup_{\alpha\leq\alpha_2}[\Psi(\frac{\alpha}{\sqrt{k}})e^{-\frac{\alpha}{\sqrt{k}}}]^k &\leq \sup_{\alpha\leq\alpha_2}[1 + |\Psi(\frac{\alpha}{\sqrt{k}})e^{-\frac{\alpha}{\sqrt{k}}} - 1|]^k \quad \text{by triangle inequality} \\ &\leq \sup_{\alpha\leq\alpha_2}\Big[1 + \frac{|\Psi(\frac{\alpha}{\sqrt{k}})e^{-\frac{\alpha}{\sqrt{k}}} - 1|}{\alpha^2}\Big]^k \\ &= \Big[1 + \frac{1}{k}\sup_{\alpha\leq\alpha_2}\frac{|\Psi(\frac{\alpha}{\sqrt{k}})e^{-\frac{\alpha}{\sqrt{k}}} - 1|}{\frac{\alpha^2}{k}}\Big]^k \\ &\leq exp\Big(\sup_{\alpha\leq\alpha_2}\frac{|\Psi(\frac{\alpha}{\sqrt{k}})e^{-\frac{\alpha}{\sqrt{k}}} - 1|}{\frac{\alpha^2}{k}}\Big) \quad \text{since } (1+\frac{x}{k})^k \leq e^k \text{ for } x > 0 \\ &\leq e^d. \end{aligned}$$

Let us denote e^d by D_2. Therefore,

$$\begin{aligned} E\big(exp(\alpha_2 S_k/\sqrt{k})\big) &= E\big(exp(\frac{\alpha_2}{\sqrt{k}}\sum_{i=1}^{k}(W^{(i)}-1))\big) \\ &= \big(E(exp(\frac{\alpha_2}{\sqrt{k}}(W'-1)))\big)^k \\ &= \big(\Psi(\frac{\alpha_2}{\sqrt{k}})e^{-\frac{\alpha_2}{\sqrt{k}}}\big)^k \\ &\leq D_2 \quad \text{by the estimate above.} \end{aligned} \tag{4.4}$$

With that, we have established an upper bound D_2 for $E\big(exp(\alpha_2 S_k/\sqrt{k})\big)$. Now turning to the difference,

$$\begin{aligned} W' - W_{N(t)} &= \lim_{v\to\infty}(W_{N(t+v)} - W_{N(t)}) \quad \text{almost surely} \\ &= \lim_{v\to\infty} e^{-u'(1)(t+v)}Z_{N(t+v)} - e^{-u'(1)t}Z_{N(t)} \\ &= e^{-u'(1)t}[\lim_{v\to\infty} e^{-u'(1)v}Z_{N(t+v)} - Z_{N(t)}] \\ &= e^{-u'(1)t}[\lim_{v\to\infty} e^{-u'(1)v}\sum_{j=1}^{Z_{N(t)}} Z^{(j)}_{N(t+v)-N(t)} - Z_{N(t)}] \quad \text{by conditioning on } Z_{N(t)} \\ &= e^{-u'(1)t}\sum_{j=1}^{Z_{N(t)}}(W^{(j)}-1). \end{aligned} \tag{4.5}$$

It follows that

$$\begin{aligned} & P(W' - W_{N(t)} > \varepsilon \,|\, Z_0, Z_1, ..., Z_{N(t)}) \\ = \ & P(W' - W_{N(t)} > \varepsilon \,|\, Z_{N(t)}) \quad \text{by Markov property} \\ = \ & P(e^{-u'(1)t}\sum_{j=1}^{Z_{N(t)}}(W^{(j)}-1) > \varepsilon \,|\, Z_{N(t)}) \quad \text{by (4.5)} \\ = \ & P(S_{Z_{N(t)}} > e^{u'(1)t}\varepsilon \,|\, Z_{N(t)}). \end{aligned}$$

We see that

$$\begin{aligned} P(S_k > \zeta) &= P(\frac{S_k}{\sqrt{k}} > \frac{\zeta}{\sqrt{k}}) \\ &= P(e^{\alpha_2 S_k/\sqrt{k}} > e^{\alpha_2\zeta/\sqrt{k}}) \\ &\leq E(e^{\alpha_2 S_k/\sqrt{k}})e^{-\alpha_2\zeta/\sqrt{k}} \quad \text{by Markov's inequality} \\ &\leq D_2 exp\big(-\frac{\alpha_2\zeta}{\sqrt{k}}\big), \end{aligned} \tag{4.6}$$

where the last equality is due to (4.4). We see the probability

$$\begin{aligned} P(W' - W_{N(t)} > \varepsilon) &= E[P(W' - W_{N(t)} > \varepsilon \,|\, Z_{N(t)})] \\ &= E[P(S_{Z_{N(t)}} > e^{u'(1)t}\varepsilon \,|\, Z_{N(t)})] \\ &\leq E[D_2 exp\big(- \frac{\alpha_2 e^{u'(1)t}\varepsilon}{\sqrt{Z_{N(t)}}}\big)] \quad \text{by (4.6)} \\ &= D_2 E\big(exp\big(- \alpha_2 \varepsilon \sqrt{e^{u'(1)t}} \frac{1}{\sqrt{W_{N(t)}}}\big)\big). \end{aligned}$$

Hence, the expectation

$$\begin{aligned} & E\big(exp\big(- \tau\big(\frac{1}{\sqrt{W_{N(t)}}}\big)\big)\big) \\ = & \int_0^\infty P\big(exp\big(- \tau\big(\frac{1}{\sqrt{W_{N(t)}}}\big)\big) \geq v\big)\, dv \\ = & \ \tau \int_0^\infty e^{-\tau u} P\big(\frac{1}{\sqrt{W_{N(t)}}} \leq u\big)\, du \\ = & \ \tau \int_0^\infty e^{-\tau u} P\big(exp(\alpha_1 W_{N(t)}) \geq exp(\frac{\alpha_1}{u^2})\big)\, du \\ \leq & \ \tau \int_0^\infty e^{-\tau u} E\big(exp(\alpha_1 W_{N(t)})\big) exp(-\frac{\alpha_1}{u^2})\, du \quad \text{by Markov's inequality} \\ \leq & \ \tau \int_0^\infty e^{-\tau u} D_1 exp(-\frac{\alpha_1}{u^2})\, du \quad \text{by Theorem 4.6} \\ = & \ D_1 \int_0^\infty e^{-r} exp\big(- \frac{\alpha_1 \tau^2}{r^2}\big)\, dr. \end{aligned}$$

Therefore,

$$P(W - W_{N(t)} > \varepsilon) \leq D_2 D_1 \cdot \int_0^\infty e^{-r} exp\big(- \frac{\alpha_1 \tau_t^2}{r^2}\big)\, dr, \tag{4.7}$$

where τ_t denote the quantity $\alpha_2 \varepsilon \sqrt{e^{u'(1)t}}$. Let $I(\tau) := \int_0^\infty e^{-r} e^{-\frac{\tau^2}{r^2}}\, dr$. Then by choosing $k(\tau) = \tau^{\frac{2}{3}}$, we can see that

$$\begin{aligned} I(\tau) &= \int_0^\infty e^{-r} e^{-\frac{\tau^2}{r^2}}\, dr \\ &= \int_0^{k(\tau)} e^{-r} e^{-\frac{\tau^2}{r^2}}\, dr + \int_{k(\tau)}^\infty e^{-r} e^{-\frac{\tau^2}{r^2}}\, dr \\ &\leq exp\big(- \frac{\tau^2}{(k(\tau))^2}\big) + e^{-k(\tau)} \\ &\leq 2exp(-\tau^{\frac{2}{3}}). \end{aligned} \tag{4.8}$$

Consequently, by (4.7) and (4.8), we have

$$\begin{aligned} P(W' - W_{N(t)} > \varepsilon) &\leq 2D_2D_1exp\big(\big(-\sqrt{\theta_1}\alpha_2\varepsilon[e^{u'(1)t}]^{\frac{1}{2}}\big)^{\frac{2}{3}}\big) \\ &= D_3exp\big(-\beta[e^{u'(1)t}]^{\frac{1}{3}}\varepsilon^{\frac{2}{3}}\big) \\ &= D_3exp\big(-\beta e^{\frac{1}{3}u'(1)t}\varepsilon^{\frac{2}{3}}\big), \end{aligned}$$

where $D_3 := 2D_2D_1$ and $\beta := (\sqrt{\alpha_1}\alpha_2)^{\frac{2}{3}}$. The argument for $P(W_{N(t)} - W' > \varepsilon)$ is essentially identical. Therefore, the result is obtained.

