Branching Processes:
Optimization, Variational Characterization,
and Continuous Approximation

Von der Fakultät für Mathematik und Informatik

der Universität Leipzig

angenommene

D I S S E R T A T I O N

zur Erlangung des akademischen Grades

DOCTOR RERUM NATURALIUM

(Dr.rer.nat.)

im Fachgebiet

Mathematik

vorgelegt

von B.Sc. Ying Wang

geboren am 09.04.1981 in Shandong (Volksrepublik China)

Die Annahme der Dissertation haben empfohlen:

1. Professor Dr. Jürgen Jost (MPIMN Leipzig)
2. Professor Dr. Wolfgang König (WIAS Berlin)

Die Verleihung des akademischen Grades erfolgt mit Bestehen der Verteidigung am

27.10.2010 mit dem Gesamtprädikat magna cum laude.

Acknowledgements

Contents

Chapter 1

Introduction

In this thesis, we use multitype Galton-Watson branching processes in random environments as individual-based models for the evolution of structured populations with both demographic stochasticity and environmental stochasticity, and investigate the phenotype allocation problem. We explore a variational characterization for the stochastic evolution of a structured population modeled by a multitype Galton-Watson branching process. When the population under consideration is large and the time scale is fast, we deduce the continuous approximation for multitype Markov branching processes in random environments.

1.1 Optimization and the allocation problem

Optimization, this beguilingly simple idea, allows biologists not only to understand current adaptations, but also to predict new designs that may yet evolve. The essence of optimization is to calculate the most efficient solution to a given problem, and then to test the prediction. The concept has already revolutionized some aspects of biology, but it has the potential for much wider application [Sut05].

Many problems in evolutionary biology involve the allocation of some limited resource among several investments [JP08]. It is often of interest to know whether, and how, allocation strategies can be optimized for the evolution of a structured population with randomness. In our work, the investments represent different types of

offspring, or alternative strategies for allocations to offspring. As payoffs we consider the long-term growth rate, the expected number of descendants with some future discount factor, the extinction probability of the lineage, or the expected survival time. Two different kinds of population randomness are considered: demographic stochasticity and environmental stochasticity.

The reasons for this particular choice of payoffs are the following. Darwin's theory of natural selection tells us that the action of natural selection results in individuals that are good at surviving and reproducing. So in a number of important biological scenarios the optimization criterions are the maximization of the long-term growth rate and the minimization of the extinction probability of the lineage. In financial mathematics, optimization criterions like the maximization of the expected total returns with some future discount factor and the maximization of expected stopping times are investigated a lot. These criterions are also relevant in evolutionary biology.

Demographic stochasticity is caused by chance realizations of individual probabilities of death and reproduction. We model the evolution of a structured population with demographic stochasticity by a multitype Galton-Watson branching process. The Galton-Watson process can be viewed as a real-time process, provided all the individuals can be assumed to have the same life span of length one (year or season). In such cases, we assume that individuals are born at the beginning of the season. The phenotype of an individual is denoted as its type. We assume that the type space has finitely many states and assume the number of offspring is finite. The theory of branching processes provides characterizations of the payoffs which are the growth rate conditioned on non-extinction, the extinction probability, the expected survival time conditioned on extinction and the expected number of descendants conditioned on extinction. Since it is very difficult to calculate explicitly the optimal offspring distributions in the general framework, we come up with a simple model I, see subsection 2.1.1, and solve the optimal allocation strategies w.r.t. the above payoff functions.

Another source of randomness is environmental stochasticity that affects the population as a whole. Models for the evolution of a population with environmental stochasticity have been much explored in the literature. We note only a small se-

6

lection that is directly relevant to the work in this thesis. Optimal strategies that maximize the long-term growth rate in various models with environmental stochasticity can be found in a survey paper of McNamara, Houston and Collins [MHC01] and the references therein. In an attempt to explain the variability of the clutch size among laying birds of the same species, Grey [Gre80] considered models in which birds have to choose a randomized strategy for the clutch size in the face of random environments in order to minimize the probability of extinction. Lande and Orzack [LO88] studied an age-structured population in a fluctuating environment and derived analytical formulae for the distribution of the time to extinction and the probability of extinction based on a diffusion approximation for the logarithm of total population size.

Analogous to the simple model I, we propose a simple model II, see subsection 2.2.1, which describes the evolution of a structured population with environmental stochasticity. In the fluctuating model, the payoffs are the long-term growth rate, the extinction probability, the expected survival time conditioned on extinction, and the discounted expected number of descendants. We solve the optimal allocation strategies w.r.t. the above payoff functions. In particular, the Bellman scheme in control theory is applied to obtain the optimal strategy that maximizes the discounted expected number of descendants.

Finally, we come to consider the evolution of a structured population with both demographic stochasticity and environmental stochasticity. The population is modeled as a multitype Galton-Watson branching process in a random environment. We propose a simple model III, see subsection 2.3.1, which is a combination of the simple models I and II. By the convergence theorem of branching processes in random environments the long-term growth rate of the population is characterized by the largest Lyapunov exponent. The random Perron-Frobenius theorem provides an alternative expression for the largest Lyapunov exponent. We use it to deduce an explicit formula for the largest Lyapunov exponent of the simple model III. We approximate the extinction probability by using a similar method developed by Wilkinson. For the simple model III, we obtain the optimal strategy that maximizes the largest Lyapunov

7

exponent and minimizes the approximate extinction probability respectively.

Recently, Dombry, Mazza and Bansaye [DMB09] also used multitype branching processes in random environments to model population growth. They studied the optimal type allocation strategy that maximizes the net growth rate of the population. We did our work independently. The subtle differences between the characterizations of their net growth rate and our long-term growth rate are analyzed in subsection 2.3.2.

1.2 Variational characterization

Evolution is often understood as an optimization problem, and there is a long tradition to look at evolutionary models from a variational perspective. For example, in deterministic population dynamics defined via the Leslie matrix, the equilibrium state of the dynamical system is characterized by using a variational principle [Dem83][AGD94][Dem97]; a variational principle in mutation-selection models is deduced in [BG07]. The purpose of our work is to deduce a variational characterization for the stochastic evolution of a structured population modeled by a multitype Galton-Watson branching process.

By the Kesten-Stigum theorem ([KS66],[AN72]) and Theorem 2 of [KLP97], we describe the asymptotic properties of the multitype Galton-Watson branching process forward in time. The asymptotic properties are determined by the principle eigenvalue ρ of the mean matrix M, the associated right eigenvector u and left eigenvector v. That is, $\log \rho$ is equal to the growth rate of the population size, u is a measure of the relative contribution made to the stationary population by each type, and v is the stationary composition of the population.

Since the Leslie matrices of classic demography are nothing but the mean matrices of special multitype branching processes, we deduce a similar variational characterization for the asymptotic growth rate $\log \rho$ of the multitype Galton-Watson branching process via suitably defined entropy and reproductive potential.

In particular, the so-called retrospective process plays an important role in the description of the equilibrium state used in the variational characterization. In our

8

work, we define the retrospective process associated with a multitype Galton-Watson branching process and identify it with the mutation process describing the type evolution along typical lineages of the multitype Galton-Watson branching process. In the mathematical literature, the retrospective process was firstly introduced by Jagers ([Jag89], p.195) for a general branching process and a version for multitype Markov branching process was defined in [GB03]. The identification result for the retrospective process associated with a multitype Markov branching process in continuous time was proved by Georgii and Baake ([GB03], Theorem3.3).

1.3   Continuous approximation

Continuous approximation of branching processes is of both practical and theoretical interest. As was shown in ([HJV05], Chap. 4.1), approximations of stochastic models can be useful for several reasons: The approximating model may be simpler than the original model; The qualitative behavior of the approximating model may be easier to understand; The number of unknown parameters (i.e., parameters that need to be estimated from data) may be smaller for the approximating model; The approximating model may be computationally more tractable. Operator semigroups, martingale problems, and stochastic equations provide approaches to the characterization of branching processes, and to each of these approaches correspond methods for proving convergence results.

A lot of research work has been done on approximation of one-dimensional branching processes. The diffusion approximation of Galton-Watson branching processes in the near-critical case was first formulated by Feller [Fel50] and made rigorous by Jiřina [Jir69]. For branching processes in random environments, Keiding [Kei75] formulated a diffusion approximation for a Galton-Watson process in a random environment that was made rigorous by Helland [Hel81] in the case of independent environments for each generation. Kurtz in [Kur78] used a semigroup approach to establish such an approximation under more general assumptions on the environments. Borovkov [Bor02] extended the Kurtz result so that the range of the initial values of the processes can be much wider, moment conditions are more general, and the approximant can be

a discontinuous process. Diffusion approximation for multitype branching processes was exploited by Joffe and Metivier [JM86]. However, there is no literature on approximation of multitype branching processes in random environments.

Inspired by the work in ([EK86], Chap. 9.3), which deduced a continuous approximation for a sequence of one-dimensional Markov branching processes in random environments, we want to obtain a continuous approximation for a sequence of multi-dimensional Markov branching processes in random environments. The former convergence result is based on the technique of random time changes. However, it does not generalize to vector-valued processes in random environments. In our work, we firstly construct a multitype Markov branching processes in random environments rigorously. Then for the parallel mutation-selection Markov branching processes in random environments, we analyze the instability property. Finally, we make use of weak convergence theorem of stochastic differential equations by Kurtz and Protter [KP91] and give conditions under which continuous approximation for a sequence of two-type parallel mutation-selection Markov branching processes in random environments can be obtained. In fact, continuous approximation for a sequence of d-type $(d \geq 2)$ parallel mutation-selection Markov branching processes in random environments can be deduced without essential difficulties. However, our result for one-dimensional case is not as strong as that in ([EK86], Chap. 9.3).

1.4 Overview

In chapter 2, we solve the phenotype-allocation problem by means of three stochastic population models depending on different kinds of population randomness. In section 2.1, we model the evolution of a structured population with demographic stochasticity by a multitype Galton-Watson branching process and propose a simple model I for the phenotype-allocation problem. The payoffs are characterized by the theory of branching processes. We explore the optimal strategies w.r.t the payoffs in the subsequent subsections. In section 2.2, we study a simple model II for the phenotype-allocation problem for the evolution in the presence of environmental stochasticity. In section 2.3, we model the evolution of a structured population with both demo-

graphic stochasticity and environmental stochasticity by a multitype Galton-Watson branching process in a random environment. In subsection 2.3.1, we come up with the simple model III. We deduce an explicit formulae of the long-term growth rate in the simple model III and obtain the optimal strategy that maximizes the growth rate in subsection 2.3.2. We approximate the extinction probability in the simple model III by a method similar to the one of Wilkinson and obtain the optimal strategy that minimizes the approximate extinction probability for an concrete example in subsection 2.3.3.

In Chapter 3 we develop a variational principle for the models that are described by multitype Galton-Watson branching processes. We take the point of view of individual lineages through time, their ancestry and genealogy. In section 3.1, we show that the asymptotic properties of the branching process forward in time are, to a large extent, determined by the principle eigenvalue ρ of the mean matrix M, the associated right eigenvector u and left eigenvector v . In section 3.2, we define the retrospective process and identify it with the mutation process describing the type evolution along typical lineages of the branching process. In section 3.3, we deduce that the asymptotic growth rate $\log \rho$ of the branching process satisfies a variational principle and show that the distribution of the retrospective process is the unique equilibrium state for the variational characterization.

In chapter 4, instead of discrete time branching processes, we study continuous time multitype Markov branching processes in random environments. In section 4.1, we give a rigorous construction of multitype Markov branching processes in random environments by employing the technique used in ([EK86], Chap. 6.4). In section 4.2, we deduce the Kolmogorov equations and the mean matrix for the branching process conditioned on environment. In section 4.3, we analyze the instability property of a parallel mutation-selection Markov branching process in a random environment. In section 4.4, we deduce a weak convergence result for a sequence of parallel branching processes in random environments. In section 4.5, we give examples for applications.

Chapter 2

Stochastic population models and the allocation problem

Many problems in evolutionary biology involve the allocation of some limited resource among several investments [JP08]. It is often of interest to know whether, and how, allocation strategies can be optimized. Here, the investments represent different types of offspring, or alternative strategies for allocations to offspring. As payoffs we consider the long-term growth rate, the expected number of descendants with some future discount factor, the extinction probability of the lineage, or the expected survival time.

We solve the allocation problem in three stochastic population models depending on different kinds of population randomness. Population randomness through individual variability is called demographic stochasticity. Another source of randomness is environmental stochasticity that affects the population as a whole. Before we study the full model, in which both demographic stochasticity and environmental stochasticity are included, we discuss the classical models used for studying evolution in the presence of demographic and environmental stochasticity, respectively.

2.1 Multitype Galton-Watson branching processes

Demographic stochasticity is caused by chance realizations of individual probabilities of death and reproduction in a finite population. The classical model for studying the evolution of a structured population with demographic stochasticity is the multitype Galton-Watson branching process [HJV05].

Let the type space be given by $S = \{1, \cdots, d\}$. The multitype Galton-Watson branching process is described as follows [KA02]: a single ancestor individual of i-type lives for exactly one unit of time, and at the moment of death it produces a random number of offspring according to a prescribed probability distribution p^i. Each of the first-generation offspring behaves, independently of each other, as the initial particle did, and so forth. Then $\{\mathbf{Z}_n = (Z_n^1, \cdots, Z_n^d)^\top\}$ is a multitype branching process defined on a probability space $(\Omega, \mathcal{F}, P)$ where Z_n^i stands for the number of type i individuals in generation n with $i = 1, \cdots, d$. Here Ω is the space of trees (for the construction see Chapter VI of [Har63]), $\mathcal{F}$ is generated by the cylinder sets of Ω, and P is the probability measure on $(\Omega, \mathcal{F})$. In particular, P^j denotes the probability measure on $(\Omega, \mathcal{F})$ when the process is initiated with $\mathbf{Z}_0 = e_j$, where e_j is the column vector with jth component equal to 1 and else 0. The expectation operator for P^j will be denoted by E^j.

In this section, we firstly come up with a simple model I of a structured population of annual organisms with demographic stochasticity. Then we define the growth rate conditioned on survival and give an analytical representation for it based on the Kesten-Stigum theorem in subsection 2.1.2. In subsection 2.1.3, the extinction probability is characterized by the theory of branching processes. In subsection 2.1.4, we show that a multitype supercritical Galton-Watson branching process conditioned on extinction is a subcritical branching process and we characterize the expected survival time conditioned on extinction. We study the expected number of descendants conditioned on extinction in subsection 2.1.5. Furthermore, we solve the allocation problem w.r.t. the above mentioned payoff functions in the simple model I.

2.1.1 Simple model I

Consider the following simple model of a structured population of annual organisms with demographic stochasticity. There are 2 possible phenotypes for each organism. The phenotype is chosen at the beginning of the development of the organism. An organism develops into phenotype 1 with probability x and phenotype 2 with probability $1 - x$, where $0 < x < 1$. An organism of phenotype i will die with probability d_i leaving no offspring and survive with probability $1 - d_i$ producing c_i offspring, where $0 < d_i < 1$ and c_i are positive integers.

We model the evolution of the population by using a multitype Galton-Watson branching process. Here the type space is given by $S = \{1, 2\}$. The offspring distributions $p^i(n_1, n_2)$ with n_i denoting the number of offspring of type i for $i \in S$ are as follows:

$$
p^1(n_1, n_2) = \begin{cases} d_1, & n_1 = n_2 = 0; \\ (1 - d_1) \begin{pmatrix} c_1 \\ n_1 \end{pmatrix} x^{n_1} (1 - x)^{n_2}, & n_1 + n_2 = c_1; \\ 0, & \text{else.} \end{cases} \tag{2.1.1}
$$

and

$$
p^2(n_1, n_2) = \begin{cases} d_2, & n_1 = n_2 = 0; \\ (1 - d_2) \begin{pmatrix} c_2 \\ n_1 \end{pmatrix} x^{n_1} (1 - x)^{n_2}, & n_1 + n_2 = c_2; \\ 0, & \text{else.} \end{cases} \tag{2.1.2}
$$

There are two one-dimensional extreme cases of the above model corresponding to $x = 0$ and $x = 1$, respectively. The former models the evolution of a non-structured population composed of organisms of phenotype 1 with offspring distribution

$$
p_1(n) = \begin{cases} d_1, & n = 0; \\ 1 - d_1, & n = c_1. \end{cases} \tag{2.1.3}
$$

The latter models the evolution of a non-structured population composed of organisms

of phenotype 2 with offspring distribution

$$p_2(n) = \begin{cases} d_2, & n = 0; \\ 1 - d_2, & n = c_2. \end{cases} \tag{2.1.4}$$

Note that we use upper indices for offspring distributions p^i of a multitype branching process while we use lower indices for offspring distributions p_i of the ith-type one dimensional branching processes.

2.1.2   The growth rate

The growth rate is defined as

$$\lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1, \text{ a.s. conditioned on non-extinction,} \tag{2.1.5}$$

where $\|\mathbf{Z}_n\|_1 = \sum_{i \in S} Z_n^i$ and non-extinction means we only consider those $\omega \in \Omega$ such that $\lim_{n \to \infty} \|\mathbf{Z}_n(\omega)\|_1 = \infty$. The theory of branching processes shows how to give an analytical representation of the growth rate. Here are some definitions and notations. The branching process $\{\mathbf{Z}_n\}_{n \geq 0}$ is said to be a singular process if each individual has exactly one offspring, otherwise it is nonsingular. Define $M = (m_{ij})_{i,j \in S}$ to be the matrix of expected numbers of offspring. Specially, $m_{ij} = E^i[Z_1^j]$ is the expected number of progeny of type j produced from an individual of type i. The branching process $\{\mathbf{Z}_n\}_{n \geq 0}$ is called positively regular if the mean matrix M is strictly positive, i.e. there is an N such that $m_{ij}^{(N)} > 0$ for all $i, j \in S$. The branching process is supercritical if $\rho > 1$, where ρ is the principle eigenvalue of M.

Theorem 2.1.1. *Assume that $\{\mathbf{Z}_n\}_{n \geq 0}$ is nonsingular, positively regular and supercritical. Assume that*

$$E^i[\mathbf{Z}_1^j \log \mathbf{Z}_1^j] < \infty \text{ for all } i, j \in S. \tag{2.1.6}$$

Then

$$\lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1 = \log \rho, \text{ a.s. conditioned on non-extinction.} \tag{2.1.7}$$

16

Proof. Firstly, the Perron-Frobenius theory tells us that M has a principle eigenvalue ρ and associated positive right and left eigenvectors u and v which will be normalized so that $\langle u, v \rangle = 1$ and $\langle v, \mathbf{1} \rangle = 1$.

Then the Kesten-Stigum theorem ([KS66],[AN72]) says that there is a scalar non-negative random variable W such that

$$\lim_{n \to \infty} \frac{\mathbf{Z}_n}{\rho^n} = Wv \text{ a.s.} \tag{2.1.8}$$

and

$$P^i\{W > 0\} = P^i\left\{\lim_{n \to \infty} \|\mathbf{Z}_n\|_1 = \infty\right\} \tag{2.1.9}$$

iff (2.1.6) holds.

By the Kesten-Stigum theorem, conditioned on non-extinction,

$$\begin{aligned}
&\lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1 \\
&= \lim_{n \to \infty} \frac{1}{n} \log \rho^n \langle v, \mathbf{1} \rangle W \\
&= \log \rho + \lim_{n \to \infty} \frac{1}{n} \log W \\
&= \log \rho.
\end{aligned}$$

$\square$

Now we explore the solution to the allocation problem w.r.t. the growth rate in the simple model I. Assume that $(1 - d_i)c_i > 1$ for $i \in S$. The mean matrix is given by

$$M = \begin{pmatrix} (1 - d_1)c_1 x & (1 - d_1)c_1(1 - x) \\ (1 - d_2)c_2 x & (1 - d_2)c_2(1 - x) \end{pmatrix}.$$

It is obvious that M is positively regular for $0 < x < 1$. By applying Theorem 2.1.1, we obtain that the growth rate is given by

$$\lambda(x) := \log \rho(x) = \log((1 - d_1)c_1 x + (1 - d_2)c_2(1 - x)), \text{ for } 0 < x < 1. \tag{2.1.10}$$

For the two extreme cases $x = 0$ and $x = 1$ of the simple model I, we can deduce that the growth rates are $\log(1 - d_1)c_1$ and $\log(1 - d_2)c_2$ respectively.