$\square$

Theorem 4.7 shows a super-exponential rate of decay for the ratio of successive generation sizes conditioned on the limiting random variable W' staying positive.

Theorem 4.7 *Assume that $E(exp(\alpha_0 Z_1)) < \infty$ for some $\alpha_0 > 0$. Then there exists positive constants, $D_5 > 0$ and $\tau > 0$ such that for any $\varepsilon > 0$ and $d > 0$, we can find some $0 < I(\varepsilon) < \infty$ such that*

$$P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, W' \geq d\big) \leq \alpha_d\big[D_5exp\big(-d\gamma I(\varepsilon)e^{u'(1)t}\big) + D_3exp\big(-\tau\big[d(1-\gamma)\big]^{\frac{2}{3}}e^{\frac{1}{3}u'(1)t}\big)\big]$$

for any $0 < \gamma < 1$, where $\alpha_d = \frac{1}{P(W' \geq d)}$.

Proof. First of all, we start with

$$\begin{aligned} & P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, W' \geq d\big) \\ = \; & P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, W' \geq d\big)\frac{1}{P(W' \geq d)} \\ = \; & \alpha_d\big[P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, W_{N(t)} < d\gamma, W' \geq d\big) \\ & + P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, W_{N(t)} \geq d\gamma, W' \geq d\big)\big]. \end{aligned}$$

Let I_1 denote $P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, W_{N(t)} < d\gamma, W' \geq d\big)$ and similarly let I_2 denote $P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, W_{N(t)} \geq d\gamma, W' \geq d\big)$.

Let us estimate I_2 first. It is clear that

$$
\begin{aligned}
I_2 &\leq P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, W_{N(t)} \geq d\gamma\big) \\
&= P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} \geq d\gamma e^{u'(1)t}\big) P\big(Z_{N(t)} \geq d\gamma e^{u'(1)t}\big) \\
&\leq P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} \geq d\gamma e^{u'(1)t}\big) \\
&= \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, Z_{N(t)} = k \,\big|\, Z_{N(t)} \geq d\gamma e^{u'(1)t}\big).
\end{aligned}
$$

Then we see that

$$
\begin{aligned}
& \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon, Z_{N(t)} = k \,\big|\, Z_{N(t)} \geq d\gamma e^{u'(1)t}\big) \\
= & \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\Big|\, Z_{N(t)} = k, Z_{N(t)} \geq d\gamma e^{u'(1)t}\big) \\
& P\big(Z_{N(t)} = k \big| Z_{N(t)} \geq a\gamma e^{u'(1)t}\big) \\
= & \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k\big) P\big(Z_{N(t)} = k \,\big|\, Z_{N(t)} \geq a\gamma e^{u'(1)t}\big) \\
\leq & \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k\big) \\
= & \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon\big) \\
= & \sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{1}{k}\sum_{i=1}^{k} Y_i\big| > \varepsilon\big),
\end{aligned}
$$

by letting $Y_i := X_i - m$. Notice that Y_i's are i.i.d..

Since Y_1 has finite exponential moment, we have a Chernoff type bound, i.e., an exponential upper bound that decays in the number of the empirical mean. Therefore, we obtain

$$
\begin{aligned}
&\sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} P\big(\big|\frac{1}{k}\sum_{i=1}^{k} Y_i\big| > \varepsilon\big) \\
\leq &\sum_{k=\lceil d\gamma e^{u'(1)t}\rceil}^{\infty} \frac{D_4}{e^{I(\varepsilon)k}} \quad \text{for some}\, D_4 > 0 \,\text{and}\, I(\varepsilon) > 0 \\
\leq &\; D_4 \frac{e^{-I(\varepsilon)d\gamma e^{u'(1)t}}}{1-e^{-I(\varepsilon)}} \quad \text{since}\; d\gamma e^{u'(1)t} \leq \lceil d\gamma e^{u'(1)t}\rceil \\
\doteq &\; D_5 exp\big(-I(\varepsilon)d\gamma e^{u'(1)t}\big) \quad \text{for some}\, D_5 > 0.
\end{aligned}
$$

Hence, we obtain

$$
I_2 \leq D_5 exp\big(-I(\varepsilon)d\gamma e^{u'(1)t}\big). \tag{4.9}
$$

Finally, by Theorem 4.6, one can see that

$$
\begin{aligned}
I_1 &\leq P(W_{N(t)} < d\gamma, W' \geq d) \\
&= P(W' - W_{N(t)} > d(1-\gamma)) \\
&\leq D_3 exp(-\tau(d(1-\gamma))^{\frac{2}{3}} e^{\frac{1}{3}u'(1)t}),
\end{aligned} \tag{4.10}
$$

for some $\tau > 0$ and $D_3 > 0$. Therefore, combining (4.9) and (4.10), we reach

$$
\begin{aligned}
P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,|\, W' \geq d\big) \;\leq\; &\alpha_d\big[D_5 exp(-I(\varepsilon)d\gamma e^{u'(1)t}) \\
&+ D_3 exp(-\tau(d(1-\gamma))^{\frac{2}{3}} e^{\frac{1}{3}u'(1)t})\big].
\end{aligned}
$$

$\square$

Chapter 5

Applications to Finance and Physics

5.1 An Application to Finance - Mean Reversion

In [14], Epps proposed a PRIBP to model the short-term behavior of stock price. He provided an empirical study for U.S. stocks markets and found that this model worked well. In his work, he explains the reasons for using a PRIBP to model stock prices. First of all, although stock price is not integer-valued, it is the multiple of tick size such as multiple of $\$\frac{1}{8}$, $\$\frac{1}{16}$, or \$0.01 in primary U.S. stock markets. That is, stock price takes values in a discrete space. In the Black-Scholes model, the stock price is assumed to follow a geometric Brownian motion, which takes positive real values, and hence the movement of stock price is continuous. Although nowadays this assumption is widely used by many researchers and practitioners, it fails to capture the discreteness of stock prices. On the other hand, PRIBP successfully captures this discrete nature of the discrete stock price fluctuation. Secondly, this model allows the possibility of bankruptcy of a firm. Under the geometric Brownian motion assumption, stock price cannot reach zero. However, in the PRIBP, the stock price is allowed to attain zero and this situation is viewed as the bankruptcy of the firm. Finally, the model shows the inverse relationship between the variance of return and the initial stock price that is suggested in the literature.

To understand the mechanism of a stock price as a PRIBP, let us assume that the tick size in a stock market is $\$\frac{1}{8}$. Then all stocks in this market take values in the space $\{j * \frac{1}{8} \mid j \in \mathbb{N} \cup \{0\}\}$. Then we translate the price of a stock $\{S(t)\}_{t\geq 0}$ into the number of individuals $\{Z_{N(t)}\}_{t\geq 0}$ by $Z_{N(t)} = S(t) * 8$. For example, if a stock price is $\$24\frac{1}{4}$ at the initial time 0, then $Z_{N(0)} = S(0) * 8 = 24\frac{1}{4} * 8 = 194$. In this case, the PRIBP starts from $Z_0 = 194$ many ancestors. The subordinator is a Poisson process which represents the occurrence of price fluctuations. This assumption implies that the shock of the information about a given stock is modeled by a Poisson process. Since in the model of a stock price as a PRIBP, the initial price is usually not 1, we need to adjust the results about a PRIBP starting from one ancestor to those starting from an arbitrary positive integer number of ancestors. That is, $Z_0 = l$, where $l \geq 1$. Let $F_{N,l}(s,t) = E(s^{Z_{N(t)}}|Z_0 = l)$ be the p.g.f. of $Z_{N(t)}$ when $Z_0 = l \geq 1$, i.e., when there are l many ancestors. Note that when $Z_0 = l$, $E(Z_{N(t)}) = le^{u'(1)t} = le^{\lambda(m-1)t}$.

We now phrase some adjusted results for a PRIBP with Z_0 taking an arbitrary positive integer. Proposition 5.1 is an extension version of proposition 3.1 in the previous chapter; Proposition 5.2 is an extension of proposition 3.2.

Proposition 5.1 $\lim_{t\to\infty} F_{N,l}(s,t) = q^l \quad$ *for all* $\quad 0 \leq s < 1$.

Remark 5.1 *When* $p_0 = 0$*, then* $q = 0$ *and hence* $\lim_{t\to\infty} F_{N,l}(s,t) = 0$ *for all* $0 \leq s < 1$.