Firstly, suppose that $(1-d_1)c_1 > (1-d_2)c_2 > 1$. Then we get the obvious conclusion that $\log(1 - d_2)c_2 < \lambda(x) < \log(1 - d_1)c_1$ for $0 < x < 1$. Hence under the above assumptions the optimal strategy for maximizing the growth rate is that all offspring develop into type-1. All mixed strategies lead to a smaller growth rate.

Secondly, suppose that $(1 - d_1)c_1 = (1 - d_2)c_2 > 1$. Then any allocation strategy results in the same growth rate. So any strategy leads to an optimal growth rate.

Finally, suppose that $(1-d_2)c_2 > (1-d_1)c_1 > 1$. Then under the above assumptions the optimal strategy for maximizing the growth rate is that all offspring develop into type-2.

2.1.3 The extinction probability

The extinction probabilities of a multitype branching process is represented as

$$q^i = P^i \left\{ \lim_{n \to \infty} \|\mathbf{Z}_n\|_1 = 0 \right\}, \text{ for } i \in S. \tag{2.1.11}$$

The theory of branching processes characterizes the extinction probabilities via the generating functions of the offspring distributions. We first introduce some notations. Let $f^i(s_1, \cdots, s_d) := E^i \left[s_1^{Z_1^1} \cdots s_d^{Z_1^d} \right]$ with $0 \leq s_i < 1$ denote the generating functions of the offspring distributions p^i. We write $\mathbf{s} = (s_1, \cdots, s_d)$, $\mathbf{f}(\mathbf{s}) = (f^1(\mathbf{s}), \cdots, f^d(\mathbf{s}))$ and $\mathbf{q} = (q^1, \cdots, q^d)$. Write $C_d = \{\mathbf{s} \mid \mathbf{0} \leq \mathbf{s} < \mathbf{1}\}$, where $\mathbf{0} \leq \mathbf{s} < \mathbf{1}$ means $0 \leq s_i < 1$, for $1 \leq i \leq d$. The main result is given below (complete proof see section II.7 of [Har63]).

Theorem 2.1.2. *Assume that $\{\mathbf{Z}_n\}_{n \geq 0}$ is nonsingular, positively regular and supercritical. Then $\mathbf{q} < \mathbf{1}$ and $\mathbf{q}$ is the only solution of*

$$\mathbf{f}(\mathbf{s}) = \mathbf{s} \text{ in } C_d. \tag{2.1.12}$$

For the reduced ith-type one dimensional branching processes, we use notations with lower indices, i.e. q_i and f_i denote the extinction probability and the generating function of p_i, respectively. A similar result, as Theorem 2.1.2, holds for the extinction probability in the one-dimensional case, see section I.6 of [Har63].

18

Now we explore the solution to the allocation problem w.r.t. the extinction probability in the simple model I. Assume that $(1 - d_i)c_i > 1$ for $i \in S$. Then for the two-type Galton-Watson process, the generating functions are

$$f^1(s_1, s_2) = d_1 + (1 - d_1)[s_1 x + s_2(1 - x)]^{c_1} \tag{2.1.13}$$

and

$$f^2(s_1, s_2) = d_2 + (1 - d_2)[s_1 x + s_2(1 - x)]^{c_2}. \tag{2.1.14}$$

The extinction probabilities satisfy the equations:

$$q^1 = f^1(q^1, q^2), \tag{2.1.15}$$

$$q^2 = f^2(q^1, q^2). \tag{2.1.16}$$

For the reduced 1st-type one-dimensional Galton-Watson process, the generating functions are

$$f_1(s) = d_1 + (1 - d_1)s^{c_1}. \tag{2.1.17}$$

The extinction probability satisfies the equation:

$$q_1 = f_1(q_1). \tag{2.1.18}$$

For the reduced 2nd-type one-dimensional Galton-Watson process, the generating function is

$$f_2(s) = d_2 + (1 - d_2)s^{c_2}. \tag{2.1.19}$$

The extinction probability satisfies the equation:

$$q_2 = f_2(q_2). \tag{2.1.20}$$

First of all, assume that $(1 - d_1)c_1 > (1 - d_2)c_2 > 1$ and $d_1 < d_2$. Then q_1, q_2, q^1 and q^2 can be compared by the following lemma.

Lemma 2.1.1. *Assume that $d_1 < d_2$ and $c_1(1 - d_1) > c_2(1 - d_2) > 1$. Then*

$$d_2 + (1 - d_2)s^{c_2} > d_1 + (1 - d_1)s^{c_1},$$

for $s \in (0, 1)$.

19

Proof. Let

$$g(s) = d_2 + (1 - d_2)s^{c_2} - d_1 - (1 - d_1)s^{c_1},$$

then it is sufficient to prove $g(s) > 0$. In order to prove this we distinguish the following two cases.

Case 1: $c_1 \geq c_2$.

$$
\begin{aligned}
g(s) &= d_2 - d_1 + (1 - d_2)s^{c_2} - (1 - d_1)s^{c_1} \\
&\geq d_2 - d_1 + (1 - d_2)s^{c_1} - (1 - d_1)s^{c_1} \\
&= (d_2 - d_1)(1 - s^{c_1}) \\
&> 0
\end{aligned}
$$

Case 2: $c_1 < c_2$.

Since $g(0) > 0$, $g(1) = 0$, we only need to prove

$$g'(s) < 0, \text{ for } s \in (0, 1).$$

Calculating the first and second derivative of $g(s)$, we get

$$g'(s) = (1 - d_2)c_2 s^{c_2 - 1} - (1 - d_1)c_1 s^{c_1 - 1},$$

and

$$g''(s) = (1 - d_2)c_2(c_2 - 1)s^{c_2 - 2} - (1 - d_1)c_1(c_1 - 1)s^{c_1 - 2}.$$

Let $g''(s) = 0$, we get

$$s = \left[\frac{(1 - d_1)c_1(c_1 - 1)}{(1 - d_2)c_2(c_2 - 1)} \right]^{\frac{1}{c_2 - c_1}}.$$

- if $(1 - d_1)c_1(c_1 - 1) \geq (1 - d_2)c_2(c_2 - 1)$, then $g''(s)$ has no zero point in the interval $(0, 1)$;

- if $(1 - d_1)c_1(c_1 - 1) < (1 - d_2)c_2(c_2 - 1)$, then $g''(s)$ has exactly one zero point in the interval $(0, 1)$ and $g''(1) > 0$.

Together with $g'(0) = 0$, $g'(1) < 0$, we get $g'(s) < 0$ for $0 < s < 1$. This implies that $g(s) > 0$ and so the prove is complete. $\quad\square$

20

Now we continue to study model I, in particular equations (2.1.13)-(2.1.19). By the lemma above, it is easy to deduce that $q^1 < q^2 < 1$ and $q_1 < q_2 < 1$. Furthermore, $q_1 < q^1$. Indeed, by (2.1.13) and (2.1.15) we have

$$q^1 = d_1 + (1 - d_1)[q^1 x + q^2(1 - x)]^{c_1} > d_1 + (1 - d_1)(q^1)^{c_1} \qquad (2.1.21)$$

and by (2.1.17) and (2.1.18) we get

$$q_1 = d_1 + (1 - d_1)(q_1)^{c_1}. \qquad (2.1.22)$$

Similarly, $q^2 < q_2$. Hence, $q_1 < q^1 < q^2 < q_2$. Under the above assumptions the optimal strategy for minimizing the extinction probability is that all offspring develop into type-1.

Secondly, assume that $(1 - d_1)c_1 > (1 - d_2)c_2 > 1$ and $d_1 = d_2$. Then

$$d_2 + (1 - d_2)s^{c_2} > d_1 + (1 - d_1)s^{c_1},$$

for $s \in (0, 1)$. Hence, under the above assumptions the optimal strategy for minimizing the extinction probability is that all offspring develop into type-1.

Finally, assume that $(1 - d_1)c_1 > (1 - d_2)c_2 > 1$ and $d_1 > d_2$. Let q be the solution of the equation

$$d_1 + (1 - d_1)s^{c_1} - (d_2 + (1 - d_2)s^{c_2}) = 0,$$

for $s \in (0, 1)$.

Case a: $q = d_1 + (1 - d_1)q^{c_1}$. Then $q_1 = q^1(x) = q^2(x) = q_2$ for each $x \in (0, 1)$. Then any allocation strategy results in the same extinction probability. So any allocation strategy is optimal.

Case b: $q < d_1 + (1 - d_1)q^{c_1}$. Then $q_1 < q^1(x) < q^2(x) < q_2$ for each $x \in (0, 1)$. Hence, under the above assumptions the optimal strategy for minimizing the extinction probability is that all offspring develop into type-1.

Case c: $q > d_1 + (1 - d_1)q^{c_1}$. Then $q_1 > q^1(x) > q^2(x) > q_2$ for each $x \in (0, 1)$. Hence, under the above assumptions the optimal strategy for minimizing the extinction probability is that all offspring develop into type-2.

21

From the discussion above, we get the conclusion that there always exists a pure strategy that has a lower or equal extinction probability than any mixed strategy. So the mixed strategy can not decrease the extinction probability.

2.1.4   The expected survival time conditioned on extinction

It is well known that a simple, supercritical Galton-Watson process turns into a subcritical one, if conditioned on extinction [AN72]. Jagers and Lagerås proved that general multitype supercritical branching processes conditioned on extinction are subcritical branching processes [JL08]. As a special case, we prove that the corresponding holds for multitype Galton-Watson processes.