Proof. Let us start with

$$\begin{aligned} F_{N,l}(s,t) &= E(s^{Z_{N(t)}} \mid Z_0 = l) \\ &= \sum_{n=0}^{\infty} E(s^{Z_n} \mid Z_0 = l) P(N(t) = n) \\ &= \sum_{n=0}^{\infty} (f_n(s))^l P(N(t) = n) \\ &= E[(f_{N(t)}(s))^l]. \end{aligned}$$

Since $\lim_{n\to\infty} f_n(s) = q$ for $0 \leq s < 1$, $\lim_{t\to\infty} f_{N(t)}(s) = q$ a.s. for $0 \leq s < 1$. By dominated convergence theorem, for $0 \leq s < 1$,

$$\begin{aligned} \lim_{t\to\infty} F_{N,l}(s,t) &= \lim_{t\to\infty} E[(f_{N(t)}(s))^l] \\ &= E[\lim_{t\to\infty} (f_{N(t)}(s))^l] \\ &= q^l. \end{aligned}$$

$\square$

Proposition 5.2 *Assume that* $m \neq 1$*. If* $p_0 \neq 0$*, then*

$$\lim_{t\to\infty} e^{-u'(q)t}[F_{N,l}(s,t) - q^l] = lq^{l-1}Q(s) = lq^{l-1} \sum_{k=0}^{\infty} q_l s^l < \infty \quad \textit{for all } 0 \leq s < 1.$$

Moreover, $Q(s)$ *is the unique solution of a functional equation*

$$Q(f(s)) = f'(q)Q(s) \quad \textit{for all } 0 \leq s < 1.$$

Remark 5.2 *In particular, when $p_0 = 0$ and $p_1 \neq 0$, then the theorem becomes*

$$\lim_{t\to\infty} e^{-\lambda(p_1-1)t} F_{N,l}(s,t) = \begin{cases} 0 & \text{if} \quad l > 1, \\ \hat{Q}(s) = \sum_{k=1}^{\infty} \hat{q}_k s^k & \text{if} \quad l = 1, \end{cases} \quad \text{for all} \quad 0 \leq s < 1.$$

Moreover, $\hat{Q}(s)$is the unique solution of $\hat{Q}(f(s)) = p_1\hat{Q}(s)$ for all $0 \leq s < 1$.

Proof. Define $A(s,t) := e^{-u'(q)t}[F_{N,l}(s,t) - q^l] = e^{-u'(q)t}[E[(f_{N(t)}(s))^l] - q^l]$. First, note that

$$\begin{aligned} \frac{\partial}{\partial s} A(s,t) &= e^{-u'(q)t} \frac{\partial}{\partial s} [\sum_{n=0}^{\infty} (f_n(s))^l P(N(t) = n)] \\ &= e^{-u'(q)t} \sum_{n=0}^{\infty} \frac{\partial}{\partial s} [(f_n(s))^l P(N(t) = n)], \end{aligned}$$

where the interchange of derivative and summation is justified as showed in the proof of proposition 3.2. Recall $V_n(s)$ defined in proposition 3.2. We can see that

$$\begin{aligned} & e^{-u'(q)t} \sum_{n=0}^{\infty} \frac{\partial}{\partial s} [(f_n(s))^l P(N(t) = n)] \\ = & e^{-u'(q)t} \sum_{n=0}^{\infty} P(N(t) = n) l (f_n(s))^{l-1} f_n'(s) \\ = & e^{-u'(q)t} \sum_{n=0}^{\infty} P(N(t) = n) l (f_n(s))^{l-1} \prod_{j=0}^{n-1} f'(f_j(s)) \\ = & e^{-\lambda f'(q)t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} l (f_n(s))^{l-1} \prod_{j=0}^{n-1} f'(f_j(s)) \\ = & \sum_{n=0}^{\infty} P(N(f'(q)t) = n) l (f_n(s))^{l-1} \prod_{j=0}^{n-1} \frac{f'(f_j(s))}{f'(q)} \\ = & \sum_{n=0}^{\infty} P(N(f'(q)t) = n) l (f_n(s))^{l-1} V_n'(s) \\ = & E[l(f_{N(f'(q)t))}(s))^{l-1} V'_{N(f'(q)t))}(s)], \end{aligned}$$

where $N(f'(q)t))$ is a Poisson random variable with parameter $\lambda f'(q)t$.

The rest of the proof follows similar lines to that for proposition 3.2. Hence we only give the parts with some difference here. First, note that

$$\begin{aligned}\lim_{t\to\infty}\frac{\partial}{\partial s}A(s,t) &= E[\lim_{t\to\infty} l(f_{N(f'(q)t))}(s))^{l-1}V'_{N(f'(q)t))}(s)]\\ &= lq^{l-1}V'(s) \quad \text{for all } 0\le s<1.\end{aligned}$$

Since $A(s,t) = A(s,t) - A(q,t) = \int_q^s \frac{\partial}{\partial v}A(v,t)\,dv$,

$$\begin{aligned}\lim_{t\to\infty} A(s,t) &= \lim_{t\to\infty}\int_q^s \frac{\partial}{\partial v}A(v,t)\,dv\\ &= \int_q^s \lim_{t\to\infty}\frac{\partial}{\partial v}A(v,t)\,dv\\ &= \int_q^s lq^{l-1}V'(v)\,dv\\ &= lq^{l-1}[V(s)-V(q)]\\ &= lq^{l-1}\{[Q(s)+C]-[Q(q)+C]\}\\ &= lq^{l-1}Q(s).\end{aligned}$$

Therefore, $\quad \lim_{t\to\infty} e^{-u'(q)t}[F_{N,l}(s,t)-q^l] = lq^{l-1}Q(s) = lq^{l-1}\sum_{k=0}^{\infty} q_k s^k < \infty \quad$ for all $\quad 0\le s<1.$

Notice that when $p_0=0$, we need the condition that $p_1\neq 0$. Since if $p_0=0$, then $q=0$, and hence $f'(q)=f'(0)=p_1$. Thus, if $p_0=0$ and $p_1=0$, then $f'(q)=f'(0)=p_1=0$, and hence Q_n and Q'_n are undefined. Therefore, under the assumptions that $p_0=0$ and $p_1\neq 0$, if $l=1$, then $\lim_{t\to\infty} e^{-\lambda(p_1-1)t}F_{N,l}(s,t)=\hat{Q}(s)$, which is the case in remark 3.3. Following the similar lines of this proof, we can see that under the assumptions that $p_0=0$ and $p_1\neq 0$, if $l>1$, then

$$\frac{\partial}{\partial s}\{e^{-\lambda(p_1-1)t}[E[(f_{N(t)}(s))^l]\} = E[l(f_{N(f'(q)t))}(s))^{l-1}\hat{V}'_{N(f'(q)t))}(s)].$$

Hence, by taking limit in t on both sides, we have

$$\begin{aligned}\lim_{t\to\infty}\frac{\partial}{\partial s}\{e^{-\lambda(p_1-1)t}[E[(f_{N(t)}(s))^l]\} &= \lim_{t\to\infty} E[l(f_{N(f'(q)t))}(s))^{l-1}\hat{V}'_{N(f'(q)t))}(s)]\\ &= l0^{l-1}\hat{V}'(s)\\ &= 0 \quad \text{if } l\neq 1.\end{aligned}$$

Consequently, $\lim_{t\to\infty} e^{-\lambda(p_1-1)t}F_{N,l}(s,t)=0$ if $l>1$.