Let $\{\mathbf{Z}_n\}_{n\geq 0}$ be a two-type supercritical Galton-Watson process defined on probability space $(\Omega, \mathcal{F}, P)$ with generating functions $f^1(s_1, s_2)$ and $f^2(s_1, s_2)$. Let $B = \{\omega \in \Omega \mid \|\mathbf{Z}_n(\omega)\|_1 = 0 \text{ for some } n\}$. Let $\tau = \inf\{n \in \mathbb{N} \mid \|\mathbf{Z}_n\|_1 = 0 \text{ for some } n\}$.

Theorem 2.1.3. *The process $\{\mathbf{Z}_n\}_{n\geq 0}$ conditioned on extinction is equivalent to a subcritical branching process with generating function $\widetilde{f}^1(s_1, s_2) = f^1(q^1 s_1, q^2 s_2)/q^1$ and $\widetilde{f}^2(s_1, s_2) = f^2(q^1 s_1, q^2 s_2)/q^2$, where q^1 and q^2 are the solutions of $q^1 = f^1(q^1, q^2)$ and $q^2 = f^2(q^1, q^2)$, respectively.*

Proof. It is obvious that the process $\{\mathbf{Z}_n\}_{n\geq 0}$ conditioned on extinction is still a Markov process. The transition mechanism of the conditioned process is deduced as

22

follows. For i_1, $i_2 \neq 0$, the generating function of the conditioned process is

$$\sum_{j_1,j_2=0}^{\infty} P\{\mathbf{Z}_{n+1} = (j_1, j_2) \mid \mathbf{Z}_n = (i_1, i_2), B\} s_1^{j_1} s_2^{j_2}$$

$$= \sum_{j_1,j_2=0}^{\infty} \frac{P\{\mathbf{Z}_{n+1} = (j_1, j_2), \mathbf{Z}_n = (i_1, i_2), B\}}{P\{\mathbf{Z}_n = (i_1, i_2), n < \tau < \infty\}} s_1^{j_1} s_2^{j_2}$$

$$= \sum_{j_1,j_2=0}^{\infty} s_1^{j_1} s_2^{j_2} \cdot$$

$$\frac{P\{\mathbf{Z}_n = (i_1, i_2)\} P\{\mathbf{Z}_{n+1} = (j_1, j_2) \mid \mathbf{Z}_n = (i_1, i_2)\} P\{B \mid \mathbf{Z}_{n+1} = (j_1, j_2), \mathbf{Z}_n = (i_1, i_2)\}}{P\{\mathbf{Z}_n = (i_1, i_2)\} P\{n < \tau < \infty \mid \mathbf{Z}_n = (i_1, i_2)\}}$$

$$= \frac{1}{(q^1)^{i_1} (q^2)^{i_2}} \sum_{j_1,j_2=0}^{\infty} P\{\mathbf{Z}_{n+1} = (j_1, j_2) \mid \mathbf{Z}_n = (i_1, i_2)\} (q^1 s_1)^{j_1} (q^2 s_2)^{j_2}$$

$$= [\widetilde{f}^1(s_1, s_2)]^{i_1} [\widetilde{f}^2(s_1, s_2)]^{i_2},$$

since $P\{\mathbf{Z}_{n+1} = (j_1, j_2) | \mathbf{Z}_n = (i_1, i_2)\}$ is the coefficient of $s_1^{j_1} s_2^{j_2}$ in $[f^1(s_1, s_2)]^{i_1} [f^2(s_1, s_2)]^{i_2}$.

Hence the conditioned process is a branching process with generating function $\widetilde{f}^1(s_1, s_2)$ and $\widetilde{f}^2(s_1, s_2)$.

Now we prove that the conditioned process is subcritical. From the definition of $\widetilde{f}^1(s_1, s_2)$, we can deduce that the conditioned offspring distributions satisfy

$$\widetilde{p}^1(k_1, k_2) = (q^1)^{k_1-1} (q^2)^{k_2} p^1(k_1, k_2) \tag{2.1.23}$$

and

$$\widetilde{p}^2(k_1, k_2) = (q^1)^{k_1} (q^2)^{k_2-1} p^2(k_1, k_2), \tag{2.1.24}$$

where $p^1(k_1, k_2)$ and $p^2(k_1, k_2)$ are the offspring distributions of the supercritical process. Furthermore,

$$\widetilde{P}^1\{\mathbf{Z}_n = (k_1, k_2)\} = (q^1)^{k_1-1} (q^2)^{k_2} P^1\{\mathbf{Z}_n = (k_1, k_2)\}.$$

Then

$$\begin{aligned}
\widetilde{E}^1[\|\mathbf{Z}_n\|_1] &= \frac{1}{q^1} E^1 \left[\|\mathbf{Z}_n\|_1 (q^1)^{\mathbf{Z}_n^1} (q^2)^{\mathbf{Z}_n^2} \right] \\
&= \frac{1}{q^1} E^1 [\|\mathbf{Z}_n\|_1 \exp(\mathbf{Z}_n^1 \log q^1 + \mathbf{Z}_n^2 \log q^2)] \\
&\leq \frac{1}{q^1} E^1 \left[\|\mathbf{Z}_n\|_1 q^{\|\mathbf{Z}_n\|_1} \right] \\
&\to 0 \quad \text{as } n \to \infty,
\end{aligned}$$

23

since $\|\mathbf{Z}_n\|_1$ must either tend to zero or to infinity, where $q = \max\{q^1, q^2\} < 1$. $\square$

Then the conditioned two-type branching process is a subcritical branching process with probability generating functions

$$\tilde{f}^1(s_1, s_2) = \frac{f^1(q^1 s_1, q^2 s_2)}{q^1}; \tilde{f}^2(s_1, s_2) = \frac{f^2(q^1 s_1, q^2 s_2)}{q^2}. \tag{2.1.25}$$

Define T_1 and T_2 as the extinction times of the conditioned branching processes starting with $(1, 0)^\top$ and $(0, 1)^\top$ respectively.

Lemma 2.1.2. *Assume that $d_1 < d_2$, $c_1 \leq c_2$ (c_1 and c_2 are positive integers) and $(1 - d_1)c_1 > (1 - d_2)c_2 > 1$. Then $ET_1 > ET_2$.*

Proof. Let $\tilde{f}^{i(n)}(s_1, s_2) = \tilde{f}^i(\tilde{f}^{1(n-1)}(s_1, s_2), \tilde{f}^{2(n-1)}(s_1, s_2))$ for $i = 1, 2$. Since $ET_i = \sum_{n=0}^\infty P(T_i > n) = \sum_{n=0}^\infty (1 - \tilde{f}^{i(n)}(0, 0))$ for $i = 1, 2$, it is enough to show $\tilde{f}^{1(n)}(0, 0) < \tilde{f}^{2(n)}(0, 0)$ in order to show $ET_1 > ET_2$. Since $\tilde{f}^{i(n)}(0, 0) = \frac{f^i(f^{1(n-1)}(0,0), f^{2(n-1)}(0,0))}{q^i}$ and $\frac{q^1}{q^2} = \frac{f^1(q^1, q^2)}{f^2(q^1, q^2)}$, we need to show that

$$\frac{f^1(f^{1(n-1)}(0,0), f^{2(n-1)}(0,0))}{f^2(f^{1(n-1)}(0,0), f^{2(n-1)}(0,0))} < \frac{f^1(q^1, q^2)}{f^2(q^1, q^2)}.$$

Let $h(s_1, s_2) = \frac{f^1(s_1, s_2)}{f^2(s_1, s_2)}$. Since we assume that $d_1 < d_2$, $c_1 \leq c_2$ (c_1 and c_2 are positive integers) and $(1 - d_1)c_1 > (1 - d_2)c_2 > 1$, it is obvious that $\frac{\partial h(s_1, s_2)}{\partial s_i} > 0$ for $i = 1, 2$. By induction, we get $f^{i(n)}(0, 0) \uparrow q^i$ as $n \to \infty$. $\square$

For $i = 1, 2$, the conditioned ith-type one-dimensional branching process is a subcritical branching process with probability generating function

$$\tilde{f}_i(s) = \frac{f_i(q_i s)}{q_i}. \tag{2.1.26}$$

Define τ_1 and τ_2 as the extinction times of the conditioned 1st-type and 2nd-type branching processes respectively.

Lemma 2.1.3. *Assume that $0 < d_1 < d_2 < 1/2$ and $c_1 = c_2 = 2$. Then $E\tau_1 < E\tau_2$.*

Proof. Since $c_1 = c_2 = 2$, we get $q_i = \frac{d_i}{1-d_i}$ for $i = 1, 2$. Then

$$\tilde{f}_i(s) = \frac{f_i(q_i s)}{q_i} = (1 - d_i) + d_i s^2.$$

24

Let $\tilde{f}_i^{(n)}(s) = \tilde{f}_i(\tilde{f}_i^{(n-1)}(s))$ for $i = 1, 2$. By induction, we get $\tilde{f}_1^{(n)}(0) > \tilde{f}_2^{(n)}(0)$.

Hence

$$
\begin{aligned}
E\tau_1 &= \sum_{n=0}^{\infty} P(\tau_1 > n) = \sum_{n=0}^{\infty}(1 - \tilde{f}_1^{(n)}(0)) \\
&< \sum_{n=0}^{\infty}(1 - \tilde{f}_2^{(n)}(0)) = \sum_{n=0}^{\infty} P(\tau_1 > n) = E\tau_2.
\end{aligned}
$$

□

Remark 1: The relation between $E\tau_i$ and ET_i is not clear. Consider the following simple example. Let $d_1 = 0.2$, $d_2 = 0.4$ and $c_1 = c_2 = 2$. Numerical calculations yield: if $x = 0.1$, or 0.2, we get the relation $ET_1 > E\tau_2 > ET_2 > E\tau_1$; if $x = 0.3, 0.4, 0.5$, or 0.6, we get $E\tau_2 > ET_1 > ET_2 > E\tau_1$; if $x = 0.7, 0.8$, or 0.9, we get $E\tau_2 > ET_1 > E\tau_1 > ET_2$.