□

Proposition 5.3 *Assume that $m \neq 1$. Let $Z_0 = l$. If $p_0 \neq 0$, then*

$$lq^{l-1}q_k = \begin{cases} \lim_{t\to\infty} e^{-u'(q)t} P(Z_{N(t)} = k \mid Z_0 = l) & \text{if} \quad k \geq 1, \\ \lim_{t\to\infty} e^{-u'(q)t} [P(Z_{N(t)} = k \mid Z_0 = l) - q^l] & \text{if} \quad k = 0. \end{cases}$$

Remark 5.3 *In particular, if $p_0 = 0$ and $p_1 \neq 0$, then for any $k \geq 1$*

$$\lim_{t\to\infty} e^{-\lambda(p_1-1)t} P(Z_{N(t)} = k \mid Z_0 = l) = \begin{cases} 0 & \text{if} \quad l > 1, \\ \hat{q}_k & \text{if} \quad l = 1. \end{cases} \tag{5.1}$$

Proof. First, consider the case that $p_0 \neq 0$. By proposition 5.2, for $0 \leq s < 1$, on one hand,

$$\begin{aligned} \lim_{t\to\infty} e^{-u'(q)t}[F_{N,l}(s,t) - q^l] &= lq^{l-1}Q(s) \\ &= lq^{l-1}\sum_{k=0}^{\infty} q_k s^k \\ &= \sum_{k=0}^{\infty} lq^{l-1} q_k s^k. \end{aligned} \tag{5.2}$$

On the other hand,

$$\begin{aligned} &\lim_{t\to\infty} e^{-u'(q)t}[F_{N,l}(s,t) - q^l] \\ = \; &\lim_{t\to\infty} e^{-u'(q)t}[\sum_{n=0}^{\infty} (f_n(s))^l P(N(t) = n) - q^l] \\ = \; &\lim_{t\to\infty} e^{-u'(q)t}[\sum_{n=0}^{\infty} \big(\sum_{k=0}^{\infty} p_n(1,k)s^k\big)^l P(N(t) = n) - q^l] \\ = \; &\lim_{t\to\infty} e^{-u'(q)t}[\sum_{n=0}^{\infty}\sum_{k=0}^{\infty} p_n(l,k)s^k P(N(t) = n) - q^l] \\ = \; &\lim_{t\to\infty} e^{-u'(q)t}[\sum_{k=0}^{\infty}\sum_{n=0}^{\infty} p_n(l,k)s^k P(N(t) = n) - q^l], \end{aligned}$$

where the interchange of two summations is justified by Fubini's Theorem because of non-negative terms. Moreover,

$$
\begin{aligned}
& \lim_{t\to\infty} e^{-u'(q)t}[\sum_{k=0}^{\infty}\sum_{n=0}^{\infty} p_n(l,k)s^k P(N(t)=n) - q^l] \\
= & \lim_{t\to\infty} e^{-u'(q)t}[\sum_{k=0}^{\infty} P(Z_{N(t)}=k|Z_0=l)s^k - q^l] \\
= & \lim_{t\to\infty} e^{-u'(q)t}\sum_{k=0}^{\infty}[P(Z_{N(t)}=k|Z_0=l)-c_k]s^k, \\
& \text{where } c_k = \begin{cases} 0 & \text{if} \quad k\geq 1, \\ q^l & \text{if} \quad k=0. \end{cases}
\end{aligned}
$$

Furthermore, since we have the uniform convergence in the interval of convergence of the power series, we can interchange of the limit and summation. Therefore,

$$
\begin{aligned}
& \lim_{t\to\infty} e^{-u'(q)t}\sum_{k=0}^{\infty}[P(Z_{N(t)}=k\,|\,Z_0=l)-c_k]s^k \\
= & \sum_{k=0}^{\infty}\lim_{t\to\infty} e^{-u'(q)t}[P(Z_{N(t)}=k\,|\,Z_0=l)-c_k]s^k.
\end{aligned}
$$

Therefore,

$$
\lim_{t\to\infty} e^{-u'(q)t}[F_{N,l}(s,t)-q^l] = \sum_{k=0}^{\infty}\lim_{t\to\infty} e^{-u'(q)t}[P(Z_{N(t)}=k\,|\,Z_0=l)-c_k]s^k. \tag{5.3}
$$

By comparing the coefficients in the equations (5.2) and (5.3),

$$
\begin{aligned}
lq^{l-1}q_k &= \lim_{t\to\infty} e^{-u'(q)t}[P(Z_{N(t)}=k\,|\,Z_0=l)-c_k] \\
&= \begin{cases} \lim_{t\to\infty} e^{-u'(q)t}P(Z_{N(t)}=k\,|\,Z_0=l) & \text{if} \quad k\geq 1, \\ \lim_{t\to\infty} e^{-u'(q)t}[P(Z_{N(t)}=k\,|\,Z_0=l)-q^l] & \text{if} \quad k=0. \end{cases}
\end{aligned} \tag{5.4}
$$

Now we turn to the case $p_0=0$ and $p_1\neq 0$. By proposition 5.2, on one hand,

$$
\lim_{t\to\infty} e^{-\lambda(p_1-1)t}F_{N,l}(s,t) = \begin{cases} 0 & \text{if} \quad l>1, \\ \hat{Q}(s)=\sum_{k=1}^{\infty}\hat{q}_k s^k & \text{if} \quad l=1. \end{cases} \tag{5.5}
$$

On the other hand,

$$\begin{aligned}
\lim_{t\to\infty} e^{-\lambda(p_1-1)t} F_{N,l}(s,t) &= \lim_{t\to\infty} e^{-\lambda(p_1-1)t} \sum_{n=0}^{\infty} (f_n(s))^l P(N(t)=n) \\
&= \lim_{t\to\infty} e^{-\lambda(p_1-1)t} \sum_{n=0}^{\infty} \big(\sum_{k=1}^{\infty} p_n(1,k)s^k\big)^l P(N(t)=n) \\
&= \lim_{t\to\infty} e^{-\lambda(p_1-1)t} \sum_{n=0}^{\infty}\sum_{k=1}^{\infty} p_n(l,k)s^k P(N(t)=n) \\
&= \lim_{t\to\infty} e^{-\lambda(p_1-1)t} \sum_{k=1}^{\infty}\sum_{n=0}^{\infty} p_n(l,k)s^k P(N(t)=n),
\end{aligned}$$

where the interchange of two summations is justified by Fubini's Theorem because of non-negative terms. In addition,

$$\begin{aligned}
& \lim_{t\to\infty} e^{-\lambda(p_1-1)t} \sum_{k=1}^{\infty}\sum_{n=0}^{\infty} p_n(l,k)s^k P(N(t)=n) \\
= & \lim_{t\to\infty} e^{-\lambda(p_1-1)t} \sum_{k=1}^{\infty} P(Z_{N(t)}=k \,|\, Z_0=l)s^k \\
= & \sum_{k=1}^{\infty} \lim_{t\to\infty} e^{-\lambda(p_1-1)t} P(Z_{N(t)}=k \,|\, Z_0=l)s^k,
\end{aligned}$$

where the interchange of the limit and summation is justified by the uniform convergence in the interval of convergence of the power series. Therefore,

$$\lim_{t\to\infty} e^{-\lambda(p_1-1)t} F_{N,l}(s,t) = \sum_{k=1}^{\infty} \lim_{t\to\infty} e^{-\lambda(p_1-1)t} P(Z_{N(t)}=k \,|\, Z_0=l)s^k. \tag{5.6}$$

By comparing the coefficients in the equations (5.5) and (5.6), we conclude that

$$\lim_{t\to\infty} e^{-\lambda(p_1-1)t} P(Z_{N(t)}=k \,|\, Z_0=i) = \begin{cases} 0 & \text{if} \quad l>1, \\ \hat{q}_k & \text{if} \quad l=1, \end{cases} \qquad \text{for any} \quad k\ge 1.$$

$\square$

Theorem 5.1 *Assume that $p_0 \neq 0$. Let $Z_0 = l$. Assume that $E(exp(\alpha_0 Z_1)) < \infty$ for some*

$\alpha_0 > 0$. *Then for any* $\varepsilon > 0$,

$$\lim_{t\to\infty} e^{-u'(q)t} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} > 0\big) = \frac{\sum_{k=1}^{\infty} \varphi(k,\varepsilon) l q^{l-1} q_k}{1-q^l} < \infty,$$

$$\textit{where} \quad \varphi(k,\varepsilon) = P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon\big).$$

Proof. By the same estimate as in Theorem 4.1, $\varphi(k,\varepsilon) \le 2(\delta_\varepsilon)^k$. Now we begin with the conditional probability,

$$\begin{aligned}
& P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} > 0\big) \\
= & \sum_{k=0}^{\infty} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k \,\big|\, Z_{N(t)} > 0\big) \\
= & \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon,\, Z_{N(t)} = k\big)}{P(Z_{N(t)} > 0)} \\
= & \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} = k\big) P\big(Z_{N(t)} = k\big)}{P(Z_{N(t)} > 0)} \quad \text{by independence} \\
= & \sum_{k=1}^{\infty} \frac{P\big(\big|\frac{1}{k}\sum_{k=1}^{k} X_i - m\big| > \varepsilon\big) P\big(Z_{N(t)} = k\big)}{P(Z_{N(t)} > 0)} \\
= & \sum_{k=1}^{\infty} \frac{\varphi(k,\varepsilon) P\big(Z_{N(t)} = k\big)}{1 - P(Z_{N(t)} = 0)}.
\end{aligned}$$