Remark 2: For the above example, we mention some observations: firstly, ET_i decreases w.r.t. x; secondly, when $x \to 0$, the relation of $ET_1 > E\tau_2 = ET_2 > E\tau_1$ holds. Hence, when the population starts with one individual of type-1, the optimal strategy is that each individual develops into type-2 with probability as close as possible to 1; when the population starts with one individual of type-2, the optimal strategy is that all individuals develop into type-2 with probability 1.

2.1.5 The expected number of descendants conditioned on extinction

The expected number of descendants of the two-type branching process conditioned on extinction is given by

$$
E\Big[\sum_{n=0}^{\infty} \|\mathbf{Z}_n\|_1 \,\Big|\, \lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0\Big]. \tag{2.1.27}
$$

Assume that $0 < d_1 < d_2 < 1/2$ and $c_1 = c_2 = 2$. By the monotone convergence theorem, we get

$$E\left[\sum_{n=0}^{\infty} \|\mathbf{Z}_n\|_1 \mid \lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0\right]$$

$$= \sum_{n=0}^{\infty} E\left[\|\mathbf{Z}_n\|_1 \mid \lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0\right]$$

$$= \sum_{n=0}^{\infty} \mathbf{Z}_0^\top \tilde{M}^n e,$$

where $\mathbf{Z}_0 = (z_1, z_2)^\top$ is the initial state of the process,

$$\tilde{M} = \begin{pmatrix} 2(q^1 x + q^2(1-x))(1-d_1)x & 2(q^1 x + q^2(1-x))(1-d_1)(1-x)q^2/q^1 \\ 2(q^1 x + q^2(1-x))(1-d_2)xq^1/q^2 & 2(q^1 x + q^2(1-x))(1-d_2)(1-x) \end{pmatrix}$$

and e is a column vector of all ones.

Lemma 2.1.4. *Assume that $0 < d_1 < d_2 < 1/2$ and $c_1 = c_2 = 2$. Then*

$$E\left[\sum_{n=0}^{\infty} \|\mathbf{Z}_n\|_1 \mid \lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0\right] = \frac{2}{1-\tilde{\rho}}\left(z_1\left(1 - \frac{d_1}{q^1}\right) + z_2\left(1 - \frac{d_2}{q^2}\right)\right), \quad (2.1.28)$$

where $(z_1, z_2)^\top$ is the initial size, (q^1, q^2) are extinction probabilities, and $\tilde{\rho}$ is the principle eigenvalue of $\tilde{M}$.

Proof. Since $\tilde{M}$ is a rank-one matrix, it is obvious that the principle eigenvalue

$$\tilde{\rho} = 2(q^1 x + q^2(1-x))((1-d_1)x + (1-d_2)(1-x))$$

and

$$\sum_{n=0}^{\infty} \mathbf{Z}_0^\top \tilde{M}^n e = z_1 + z_2 + \frac{2(q^1 x + q^2(1-x))^2 \left(\frac{z_1(1-d_1)}{q^1} + \frac{z_2(1-d_2)}{q^2}\right)}{1 - \tilde{\rho}}$$

From the equations $(2.1.13) - (2.1.16)$ we deduce that

$$[q^1 x + q^2(1-x)]^2 = \frac{q^1 - d_1}{1 - d_1} = \frac{q^2 - d_2}{1 - d_2};$$

$$q^2 = \frac{1 - d_2}{1 - d_1}(q^1 - d_1) + d_2;$$

$$q^1 = \frac{(1-d_1)(d_1 x + d_2(1-x)) - ((1-d_1)x + (1-d_2)(1-x))(d_2 - d_1)(1-x)}{((1-d_1)x + (1-d_2)(1-x))^2}.$$

26

Then

$$E\Big[\sum_{n=0}^{\infty} \|\mathbf{Z}_n\|_1 \ \Big| \ \lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0\Big] = \frac{2}{1-\tilde{\rho}} \left(z_1 \left(1 - \frac{d_1}{q^1} \right) + z_2 \left(1 - \frac{d_2}{q^2} \right) \right).$$

$\square$

By basic calculations, we get that q^1, q^2 and $\tilde{\rho}$ all decrease w.r.t. x. Hence $E\Big[\sum_{n=0}^{\infty} \|\mathbf{Z}_n\|_1 \ \Big| \ \lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0\Big]$ decreases w.r.t. x. Under the above assumptions the optimal strategy for maximizing the expected number of descendants conditioned on extinction is that all offspring develop into type-2.

2.2 Population models in fluctuating environments

In contrast to the last section we will consider the evolution of a population with environmental stochasticity. In this section we focus on models where some random events may simultaneously affect all the individuals in a population. Typical examples are large-scale fluctuations in the overall environmental conditions, where all population members are subject to identical environmental conditions at a given time. We refer to this form of stochasticity as environmental stochasticity.

The stochastic model for a non-structured population in a fluctuating environment [HI95] can be described as follows. Let Y_n be the population size just before period $n + 1$. Let ξ_n denote the expected number of offspring per individual in period n. Then Y_n can be recursively expressed as

$$Y_n = Y_{n-1}\xi_{n-1}. \tag{2.2.1}$$

In this section, we first give a simple biological model of a population in a fluctuating environment. Then we characterize the growth rate, the extinction probability, the expected survival time conditioned on extinction, and the discounted expected number of descendants in the fluctuating model. Finally we solve the allocation problems w.r.t. the above payoff functions.

2.2.1 Simple model II

Consider a simple model of organisms living in a fluctuating environment [BL04]. The environment sequence are independent and identically distributed. Each environment has two possible states, state 1 and state 2, that occur with probability p and $1 - p$ respectively. All individuals encounter exactly the same environment in a given period. There are two kinds of phenotypes for organisms: one suited to environment 1, and the other suited to 2. An organism can only survive and reproduce if its phenotype properly adapted to the current environment. The organism's fitness is given by the following matrix:

$$
\begin{array}{ccc}
 & Phenotype\ 1 & Phenotype\ 2 \\
Environment\ 1 & w_1 & 0 \\
Environment\ 2 & 0 & w_2
\end{array}
$$

These organisms will be selected to hedge their bets, developing into phenotype 1 with frequency x and phenotype 2 otherwise.

Since for all generations the proportion of individuals of phenotype 1 and 2 is equal to $x/(1 - x)$, we can simply study the evolution of of the population size by using (2.2.1). In this model, ξ_n are independent and identically distributed with the same distribution as ξ where $P\{\xi = w_1 x\} = p$ and $P\{\xi = w_2(1 - x)\} = 1 - p$.

2.2.2 The growth rate

The long-term geometric growth rate γ is defined as

$$
\begin{aligned}
\gamma &= \lim_{n \to \infty} \frac{1}{n} \log\left(\frac{Y_n}{Y_0}\right) \\
&= \lim_{n \to \infty} \frac{1}{n} \sum_{i=0}^{n-1} \log \xi_i \\
&= E[\log \xi] \\
&= p \log(w_1 x) + (1 - p) \log(w_2(1 - x)).
\end{aligned}
\tag{2.2.2}
$$

By a simple calculation, we get that γ is maximized when $x = p$. Note that the optimal strategy is independent of w_1 and w_2. Thus for almost all sequences

of environments, the strategy that develops with frequency p into phenotype 1 will maximize the long-term geometric growth rate. For this strategy, the growth rate is given by $\gamma(p) = p \log(w_1 p) + (1 - p) \log(w_2(1 - p))$.

2.2.3 The extinction probability

We consider three cases of the extinction probability depending on the initial population size.

Initial size is one

We will consider what happens to this model if Y_0 equals one and define the extinction probability in this case as the chance that Y_n becomes less than one. Since $\log(Y_n) = \log(Y_0) + \sum_{i=1}^{n} \log(\xi_i)$, where $\{\log(\xi_i)\}$ are i.i.d. random variables, the sequence $\{\log(Y_n)\}$ can be viewed as a random walk. From the theory of random walks (see, e.g., [Fel71], Section XII.7) we can infer that the extinction probability of the process equals

$$q = P\left\{\inf_n \log(Y_n) < 0\right\} = 1 - \exp\left[-\sum_{n=1}^{\infty} \frac{1}{n} P\{\log(Y_n) < 0\}\right], \qquad (2.2.3)$$

and that the following relations hold:

$$\gamma \leq 0 \Leftrightarrow q = 1, \ \gamma > 0 \Leftrightarrow q < 1. \qquad (2.2.4)$$

We need calculate the probability $P\{\log(Y_n) < 0\}$. By the generating function, we get

$$
\begin{aligned}
E\left[s^{\log(Y_n)}\right] &= \prod_{i=1}^{n} E\left[s^{\log \xi_i}\right] \\
&= \left[p s^{\log(w_1 x)} + (1 - p) s^{\log(w_2(1-x))}\right]^n \\
&= \sum_{k=0}^{n} \binom{n}{k} p^k (1 - p)^{n-k} s^{k \log(w_1 x) + (n-k) \log(w_2(1-x))}, \quad (2.2.5)
\end{aligned}
$$

for $0 \leq s < 1$. Hence, the probability

$$P\{\log(Y_n) < 0\} = \sum_{k \in K} \binom{n}{k} p^k (1 - p)^{n-k}, \qquad (2.2.6)$$

29

where K denotes the set $\{0 \le k \le n \mid k \log(w_1 x) + (n-k) \log(w_2(1-x)) < 0\}$.

Write $b(x) = \log(w_1 x)$ and $c(x) = \log(w_2(1-x))$. Let $\lfloor y \rfloor$ denote the largest integer which is less than or equal to y for any $y \in \mathbb{R}$. There are four cases:

1. if $b(x) < 0$ and $c(x) > 0$, then $\lfloor \frac{cn}{c-b} \rfloor \le k \le n$;

2. if $b(x) > 0$ and $c(x) < 0$, then $1 \le k \le \lfloor \frac{cn}{c-b} \rfloor$;

3. if $b(x) > 0$ and $c(x) \ge 0$ or $b(x) \ge 0$ and $c(x) > 0$, then $K = \emptyset$ and $q(x) = 0$;

4. if $b(x) \le 0$ and $c(x) \le 0$, then $1 \le k \le n$ and $q(x) = 1$.

Recall that $\gamma(x) = p \log(w_1 x) + (1-p) \log(w_2(1-x))$ and $\gamma(x) \le \gamma(p)$. Let p_0 be the minimal solution in the interval $(0,1)$ of the equation

$$\gamma(p) = 0. \tag{2.2.7}$$

Let $f(x) = b(x)/c(x)$. Then the derivative of $f(x)$ is

$$f'(x) = \frac{\frac{1}{x} \log(w_2(1-x)) + \frac{1}{1-x} \log(w_1 x)}{(\log(w_2(1-x)))^2}.$$

Let x_0 be the minimal solution in the interval $(0,1)$ of the equation

$$f'(x) = 0. \tag{2.2.8}$$

It is clear that $p_0 = x_0$ if they exist.

Example 1: Assume that $w_1 = 1$ and $w_2 = 10$. Then $p_0 = x_0 \approx 0.7613$. If $p_0 \le p < 1$, we get $\gamma(p) \le 0$. Hence $\gamma(x) \le 0$ and $q(x) = 1$ for all $x \in (0,1)$. If $0 < p < p_0$, we get $\gamma(p) > 0$. Furthermore, for a fixed $p \in (0, p_0)$, there exist two solutions x_- and x_+ in $(0,1)$ for the equation

$$\gamma(x) = 0. \tag{2.2.9}$$

It is deduced that $\gamma(x) \le 0$ and $q(x) = 1$ when $0 < x < x_-$ or $x_+ \le x < 1$; $\gamma(x) > 0$, $q(x) < 1$ and $q(x)$ decreases w.r.t. x when $x_- < x \le x_0$; and $\gamma(x) > 0$, $q(x) < 1$ and $q(x)$ increases w.r.t. x when $x_0 \le x < x_+$.

30

Example 2: Assume that $w_1 = 2$ and $w_2 = 5$. Then $\gamma(p) > 0$ for all $p \in (0, 1)$. Let x_b be the solution of

$$b(x) = 0. \tag{2.2.10}$$

Let x_c be the solution of

$$c(x) = 0. \tag{2.2.11}$$

Furthermore, for a fixed $p \in (0, 1)$, there exist two solutions x_- and x_+ in (0,1) for the equation

$$\gamma(x) = 0 \tag{2.2.12}$$

since $\gamma(0+) = -\infty$ and $\gamma(1-) = -\infty$. It is obtained that $\gamma(x) \leq 0$ and $q(x) = 1$ when $0 < x < x_-$ or $x_+ \leq x < 1$; $\gamma(x) > 0$, $q(x) < 1$, and $q(x)$ decreases w.r.t. x when $x_- < x < x_b$; $\gamma(x) > 0$, $q(x) < 1$, and $q(x)$ increases w.r.t. x when $x_c \leq x < x_+$; $q(x) = 0$ when $x_b \leq x \leq x_c$.

Initial size is sufficiently large

Assume that $\log(Y_0) = a \geq 0$. Recall that $\log(Y_n) = \log(Y_0) + \sum_{i=1}^{n} \log(\xi_i)$, where $\{\log(\xi_i)\}$ are i.i.d. random variables. We are interested in the probability that the random walk $\{\log(Y_n)\}$ will reach 0 [Gre80]. The following theorem can be found in Chapter XII of [Fel71]. It is stated in terms of the Laplace transform $\hat{F}(u)$ of $\log(\xi_i)$,

$$\hat{F}(u) = E\left[e^{-u\log(\xi_i)}\right].$$

Theorem 2.2.1. *Suppose that there exists $\kappa > 0$ such that $\hat{F}(\kappa) = 1$ and $\hat{F}'(\kappa)$ is finite. For $a \geq 0$, let $P(a)$ be the probability that the random walk $\{\log(Y_n)\}$ ever crosses level 0, that is,*

$$P(a) = P\left\{\inf_n\{\log(Y_n)\} \leq 0\right\} = P\left\{\inf_n\left\{\sum_{i=1}^{n}\log(\xi_i)\right\} \leq -a\right\}.$$

Then $P(a) \sim Ce^{-\kappa a}$ as $a \to \infty$, for some constant $C > 0$.

The above theorem tells us that for two strategies with different values of κ, the one with the larger κ-value has a smaller probability of extinction, provided that the initial

31

population size is sufficiently large. Among all strategies for which the parameter κ exists, the strategy that maximizes κ minimizes the probability of extinction for a large population.

Example 3: Assume that $w_1 = 1$, $w_2 = 10$ and $p = 0.6$. Then

$$
\begin{aligned}
\hat{F}(\kappa) &= E\left[e^{-\kappa \log(\xi_i)}\right] \\
&= pe^{-\kappa \log(w_1 x)} + (1-p)e^{-\kappa \log(w_2(1-x))} \\
&= 0.6(x)^{-\kappa} + 0.4(10(1-x))^{-\kappa}.
\end{aligned}
$$

By numerical computation we get that there exists $\kappa > 0$ such that $\hat{F}(\kappa) = 1$ and $\hat{F}'(\kappa)$ is finite when $x \in (0.25, 0.88)$ at intervals of 0.01 and κ is maximized when $x = 0.84$.

Initial size is a constant

Recall that $\{\log(\xi_i)\}$ are i.i.d. random variables of two values $\log(w_1 x)$ and $\log(w_2(1-x))$ with probability p and $1 - p$ respectively. Then the expectation of $\log(\xi_i)$ is

$$
\gamma = E[\log(\xi_i)] = p \log(w_1 x) + (1-p) \log(w_2(1-x)),
$$

and the variance is

$$
\sigma^2 = Var(\log(\xi_i)) = p[\log(w_1 x) - \gamma]^2 + (1-p)[\log(w_2(1-x)) - \gamma]^2.
$$

The central limit theorem implies that for each $x \in (0,1)$,

$$
\lim_{n \to \infty} P\left\{ \frac{\log\left(\frac{Y_n}{Y_0}\right) - \gamma n}{(\sigma^2 n)^{1/2}} \leq z \right\} = \frac{1}{(2\pi)^{1/2}} \int_{-\infty}^{z} e^{-u^2/2} du, \tag{2.2.13}
$$

if $\sigma^2 \neq 0$. Then for large n the quantity $\log\left(\frac{Y_n}{Y_0}\right)$ behaves approximately as a Wiener process W with mean γ and variance σ^2. For such a Wiener process it is well known [KT81] that the probability of ever reaching zero starting from a positive w_0 is equal to 1 if $\gamma \leq 0$, and is $\exp(-2\gamma w_0/\sigma^2)$ if $\gamma > 0$.

The convergence rate of the central limit theorem is of order $n^{-1/2}$ [Hal82] for $0 < x < 1$. The convergence rate is sufficiently fast that even on a short timescale

32

the changes in the logarithmic population size can be accurately approximated as a Wiener process W with constant infinitesimal mean γ and infinitesimal variance σ^2. Hence the extinction probability $P(a) = P\{\inf_n\{\log(Y_n)\} \leq 0\}$ can be accurately approximated by $\exp(-2\gamma a/\sigma^2)$ if $\gamma > 0$ and $P(a) = 1$ if $\gamma \leq 0$.

Remark: Here we considered the one-dimensional model (2.2.1) of the population size. Alternatively, we could consider the evolution of a two-type structured population in fluctuating environment. For further detail, we refer to [TO80] and [LO88].

Example 4: Assume that $a = 2$, $w_1 = 1$, $w_2 = 10$ and $p = 0.6$. By numerical computation we get that $P(2)$ is minimized when $x = 0.84$ at intervals of 0.01.

2.2.4 The mean survival time conditioned on extinction

According to the central limit theorem, the probability distribution of the natural logarithm of the population size asymptotically approaches a normal distribution. The rate of approach to normality will be sufficiently fast that even for short times the changes in logarithmic population size can be accurately approximated as a Wiener process W with constant infinitesimal mean γ and infinitesimal variance σ^2.

As was shown in [LO88], when $\gamma \neq 0$, the mean survival time conditioned on extinction is

$$E[T \mid T < \infty] = a/|\gamma|, \tag{2.2.14}$$

where $T = \inf\{t \geq 0 \mid W(t) \leq 0\}$ and $W(0) = \log(Y_0) = a \geq 0$.

We want to find the optimal allocation strategy to obtain the supremum of the mean survival time conditioned on extinction, i.e.,

$$\sup_{\{x \in [0,1] \mid \gamma(x) \neq 0\}} E[T \mid T < \infty]. \tag{2.2.15}$$

Example 5: Assume that $w_1 = 1$, $w_2 = 2$ and $p = 0.5$. Then the optimal strategy is $x^* = p = 0.5$.

2.2.5 The discounted expected number of descendants

The discounted expected number of descendants is given by

$$E\left[\sum_{n=0}^{\tau-1}\alpha^n Y_n\right], \tag{2.2.16}$$

where $\tau = \inf\{n \geq 0 \mid Y_n \leq 1\}$ and α is a constant in the interval $(0,1]$ satisfying $\alpha \max_{0 \leq x \leq 1}(pw_1 x + (1-p)w_2(1-x)) < 1$.

In order to use dynamical programming, we rewrite the population process as

$$Y_{n+1} = Y_n\left(w_1 x \mathbf{1}_{\{\eta_n=1\}} + w_2(1-x)\mathbf{1}_{\{\eta_n=2\}}\right), \tag{2.2.17}$$

where $\{\eta_n\}$ are i.i.d. random variables of two values 1 and 2 with probability p and $1-p$ respectively and

$$\mathbf{1}_{\{\eta_n=1\}}(\omega) = \begin{cases} 1, & \eta_n(\omega) = 1; \\ 0, & \text{otherwise.} \end{cases}$$

Furthermore, we generalize the original process by assuming that the allocation probability to two phenotypes can change from one generation to the next generation, i.e., we replace x with x_n. Then the generalized process evolves according to the equation

$$Y_{n+1} = Y_n\left(w_1 x_n \mathbf{1}_{\{\eta_n=1\}} + w_2(1-x_n)\mathbf{1}_{\{\eta_n=2\}}\right). \tag{2.2.18}$$

Assume that $Y_0 = y_0 \geq 1$.