Let us define $h_t(k) := \frac{e^{-u'(q)t}\varphi(k,\varepsilon)P(Z_{N(t)}=k)}{1-P(Z_{N(t)}=0)}$ and let $r_t(k) := \frac{e^{-u'(q)t}2(\delta_\varepsilon)^k P(Z_{N(t)}=k)}{1-P(Z_{N(t)}=0)}$. Then for each $k \ge 1, h_t(k) \le r_t(k)$ for all $t \ge 0$. Note that $P(Z_{N(t)} = 0) = F_{N,l}(0,t)$. By proposition 5.3, taking limits in t, we have

$$\begin{aligned}
\lim_{t\to\infty} h_t(k) &= \frac{\varphi(k,\varepsilon)\lim_{t\to\infty} e^{-u'(q)t} P(Z_{N(t)} = k)}{\lim_{t\to\infty}[1 - P(Z_{N(t)} = 0)]} \\
&= \frac{\varphi(k,\varepsilon) l q^{l-1} q_k}{1-q^l} \quad \text{and}
\end{aligned}$$

$$\begin{aligned}
\lim_{t\to\infty} r_t(k) &= \frac{2(\delta_\varepsilon)^k \lim_{t\to\infty} e^{-u'(q)t} P(Z_{N(t)} = k)}{\lim_{t\to\infty}[1 - P(Z_{N(t)} = 0)]} \\
&= \frac{2(\delta_\varepsilon)^k l q^{l-1} q_k}{1-q^l}.
\end{aligned}$$

On the otehr hand, the sum of $r_t(k)$ in k gives

$$\begin{aligned}
\sum_{k=1}^{\infty} r_t(k) &= \frac{2e^{-u'(q)t}}{1-P(Z_{N(t)}=0)} \sum_{k=1}^{\infty} (\delta_\varepsilon)^k P(Z_{N(t)}=k) \\
&= \frac{2e^{-u'(q)t}}{1-P(Z_{N(t)}=0)} \big[\sum_{k=0}^{\infty} (\delta_\varepsilon)^k P(Z_{N(t)}=k) - (\delta_\varepsilon)^0 P(Z_{N(t)}=0) \big] \\
&= \frac{2e^{-u'(q)t}}{1-P(Z_{N(t)}=0)} \big[F_{N,l}(\delta_\varepsilon, t) - P(Z_{N(t)}=0) \big] \\
&= \frac{2e^{-u'(q)t}}{1-P(Z_{N(t)}=0)} \big[F_{N,l}(\delta_\varepsilon, t) - F_{N,l}(0, t) \big].
\end{aligned}$$

By proposition 5.2, taking limit on $\sum_{k=1}^{\infty} r_t(k)$ in t, we have

$$\begin{aligned}
\lim_{t\to\infty} \sum_{k=1}^{\infty} r_t(k) &= \lim_{t\to\infty} \frac{2e^{-u'(q)t}}{1-P(Z_{N(t)}=0)} \big\{ [F_{N,l}(\delta_\varepsilon, t) - q^l] - [F_{N,l}(0,t) - q^l] \big\} \\
&= \frac{2\big\{ \lim_{t\to\infty} e^{-u'(q)t} [F_{N,l}(\delta_\varepsilon, t) - q^l] - \lim_{t\to\infty} e^{-u'(q)t} [F_{N,l}(0,t) - q^l] \big\}}{\lim_{t\to\infty} [1 - P(Z_{N(t)}=0)]} \\
&= \frac{2\big\{ \lim_{t\to\infty} e^{-u'(q)t} [F_{N,l}(\delta_\varepsilon, t) - q^l] - \lim_{t\to\infty} e^{-u'(q)t} [F_{N,l}(0,t) - q^l] \big\}}{1-q^l} \\
&= \frac{2lq^{l-1}[Q(\delta_\varepsilon) - q_0]}{1-q^l} < \infty.
\end{aligned}$$

Also note that

$$\begin{aligned}
\sum_{k=1}^{\infty} \lim_{t\to\infty} r_t(k) &= \frac{2}{1-q^l} \sum_{k=1}^{\infty} (\delta_\varepsilon)^k l q^{l-1} q_k \\
&= \frac{2}{1-q^l} \big[\sum_{k=0}^{\infty} (\delta_\varepsilon)^k l q^{l-1} q_k - (\delta_\varepsilon)^0 l q^{l-1} q_0 \big] \\
&= \frac{2lq^{l-1}}{1-q^l} \big[\sum_{k=0}^{\infty} (\delta_\varepsilon)^k q_k - (\delta_\varepsilon)^0 q_0 \big] \\
&= \frac{2lq^{l-1}}{1-q^l} \big[Q(\delta_\varepsilon) - q_0 \big] \\
&= \lim_{t\to\infty} \sum_{k=1}^{\infty} r_t(k) < \infty.
\end{aligned}$$

By **(***)**, it follows that

$$\lim_{t\to\infty}\sum_{k=1}^{\infty}h_t(k)=\sum_{k=1}^{\infty}\lim_{t\to\infty}h_t(k)=\frac{\sum_{k=1}^{\infty}\varphi(k,\varepsilon)lq^{l-1}q_k}{1-q^l}<\infty. \tag{5.7}$$

Note that

$$\begin{aligned}\lim_{t\to\infty}\sum_{k=1}^{\infty}h_t(k) &= \lim_{t\to\infty}\sum_{k=1}^{\infty}\frac{e^{-u'(q)t}\varphi(k,\varepsilon)P(Z_{N(t)}=k)}{1-P(Z_{N(t)}=0)}\\ &= \lim_{t\to\infty}e^{-u'(q)t}P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}}-m\big|>\varepsilon\,\big|\,Z_{N(t)}>0\big).\end{aligned} \tag{5.8}$$

Therefore, from equations (5.7) and (5.8) above, we conclude that

$$\lim_{t\to\infty}e^{-u'(q)t}P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}}-m\big|>\varepsilon\,\big|\,Z_{N(t)}>0\big)=\frac{\sum_{k=1}^{\infty}\varphi(k,\varepsilon)lq^{l-1}q_k}{1-q^l}<\infty.$$

$\square$

Our large deviation behavior is the large-time asymptotics since we let time t goes to infinite. However, the empirical study by Epps [14] is for short-term stock price as a PRIBP. Therefore, before further evidence of empirical study of long-term stock price as a PRIBP, we would better strict ourself to short-term dynamics of the stock price. Based on this reason, we need to adjust our large-time asymptotics to a small one. A method that allows us to study asymptotics in a short fixed time horizon is to fix a time horizon $[0,t]$ for some small positive t. and then let the intensity λ of the Poisson process goes to infinity. As λ goes to infinity, the number of occurrences of an event goes to infinity in the interval $[0,t]$. This corresponds to infinitely many changes in the stock prices in the time horizon $[0,t]$.

It is not hard to see that proposition 5.1 becomes

$$\lim_{\lambda\to\infty}F_{N,l}(s,t)=q^l \quad \text{for all } 0\le s<1.$$

Proposition 5.2 turns out to be

$$\lim_{\lambda\to\infty}e^{-u'(q)t}[F_{N,l}(s,t)-q^l]=lq^{l-1}Q(s)=lq^{l-1}\sum_{k=0}^{\infty}q_ls^l<\infty \quad \text{for all } 0\le s<1.$$

Proposition 5.3 translates to

$$lq^{l-1}q_k = \begin{cases} \lim_{\lambda\to\infty} e^{-u'(q)t} P(Z_{N(t)} = k \,|\, Z_0 = l) & \text{if} \quad k \geq 1, \\ \lim_{\lambda\to\infty} e^{-u'(q)t} [P(Z_{N(t)} = k \,|\, Z_0 = l) - q^l] & \text{if} \quad k = 0. \end{cases}$$

We then have the following theorem that is analogous to Theorem 5.1. Since the proof is almost identical to that for Theorem 5.1, we are not going to supply it here.

Theorem 5.2 *Assume that* $p_0 \neq 0$. *Let* $Z_0 = l$. *Assume that* $E(exp(\alpha_0 Z_1)) < \infty$ *for some* $\alpha_0 > 0$. *Then for any* $\varepsilon > 0$,

$$\lim_{\lambda\to\infty} e^{-u'(q)t} P\big(\big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\big| > \varepsilon \,\big|\, Z_{N(t)} > 0\big) = \frac{\sum_{k=1}^{\infty} \varphi(k,\varepsilon) l q^{l-1} q_k}{1 - q^l} < \infty,$$

$$\textit{where} \quad \varphi(k,\varepsilon) = P\big(\big|\frac{1}{k}\sum_{i=1}^{k} X_i - m\big| > \varepsilon\big).$$

□

Financial Interpretation of Theorem 5.2:

There is an interesting phenomenon called mean reversion, in the study of financial market. The essence of the concept is that the extreme high or low values for the price of a stock are just temporary and that the price does have a tendency to hang around the long-term average price over time. In the literature, some research also discusses this phenomenon in the sense of the stock returns. For example, auto-correlation of stock returns is a well-known attribute of certain discrete-time stock price models that are often referred to as mean reversion. Some researchers model it as

$$R_t = a(R_{t-1} - \mu) + \mu + \sigma W_t,$$

where R_t is the return of a stock in the period t, μ the mean stock return, σ the volatility, W_t a standard Brownian motion, and $a < 1$ the auto-correlation coefficient which is assumed negative. More precisely, the model with mean reversion can be formulated as

$$\frac{dS_t}{S_t} = \mu dt + \theta \frac{S_t^* - S_t}{S_t} dt + \sigma dW_t,$$

where S_t is the stock price at time t, $S_t^* = S_0 e^{\mu t}$ is the theoretical price, and $\theta > 0$ is the rate which controls the speed of mean reversion.

Now notice that we can rewrite the following probability

$$
\begin{aligned}
& e^{-u'(q)t}P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}}-m\Big|>\varepsilon\,\Big|\,Z_{N(t)}>0\Big) \\
= \;& e^{-u'(q)t}P\Big(\Big|\frac{Z_{N(t)+1}-Z_{N(t)}}{Z_{N(t)}}-(m-1)\Big|>\varepsilon\,\Big|\,Z_{N(t)}>0\Big).
\end{aligned}
$$

In this setting, $\frac{Z_{N(t)+1}-Z_{N(t)}}{Z_{N(t)}}$ is the tick-by-tick rate of return and $m-1$ is viewed as the average tick-by-tick rate of return. Since we take $\lambda \to \infty$, this is a result of "high-frequency stock in a finite time horizon". This is because the Poisson intensity goes to ∞. Therefore, for a fixed time horizon $[0,t]$, where t is not too large, Theorem 5.2 says that the probability that the high-frequency tick-by-tick rate of return deviating from the average rate of return decays at an exponential rate asymptotically. Therefore, Theorem 5.2 suggests a special form of mean reversion in a stock market – short-term mean reversion for high-frequency tick-by-tick rate of return.

5.2 An Application to Physics - Neutron Fluctuations

A model for neutron fluctuation via a branching process has been studied since 1960s. In the literature, neutron fluctuation is often modeled using a continuous-time Markov branching process (see [29]). In the most classical model in this direction, $Z(t)$ presents the number of neutrons at time t in the system. p_0 is the probability of absorption in a reaction process and it includes the event of capture by other material and the case when a fission leads to zero neutron. p_1 is the probability of renewal in a reaction process and it includes the event of scattering and a fission leading to one neutron. p_k, where $k \geq 2$, is the probability of multiplication, i.e., a fission reaction that results in more than one neutrons. Here we would like to apply our large deviation results to study on the fluctuation in neutron count. Therefore, in order to study large deviation behavior of neutron fluctuations, we need to extend our results to a continuous-time Markov branching process.

Recall that in a continuous-time Markov branching process $\{Z(t)\}_{t\geq 0}$, the particles alive have the i.i.d. life-time distribution with a common exponential distribution with some parameter $a > 0$. They have the same offspring distribution $\{p_k\}_{k=0}^{\infty}$ and the numbers of their children are independent of each other and of the whole history. At any given time t, $Z(t)$ represents the number of particles existing in the branching process at time t. A continuous-time Markov branching process is a continuous-time counterpart of a Galton-Watson branching process. They share certain features. For instance, they are both Markovian and the reproduction of particles in both processes are not effected by one another and by the history. Consequently, they both have the branching property, i.e., a process starting from l many ancestors is the same as l independent copies of the process starting from only one ancestor. The difference between a continuous-time Markov branching process and a Galton-Watson branching process is that in the latter, the particles in the same generation give births together at the end of their common one unit of life-time, whereas in the former the particles have their own life-times that are i.i.d. exponentially distributed random variables and hence the particles may give births at different times. Notice that by the branching property, the transition probability $p_{ij}(t)$ satisfies

$$\sum_{j=0}^{\infty} p_{ij}(t)s^j = [\sum_{j=0}^{\infty} p_{1j}(t)s^j]^i.$$

As in a Galton-Watson branching process, the probability generating function plays a key role in a continuous-time Markov branching process. Let

$$f(s) = \sum_{j=0}^{\infty} p_j s^j \quad \text{be the p.g.f. of the offspring distribution.}$$

Define $u(s) := a[f(s) - s]$, where a is the parameter in the life-time exponential distribution. Suppose $Z(0) = 1$. Define

$$F(s,t) := E(S^{Z(t)}|Z(0) = 1) = \sum_{k=0}^{\infty} p_{1k}(t)s^k.$$

Some of the classical preliminary results on this can be found in [4] and [20]. First of all, the extinction probability is q, the smallest root of $f(s) = s$ in $[0,1]$. Secondly, the first moment of $Z(t)$ is $E(Z(t)) = e^{u'(1)t} = e^{a(m-1)t}$. Thirdly, $F(s,t)$ converges to q as $t \to \infty$ for all $0 \le s < 1$. Finally, under the assumption that $m \ne 1$, there exist real numbers $\{a_k\}_{k=0}^{\infty}$ such that $\lim_{t\to\infty} e^{-u'(q)t}[F(s,t) - q] = \sum_{k=0}^{\infty} a_k s^k = A(s) < \infty$ for all $0 \le s < 1$. Notice that some of the properties about a continuous-time Markov branching process are quite similar to those on a Galton-Watson branching process.

Since a continuous-time Markov branching process has the branching property, the results mentioned above can be naturally extended to the situation when $Z(0) = l$, where i is arbitrary positive integer. Define

$$F_l(s,t) := E(S^{Z(t)}|Z(0) = l) = \sum_{k=0}^{\infty} p_{lk}(t)s^k.$$

It is clear that $F_l(s,t) = [F(s,t)]^l$ from independence. The extinction probability is now q^l and $F_l(s,t)$ converges to q^l as $t \to \infty$ for all $0 \leq s < 1$. The expected number of particles at time t is now $E(Z(t)) = le^{u'(1)t} = le^{a(m-1)t}$. Also, under the assumption that $m \neq 1$, now $\lim_{t\to\infty} e^{-u'(q)t}[F_l(s,t) - q^l] = lq^{l-1}\sum_{k=0}^{\infty} a_k s^k = lq^{l-1}A(s) < \infty$ for all $0 \leq s < 1$. The following result is necessary in proving our large deviation result stated later. Since the proof is essentially identical to the proof for the case of the PRIBP in the previous chapter, we only state the result.

Proposition 5.4 *Assume that $m \neq 1$. Let $Z_0 = l$. If $p_0 \neq 0$, then*

$$lq^{l-1}a_k = \begin{cases} \lim_{t\to\infty} e^{-u'(q)t} P(Z(t) = k \,|\, Z(0) = l) & \text{if} \quad k \geq 1, \\ \lim_{t\to\infty} e^{-u'(q)t} [P(Z(t) = k \,|\, Z(0) = l) - q^l] & \text{if} \quad k = 0. \end{cases}$$

□

Notice that for any $v \geq 0$,

$$Z(t+v) = \begin{cases} \sum_{i=1}^{Z(t)} Z_t^{(i)}(v) & \text{if} \quad Z(t) > 0, \\ 0 & \text{if} \quad Z(t) = 0, \end{cases}$$

where $Z_t^{(i)}(v)$ is the number of particles alive at time $t+v$ that are offspring of the i-th particle existing at time t. Note that $Z_t^{(i)}(v)$ for $i = 1, 2, \cdots, Z(t)$ are i.i.d. with a common probability law, $\{p_{1k}(v)\}_{k=0}^{\infty}$, i.e. $Z_t^{(i)}(v)$ are i.i.d. copies of $Z(v)$. The next theorem shows a large deviation behavior for the probability that the ratio of population sizes between any two time points deviating away from the corresponding expectation. Since the proof is not too different from that for the case of the PRIBP, we state it without supply the proof.