The problem is to determine the optimal allocation strategy that maximizes the discounted expected number of descendants. It may be cast within the total cost, infinite horizon framework, where we consider maximization instead of minimization [Ber05]. The state space is the set $I = [0, \infty) \cup \{\Delta\}$, where Δ is the termination state to which the system moves with certainty from states in $[0, 1]$ with corresponding reward equal to 0. When $Y_n \in (1, \infty)$, the system evolves according to (2.2.18). The reward per stage when $Y_n \in (1, \infty)$ is Y_n. The control constraint set is $U(y) = [0, 1] = C$ for all $y \in I$. Given an initial state y_0, we want to find a strategy $\pi = \{\mu_0, \mu_1, \cdots\}$, where $\mu_n : I \to C$, $\mu_n(y_n) = x_n \in U(y_n)$, for all $y_n \in I$, $n = 0, 1, \cdots$, that maximizes the return function

$$J_\pi(y_0) = E\left[\sum_{n=0}^{\infty}\alpha^n Y_n \mathbf{1}_{\{Y_n>1\}}\right]. \tag{2.2.19}$$

We denote by Π the set of all admissible policies π, i.e., the set of all sequences of functions $\pi = \{\mu_0, \mu_1, \cdots\}$, where $\mu_n : I \to C$, $\mu_n(y_n) = x_n \in U(y_n)$, for all $y_n \in I$, $n = 0, 1, \cdots$. The optimal return function J^* is defined by

$$J^*(y) = \max_{\pi \in \Pi} J_\pi(y), \; y \in I. \tag{2.2.20}$$

An optimal strategy, for a given initial state y, is one that attains the optimal return $J^*(y)$. A strategy of the form $\pi = \{\mu, \mu, \cdots\}$ is referred to as a stationary strategy μ, and $J_{\pi=\{\mu,\mu,\cdots\}}(y)$ is denoted by $J_\mu(y)$. We say that μ is optimal if $J_\mu(y) = J^*(y)$ for all states y.

The mapping T that defines the dynamic programming algorithm takes the form

$$(TJ)(y) = \begin{cases} \max_{0 \leq x \leq 1}(y + \alpha(pJ(yw_1 x) + (1-p)J(yw_2(1-x)))), & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1], \end{cases}$$

for any function $J : (0, \infty) \to [0, \infty]$.

By applying Prop. 3.1.1 in [Ber05], the optimal return function J^* satisfies Bellman's equation.

Theorem 2.2.2. *The optimal return function J^* satisfies*

$$J^*(y) = \begin{cases} \max_{0 \leq x \leq 1}(y + \alpha(pJ^*(yw_1 x) + (1-p)J^*(yw_2(1-x)))), & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1]. \end{cases}$$

Or, equivalently,

$$J^* = TJ^*.$$

We now give a characterization of an optimal stationary policy by Prop. 3.1.4 in [Ber05]:

Theorem 2.2.3. *A stationary policy μ is optimal if and only if*

$$TJ_\mu = T_\mu J_\mu,$$

where

$$T_\mu J_\mu(y) = \begin{cases} y + \alpha(pJ_\mu(yw_1 \mu(y)) + (1-p)J_\mu(yw_2(1-\mu(y)))), & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1]. \end{cases}$$

35

Example 6: Assume that $w_1 = 1$, $w_2 = 10$ and $p = 0.6$. Then for $\alpha < 0.25$ the optimal stationary policy μ^* is given by

$$\mu^*(y) = 0, \text{ for all } y \in I. \tag{2.2.21}$$

We will show that the strategy μ^* is indeed an optimal policy. To this end it is sufficient to show that

$$TJ_{\mu^*} = T_{\mu*}J_{\mu*}.$$

In fact,

$$J_{\mu^*}(y) = \begin{cases} c(\alpha)y, & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1], \end{cases}$$

where $c(\alpha) = \frac{1}{1-4\alpha}$. Then

$$
\begin{aligned}
T_{\mu*}J_{\mu^*}(y) &= \begin{cases} y + \alpha(pJ_{\mu^*}(yw_1\mu^*(y)) + (1-p)J_{\mu^*}(yw_2(1-\mu^*(y)))), & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1], \end{cases} \\
&= \begin{cases} c(\alpha)y, & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1]. \end{cases}
\end{aligned}
$$

$$
\begin{aligned}
TJ_{\mu^*}(y) &= \begin{cases} \max_{0 \le x \le 1}(y + \alpha(pJ_{\mu^*}(yw_1x) + (1-p)J_{\mu^*}(yw_2(1-x)))), & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1], \end{cases} \\
&= \begin{cases} c(\alpha)y, & \text{if } y \in (1, \infty), \\ 0, & \text{if } y \in [0, 1]. \end{cases}
\end{aligned}
$$

Hence,

$$TJ_{\mu^*} = T_{\mu*}J_{\mu*}.$$

Consequently, the stationary policy μ^* is optimal.

36

2.3 Multitype Galton-Watson branching processes in random environments

In this section, we consider the evolution of a structured population with demographic and environmental stochasticity. The population is modeled as a multitype Galton-Watson branching process in a random environment (MBPRE).

Assume that the type space is $S = \{1, 2, \cdots, d\}$ and that $\{\mathbf{Z}_n = (Z_n^1, \cdots, Z_n^d)^\top\}$ defined on the probability space $(\Omega, \mathcal{F}, P)$ is a multitype Galton-Watson branching process in a random environment introduced by Athreya and Karlin [AK70] [AK72a] [AK72b]. Assume that $\zeta = (\zeta_0, \cdots, \zeta_n, \cdots)$ is the environmental sequence of the process defined on the same probability space $(\Omega, \mathcal{F}, P)$. Throughout this section we will assume that ζ is stationary and ergodic. The left shift transformation is denoted by θ. The expectation operator for P will be denoted by E.

We propose a simple model III which is a combination of the simple model I and the simple model II. By the convergence theorem of branching processes in random environments the long-term growth rate of the population is characterized by the largest Lyapunov exponent. The Random Perron-Frobenius theorem provides an alternative expression for the largest Lyapunov exponent, which is used to deduce an explicit formula of the largest Lyapunov exponent in the simple model III. The extinction probability can be approximated using a similar method of Wilkinson. Optimal strategies that maximize the largest Lyapunov exponent and minimize the approximate extinction probability are obtained for the simple model III. For simplicity, we consider the two-dimensional case. However, all results can be easily generalized to the multi-dimensional case.

2.3.1 Simple model III

Consider the following simple model of a structured population of annual organisms living in a random environment. The state of the environment in each year is an independent random variable with 2 states, that is, the environmental states $\{\zeta_l\}$ are

i.i.d. and for each l, ζ_l has two states y_1 and y_2 that occur with probabilities p_1 and p_2 respectively. All individuals encounter exactly the same environment in a given year. At the beginning of its development, each organism has to make a developmental decision between two possible phenotypes, each of which is best adapted to one of the two environments. The organisms fitness is given by the following matrix:

$$
\begin{array}{ccc}
& Phenotype\ 1 & Phenotype\ 2 \\
Environment\ 1 & w_{11} & w_{12} \\
Environment\ 2 & w_{21} & w_{22}
\end{array}\quad ,
$$

where $w_{ij} > 0$. These organisms will be selected to hedge their bets during their development, developing into phenotypes $1, 2$ with positive probabilities x_1 and x_2 respectively. An organism of phenotype i in environment k will die with probability d_{ki} leaving no offspring and survive with probability $1 - d_{ki}$ producing c_{ki} offspring, where $0 < d_{ki} < 1$ and c_{ki} are positive integers.

We model the evolution of the population by using a multitype Galton-Watson branching process in a random environment. Here $S = \{1, 2\}$ is the type space for the phenotype and for the environment. For $i, k \in S$, the fitness of an organism of phenotype i in environment k is $w_{ki} = E[\|\mathbf{Z}_1\|_1 \mid \mathbf{Z}_0 = \mathbf{e}_i, \zeta_0 = y_k]$. And for $i, k \in S$, the offspring distribution $p_k^{(i)}$ of an organism of type i in environment k is as follows:

$$
p_k^{(i)}(n_1, n_2) = \begin{cases}
d_{ki}, & n_1 = n_2 = 0; \\[2mm]
(1 - d_{ki}) \begin{pmatrix} c_{ki} \\ n_1 \end{pmatrix} x_1^{n_1} x_2^{n_2}, & n_1 + n_2 = c_{ki}; \\[2mm]
0, & \text{else.}
\end{cases}
$$

The corresponding multi-variant probability generating function of $p_k^{(i)}$ is

$$
\phi_k^{(i)}(s_1, s_2) \;=\; d_{ki} + (1 - d_{ki})(s_1 x_1 + s_2(1 - x_2))^{c_{ki}}.
$$

2.3.2 The growth rate

The growth rate is defined as

$$
\lim_{n \to \infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1, \text{ a.s. conditioned on non-extinction.} \tag{2.3.1}
$$

Tanny [Tan81] showed that the growth rate is equal to the largest Lyapunov exponent.

First of all, we give some notations. Define $M(\theta^n\omega) = [m_{ij}((\theta^n\omega)_0)]_{i,j\in S}$ to be the matrix of expected numbers of progeny conditioned on the environmental sequence $\omega = (\omega_0, \cdots, \omega_n, \cdots)$. Specially, $m_{ij}((\theta^n\omega)_0) = m_{ij}(\omega_n) = E[Z^j_{n+1} \mid \mathbf{Z}_n = \mathbf{e}_i, \zeta = \omega]$ is the expected number of progeny of type j produced from an individual of type i in generation n with the environment ω_n. We will throughout assume that the random variable M takes values in $\mathcal{M}_+$, where $\mathcal{M}_+$ is the semigroup of positive $d \times d$ matrices (a matrix $M = [m_{ij}]$ is positive, in symbols $M > 0$, if $m_{ij} > 0$ for all $1 \le i, j \le d$).