Theorem 5.3 *Assume that $p_0 \neq 0$. Let $Z(0) = l$. Assume that $E(exp(\alpha_0 Z(v))) < \infty$ for some $\alpha_0 > 0$ and $v \geq 0$. Then for any $\varepsilon > 0$,*

$$\lim_{t\to\infty} e^{-u'(q)t} P\big(\big|\frac{Z(t+v)}{Z(t)} - e^{a(m-1)v}\big| > \varepsilon \,\big|\, Z(t) > 0\big) = \frac{\sum_{k=1}^{\infty} \varphi_v(k,\varepsilon) lq^{l-1} a_k}{1 - q^l} < \infty,$$

$$\textit{where} \quad \varphi_v(k,\varepsilon) = P\big(\big|\frac{1}{k}\sum_{i=1}^{k} Z_t^{(i)}(v) - e^{a(m-1)v}\big| > \varepsilon\big).$$

□

In a nuclear chain reaction, the number of neutrons is a main concern because it provides information about how far the reaction has progressed. For example, it is especially useful to know the joint distribution of the number of neutrons at different time points and the auto-correlation function for the number of neutrons. In the study of neutron fluctuation as a branching process, the extinction and survival probabilities and the expectation and variance of the number of neutrons have received intensive attention under various setting of the models they use. Therefore, the ratio of the number of neutrons at different time points holds the most essential and relevant information in asymptotic sense. Here the result of Theorem 5.3 can explain the behavior of the ratio of the number of neutrons at two arbitrary time instants. It is known that since the typical neutron's lifetime in the reactor is as short as on the order of 10^{-7} to 10^{-4} seconds, the large-time asymptotic behavior is quite suitable under the time scales of the reaction in real time. Therefore, asymptotic results are useful, for example, when we need to control the growth rate of the neutrons in nuclear fission between some specific time interval due to security issues and technical aspects. In this regard, Theorem 5.3 says that, for instance, if we want to avoid the ratio between a given time interval being larger than, say k, then by solving $e^{a(m-1)v} = k$ for v, where the parameters a and m are known from various estimates based on experiments, we obtain the estimated time interval that we need to check or to apply some actions to the system periodically.

Chapter 6

Conclusion and Future Research

In the literature, to the best knowledge of the author, large deviation aspects of a RIBP has never been formally investigated. In this dissertation, we have studied various large deviation rates and other related issues concerning the ratio of successive numbers of generations for a PRIBP. We investigate not only the case in which the probability of giving no birth is zero but also extend to the case which allows non-trivial probability for no birth. Under various moment conditions, the decay rates are exponential for all conditional and unconditional cases although the limiting probabilities are slightly different. Besides, our results are applied to finance and physics. The application to finance relates to the asymptotics for high-frequency stock return in a short time horizon. The results applied to the physics direction suggest a possible scenario that may be useful in the estimation of neutron fluctuation control problems.

6.1 Theoretical Research

There are some direct extensions of a PRIBP that are worth considering. First of all, instead of the large deviation behavior of the ratio of consecutive generation sizes $Z_{N(t)+1}/Z_{N(t)}$, we can study that of other meaningful quantities. For example, the tail behavior of W' in both right extreme and right extreme, i.e, large deviations for

$$P(W' \geq x) \quad \text{as } x \to \infty \quad \text{and} \quad P(W' \leq x) \quad \text{as } x \to 0.$$

Another one is about a large deviation behavior of normalized random sums defined by

$$R_{N(t)} := \frac{S_{Z_{N(t)}}}{Z_{N(t)}} = \frac{\sum_{i=1}^{Z_{N(t)}} X_i}{Z_{N(t)}},$$

where $\{X_n\}_{n=1}^{\infty}$ is a sequence of i.i.d. real-valued random variables with zero mean. Notice that on the event $\{Z_{N(t)} > 0\}$, the random variable $R_{N(t)}$ is well-defined. $R_{N(t)}$ has also found many applications, for example, it arise in applications of polymerase chain reactions with mutations. Inspired by [17], we are curious about the large deviations of $R_{N(t)}$, especially, the large deviation behavior of the probability

$$P(R_{N(t)} \geq \varepsilon_t \,|\, Z_{N(t)} > 0) \quad as\ t \to \infty,$$

for some sequence $\{\varepsilon_t\}_{t\geq 0}$ that converges to 0. It is worth mentioning that if X_1 coincides in law with $Z_1 - m$, then $R_{N(t)}$ coincides in law with $\frac{Z_{N(t)+1}}{Z_{N(t)}} - m$. Also notice that if $\{\varepsilon_t\}_{t\geq 0} := \{\varepsilon\}$, i.e., independent of time and if $R_{N(t)}$ coincides in law with $\frac{Z_{N(t)+1}}{Z_{N(t)}} - m$, then it is exactly the case studied in this dissertation. That is, the study of $P(R_{N(t)} \geq \varepsilon_t \,|\, Z_{N(t)} > 0)$ is a generalization of the large deviation study about the ratio away from its mean for our PRIBP. Another possible large deviation study regarding $R_{N(t)}$ is the asymptotic behavior of

$$P(R_{N(t)} \in \cdot \,|\, Z_{N(t)} > v_t) \quad as\ t \to \infty,$$

for some sequence $\{v_t\}_{t\geq 0}$ that goes to ∞.

Secondly, in fact, RIBP includes PRIBP as a special case in which the indexing process is a Poisson process. Thus, we could consider a more general renewal process as the subordinator process and then investigate the large deviation behavior of the ratio $Z_{N(t)+1}/Z_{N(t)}$ and that of other quantities mentioned above. In the study of RIBP, [27] and [26] have focused on some critical RIBPs and have studied the asymptotic formulas for the moments, $Var(Z_{N(t)})$, that for the probability of non-extinction $P(Z_{N(t)} > 0)$, and limiting distributions of properly normalized $Z_{N(t)}$. All these results are restricted in the critical case, therefore, more thorough and general results for non-critical cases are necessary to provide a complete picture mathematically.

Thirdly, we can consider a PRIBP with immigration. Let

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_i + Y_{n+1} \quad \text{with } Z_0 = 1,$$

where $\{X_i\}$ are i.i.d. with a p.g.f. $f(s)$ and $\{Y_n\}$ are i.i.d. with a p.g.f. $h(s)$ and X's and Y's are independent. That is, at the birth time of the n-th generation, there is an immigration of Y_n particles into the system. Suppose that $E(X_1) = m$ and $E(Y_1) = \mu$. Notice that the expectation $E(Z_1) = m + \mu$. For a super-critical branching process, Z_{n+1}/Z_n converges to m a.s. as $n \to \infty$ when conditioning on the event $\{Z_{N(t)} > 0\}$. Therefore, we can study the large deviation rates for the probability

$$P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m\Big| > \varepsilon \,\Big|\, Z_{N(t)} > 0\Big).$$

Fourthly, we can consider a PRIBP in a random environment. Let $Z_0 = 1$. Assume that all Z_n particles in the n-th generation reproduce according to a common offspring distribution with p.g.f. $f_{\xi_n}(s)$, where $f_{\xi_n}(s)$ is chosen randomly from a collection Φ of environmental p.g.f.s according to some law. Assume that $\{f_{\xi_n}(s)\}$ are i.i.d.., i.e., i.i.d. environment. Let $m(\xi_\cdot) = \sum_{i=1}^{\infty}(\sum_{k=1}^{\infty} kp_k(\xi_i))p_i$ where $p_k(\xi_i)$ is the probability of having k children under the offspring distribution $f_{\xi_i}(s)$ and p_i is the probability of the offspring distribution of which p.g.f. is $f_{\xi_i}(s)$. We can study the large deviation rate for

$$P\Big(\Big|\frac{Z_{N(t)+1}}{Z_{N(t)}} - m(\xi_\cdot)\Big| > \varepsilon \,\Big|\, Z_{N(t)} > 0\Big).$$

Fifthly, we can consider a multi-type super-critical PRIBP inspired by the work of Athreya and Vidyashankar in [6]. Let $\{Z_n\}_{n=0}^{\infty}$ be a super-critical 2-type Galton-Watson branching process with type-dependent offspring generating functions $f^{(i)}(s)$, $i = 1, 2$, and the mean matrix $M = ((\frac{\partial f^{(i)}(s)}{\partial s_j}|_{s=(1,1)})_{ij})_{2\times 2}$. Assume that $f(0,0) = 0$, i.e., the probability of having no child is zero. Let ρ be the maximal eigenvalue of M with the corresponding left and right eigenvectors $v^{(1)}$ and $u^{(1)}$, respectively. Let $l = (l_1, l_2)$ be any non-zero vector with $l_1 \neq l_2$. Under some assumptions, we would like to study the large deviation rates for

$$P\Big(\Big|\frac{l \cdot Z_{N(t)+1}}{\mathbf{1} \cdot Z_{N(t)}} - \frac{l \cdot (Z_{N(t)}M)}{\mathbf{1} \cdot Z_{N(t)}}\Big| > \varepsilon \,\Big|\, Z_0 = e_i\Big)$$

and

$$P\Big(\Big|\frac{l \cdot Z_{N(t)}}{\mathbf{1} \cdot Z_{N(t)}} - \frac{l \cdot v^{(1)}}{\mathbf{1} \cdot v^{(1)}}\Big| > \varepsilon \,\Big|\, Z_0 = e_i\Big).$$

6.2 Applied Research

For financial applications, we consider the pricing of exotic options when the price of the underlying asset is modeled by a PRIBP. For example, we can study the pricing of lookback option. It is path dependent and the payoff depends on the maximum or minimum underlying asset's prices over the life of the option. The holder of this option is allowed to look back over the lifetime of the option to determine the payoff. We consider the lookback option with a fixed strike. The payoff functions for the fixed strike lookback call and the fixed strike lookback put are, respectively, given by:

$$LC_{fix}(T) = max(S_{max} - K, 0) \quad \text{and} \quad LP_{fix} = max(K - S_{min}, 0),$$

where K is the the fixed strike price, S_{max} is the maximum underlying asset's price during the lifetime of the option, and S_{min} is the minimum underlying asset's price during the lifetime of the option.

We would like to derive analytic formulas for pricing this call and put options. Applying martingale pricing, the prices of the call and put options at the current moment are

$$LC_{fix}(0) = e^{-rT}E[LC_{fix}(T)] = e^{-rT}E[max(S_{max} - K, 0)]$$

and

$$LP_{fix}(0) = e^{-rT}E[LP_{fix}(T)] = e^{-rT}E[max(K - S_{min}, 0)].$$

Therefore, we need to derive the exact formulas for $LC_{fix}(0)$ and $LP_{fix}(0)$ in terms of PRIBP. Notice that in order to obtain the exact formulas, we need to calculate the probabilities of maximum and minimum population size up to each generation. After obtaining the formulas, we would like to give numerical results to compare the prices of the options whose underlying asset is modeled by a PRIBP with the prices of those whose underlying asset is modeled by a lognormal distribution.

REFERENCES

[1] Athreya, K. B. *Large deviation rates for branching processes. I. Single type case.* (1994) Ann. Appl. Probab. 4, no. 3, 779–790.

[2] Athreya, K. B., Karlin, S. *On branching processes in random environments: I. Extinction probability.* (1971) AMS 42, 1449–1520.

[3] Athreya, K. B., Karlin, S. *On branching processes in random environments: II. Limit theorems.* (1971) AMS 42, 1843–1858.

[4] Athreya, K. B., Ney, P. E. *Branching processes.* (1972) Springer, Berlin.

[5] Athreya, K. B., Vidyashankar, A. N. *Large deviation rates for supercritical and critical branching processes.* (1997) In Classical and modern branching processes, 1–18, IMA Vol. Math. Appl., 84, Springer, New York.

[6] Athreya, K. B., Vidyashankar, A. N. *Large deviation rates for branching processes. II. The multitype case.* (1995) Ann. Appl. Probab. 5, no. 2, 566–576.

[7] Athreya, K. B., Vidyashankar, A. N. *Large deviation results for branching processes.* (1993) In Stochastic Processes: A Festschrift in Honour of Gopinath Kallianpur, Springer, New York.

[8] Boiko, R. V. *Limit theorems for a branching process with a varying environment which describes the development of a population in a limiting environment.* (1982) Ukrainskyj Matematychnyj Zhurnal 34, 681–687.

[9] Cont, R., Tankov, P. *Financial modelling with jump processes.* (2004) Chapman and Hall/CRC

[10] Chow, Y. S., Teicher, H. *Probability Theory. Independence, Interchangeability, Martingales.* (1988) 2nd ed. Springer, New York.

[11] Daley, D. J. *Extinction conditions for certain bisexual Galton-Watson branching processes.* (1968) ZW 9, 315–322.

[12] Dembo, A., Zeitouni, O. *Large deviations techniques and applications.* (1998) 2nd ed. Springer, New York.

[13] Dion, J. P., Epps, T. W. *Stock prices as branching processes in random environments: estimation.* (1999) Comm. Statist. Simulation Comput. 28, no. 4, 957–975.

[14] Epps, T. W. *Stock prices as branching processes.* (1996) Stochastic Models 12, no. 4, 529–558.

[15] Embrechts, P., Kluppelberg, C., Mikosch, T. *Modelling extremal events for insurance and finance.* (1997) Springer, Berlin.

[16] Feng, J., Forde, M., Fouque, J.-P. *Short maturity asymptotics for a fast mean reverting Heston stochastic volatility model.* (2010) SIAM Journal on Financial Mathematics, Vol. 1, 126–141.

[17] Fleischmann, K., Wachtel, V. *Large deviations for sums indexed by the generations of a GaltonWatson process.* (2008) Probab. Theory Relat. Fields, 141, 445–470.

[18] Galton, F. *problem 4001.* (1873) Educational Times 1 April, 17.

[19] Haccou, P., Jagers, P., Vatutin, A. *Branching processes: variation, growth, and extinction of populations.* (2005) Cambridge University Press, Cambridge.

[20] den Hollander, F. *Large deviations.* (2000) Fields Institute Monographs, 14. American Mathematical Society, Providence.

[21] Jirina, M. *Stochastic branching processes with continuous statespace.* (1958) Czechosolovak. Math. J. 8, 292–313.

[22] Klebaner, F. C. *Population-size-dependent branching process with linear rate of growth.* (1983) J. of Applied Probability 20, 242–250.

[23] Kingman, J. F. C. *Ergodic properties of continuous-time Markov processes and their discrete skeletons.* (1963) Proc. London Math. Soc. (3) 13, 593–604.

[24] Kolmogorov, A., Dmitriev, N. *Branching stochastic processes.* (1947) Dokl. Acad. Nauk. SSSR 56, 5–8.

[25] Mitov, G. K., Mitov, K. V. *Option pricing by branching process.* (2007) Pliska Stud. Math. Bulgar. 18, 213–224.

[26] Mitov, G. K., Mitov, K. V., Yanev, N. M. *Limit theorems for critical randomly indexed branching processes.* (2010) In Workshop on Branching Processes and their Applications, 95–108, Lect. Notes Stat. Proc., 197, Springer, Berlin.

[27] Mitov, G. K., Mitov, K. V., Yanev, N. M. *Critical randomly indexed branching processes.* (2009) Statist. Probab. Lett. 79, no. 13, 1512–1521.

[28] Mitov, G. K., Rachev, S. T., Kim, Y. S., Fabozzi, F. J. *Barrier option pricing by branching processes.* (2009) Int. J. Theor. Appl. Finance 12, no. 7, 1055–1073.

[29] Pal, L., Pazsit, I. *Neutron Fluctuations: A Treatise on the Physics of Branching Processes.* (2008) Elsevier Science, Landon.

[30] Royden, H. L. *Real analysis.* (1988) 3rd ed. Macmillan, New York.

[31] Sevastyanov B. A. *The theory of branching random processes.* (1951) Uspehi Mat. Nauk 6, 47–99.

[32] Smith, W. L., Wilkinson, W. E. *On branching processes in random environments.* (1969) AMS 40, 814–827.

[33] Tankov, P., Voltchkova, E. *Jump-diffusion models: a practitioner's guide.* (2009) Banque et Marchs, No. 99, March-April 2009.

[34] Watson, H. W. *Solution to problem 4001.* (1873) Educational Times, 1. aUG., 115–116.

[35] Watson, H. W., Galton, F. *On the probability of the extension of families.* (1874) J. Anthropol. Inst. Great B. and Ireland 4, 138–144.

[36] Williams, T. A. *Option pricing and branching processes.* (2001) Ph.D. Thesis, University of Virginia.