It is well-known [AK70][AK72a][AK72b] that

$$E[\mathbf{Z}_n^\top \mid \zeta = \omega, \mathbf{Z}_0] = \mathbf{Z}_0^\top M(\omega) \cdots M(\theta^{n-1}\omega).$$

Then it is important to study the matrix cocycle as in [FK60],

$$\Psi_n(\omega) := M(\omega) \cdots M(\theta^{n-1}\omega), \ n \in \mathbb{N}, \tag{2.3.2}$$

over θ generated by the random variable $M \colon \Omega \to \mathcal{M}_+$.

As is shown in the following theorem by Tanny [Tan81], the largest Lyapunov exponent $\lambda := \lim_{n\to\infty} \frac{1}{n} \log \|\Psi_n\|$ is the potential long-term growth rate of the total population, where $\|\cdot\|$ denotes the norm of a matrix.

Theorem 2.3.1. *Let $\mathbf{Z}_n$ be an MBPRE with ζ stationary and ergodic and M taking values in $\mathcal{M}_+$. Assume that $E[\log^+ \|M\|] < \infty$, where $\log^+ y$ denotes $\max\{\log y, 0\}$ for $y > 0$. Then the following holds:*
(i) If $\lambda < 0$, the population will become extinct with probability 1.
(ii) If $\lambda > 0$, there exists a positive probability that the population never becomes extinct. Furthermore, conditioned on survival, the growth rate is, with probability 1, given by $\lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1 = \lambda$.

Usually, it is difficult to calculate the largest Lyapunov exponent λ explicitly. However, the following random Perron-Frobenius theorem [AGD94] provides an alternative expression of λ. Put for $M > 0$,

$$a = \min_{1\le i,j\le d} m_{ij}, \ A = \max_{1\le i,j\le d} m_{ij}.$$

Theorem 2.3.2 (Random Perron-Frobenius Theorem)**.** *Consider the product of positive random matrices* (2.3.2) *and assume that*

$$\log^+ \frac{1}{a} \in L^1(P) \ and \ \log^+ A \in L^1(P),$$

where $L^1(P)$ denotes the class of all measurable functions $f : \Omega \to \mathbb{R}$ with $\int_\Omega |f(\omega)| P(d\omega) < \infty$. Then there is a θ-invariant set $\tilde{\Omega} \subseteq \Omega$ of full P-measure on which the following holds:

(i) There exist a unique positive random unit vector l and a positive random scalar ρ with $\log \rho \in L^1(P)$ such that

$$l(\omega)^\top M(\omega) = \rho(\omega) l(\theta\omega)^\top.$$

(ii) $\lambda = E \log \rho$.

Now we explore the solution to the allocation problem w.r.t. the growth rate in the simple model III. Recall that $m_{ij}(k) = E[Z_1^j \mid \mathbf{Z}_0 = \mathbf{e}_i, \zeta_0 = y_k]$.

Lemma 2.3.1. $m_{ij}(k) = w_{ki} x_j$.

Proof. The result follows from the definitions of w_{ki} and x_j and the properties of the binomial distribution. □

Lemma 2.3.2. *For each k, the conditional mean matrix $M(k) = [m_{ij}(k)]_{i,j \in S}$ is a rank one matrix and its dominant eigenvalue is given by $\rho_k = \sum_{i \in S} w_{ki} x_i$ with corresponding left eigenvector $l = (x_1, x_2)$.*

Proof. The result follows from properties of rank one matrix. □

Lemma 2.3.3. $\lambda = E \log \rho = \sum_{k \in S} p_k \log(\sum_{i \in S} w_{ki} x_i)$.

Proof. The result follows from properties of the product measure on the product space. □

Remark 1: Dombry, Mazza and Bansaye [DMB09] used a multitype branching process in a random environment to model the population growth and looked for the

optimal type allocation strategy that maximizes the net growth rate. They character-
ized the net growth rate as $\lim_{n\to\infty} \frac{1}{n} \log E[\|\mathbf{Z}_n\|_1 \mid \zeta]$ a.s.. In our example, under the
assumption that $w_{ij} > 0$ for each $i, j \in S$, $x_k > 0$ for $k \in S$ and ζ is i.i.d, it is obvious
that $\lim_{n\to\infty} \frac{1}{n} \log E[\|\mathbf{Z}_n\|_1 \mid \zeta] = \lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1 = \sum_{k\in S} p_k \log(\sum_{i\in S} w_{ki} x_i)$
a.s.. However, these two characterizations of the growth rate are not always the
same. For example, when $w_{12} = w_{21} = 0$ and $w_{11} > 0, w_{22} > 0$, Prop. 1 in [DMB09]
shows that $\lim_{n\to\infty} \frac{1}{n} \log E[\|\mathbf{Z}_n\|_1 \mid \zeta] = p_1 \log w_{11} x_1 + p_2 \log w_{22} x_2$ a.s. while in our
work the largest Lyapunov exponent $\lambda := \lim_{n\to\infty} \frac{1}{n} \log \| \Psi_n \| = -\infty$ and the relation
between λ and $\lim_{n\to\infty} \frac{1}{n} \log \|\mathbf{Z}_n\|_1$ is unsolved.

Remark 2: Similarly as in the simple model I, there are also two extreme cases of
the simple model III corresponding to $x_i = 1$ respectively. The extreme cases can be
modeled by ith-type one-dimensional branching process in a random environment. As-
sume that $w_{ij} > 0$ for $i, j \in S$, then the expression $\lambda = \sum_{k\in S} p_k \log(\sum_{i\in S} w_{ki} x_i)$ still
holds when $x_i = 1$. So we want to find the strategy that maximizes $\sum_{k\in S} p_k \log(\sum_{i\in S} w_{ki} x_i)$
subject to the constraints that the fractional investments in the various phenotypes
sum to one, i.e. $\sum_{i\in S} x_i = 1$ with $x_i \geq 0$.

Example 7: Let the fitness matrix be given by $\begin{pmatrix} a & 1 \\ 1 & b \end{pmatrix}$, where $1 < b < a$. Let
the probability of environment 1 be given by p. Then the growth rate is

$$\lambda(x) = p \log[ax + (1 - x)] + (1 - p) \log[x + b(1 - x)], \qquad (2.3.3)$$

for $0 \leq x \leq 1$. By a direct calculation, we get the choice of $x^*(p)$ that maximizes the
growth rate given the probability p of environment 1 is :

$$x^*(p) = \begin{cases} \frac{p(ab-1)-(b-1)}{(a-1)(b-1)}, & \frac{b-1}{ab-1} < p < \frac{a(b-1)}{ab-1}; \\ 0, & 0 < p \leq \frac{b-1}{ab-1} ; \\ 1, & \frac{a(b-1)}{ab-1} \leq p < 1. \end{cases} \qquad (2.3.4)$$

2.3.3 The extinction probability

Let $\mathbf{Z}_n$ be a two-type branching process in an i.i.d. random environment with two
states. Let $\mathbf{T}$ denote the set of all two-dimensional vectors whose components are
non-negative integers.

Definition 2.3.1. *For $C \geq 1$ and N a positive integer, let $\mathcal{M}(C, N)$ be a class of 2×2 non-negative, finite matrices such that if $M_i \in \mathcal{M}(C, N)$ for $i = 1, \cdots, N$ and $M = M_1 M_2 \cdots M_N$, then*

(a) M is positive and

(b) $1 \leq [\max_{i,j} M(i, j)] / [\min_{i,j} M(i, j)] \leq C < \infty$.

Definition 2.3.2. *For $\mathbf{a} \in \mathbf{T}$, let $q(\mathbf{a})$ denote the extinction probability of the MBPRE given that $\mathbf{Z}_0 = \mathbf{a}$; that is,*

$$q(\mathbf{a}) = P\{\mathbf{Z}_n = \mathbf{0} \text{ for some } n \mid \mathbf{Z}_0 = \mathbf{a}\}.$$

Definition 2.3.3. *Consider the discrete parameter vector Markov process $\boldsymbol{X}_n$ taking values in $[0, 1]^2$ and defined as follows: for arbitrary, but fixed $\mathbf{s} \in [0, 1]^2$,*

$$
\begin{aligned}
\boldsymbol{X}_0 &= \mathbf{s}, \\
\boldsymbol{X}_{n+1} &= \phi(\zeta_n, \boldsymbol{X}_n),
\end{aligned}
$$

where $\phi = (\phi^{(1)}, \phi^{(2)})$ is the multivariate probability generating functions (mpgf) of offspring distribution. The stochastic process $\boldsymbol{X}_n$ will be called the dual process associated with the MBPRE.

Since the dual process takes only values in the unit square-cube, $\mathbf{X}_n$ converges in distribution to some random vector $\mathbf{X}$ as $n \to \infty$. Further, $\mathbf{X}$ takes only values in $[0, 1]^2$ and

$$q(\mathbf{a}) = E\left[\prod_{j=1}^{2} X_j^{a_j} \right] =: E[\mathbf{X^a}], \ \mathbf{a} \in \mathbf{T}.$$

We count the set $\mathbf{T}$ along the diagonal, i.e., $\mathbf{T} = \{(0, 0), (1, 0), (0, 1), (2, 0), (1, 1), (0, 2), \cdots\}$, renumber them as $\{0, 1, 2, 3, 4, 5, \cdots\}$. Denote $q(\mathbf{a})$ as q_i with $i = 1 + 2 + \cdots + \|\mathbf{a}\|_1 + a_2$ and write $P_k^{(i)} = P\{\mathbf{Z}_1 = \mathbf{b} \mid \mathbf{Z}_0 = \mathbf{a}\}$ with $k = 1 + 2 + \cdots + \|\mathbf{b}\|_1 + b_2$ and $i = 1 + 2 + \cdots + \|\mathbf{a}\|_1 + a_2$.

It is easily seen that the sequence $q_1, q_2, \cdots$ of extinction probabilities satisfies the

following system of equations [Wil69]

$$q_1 = P_0^{(1)} + P_1^{(1)}q_1 + P_2^{(1)}q_2 + \cdots + P_n^{(1)}q_n + \cdots,$$

$$q_2 = P_0^{(2)} + P_1^{(2)}q_1 + P_2^{(2)}q_2 + \cdots + P_n^{(2)}q_n + \cdots,$$

$$\cdots,$$

$$q_n = P_0^{(n)} + P_1^{(n)}q_1 + P_2^{(n)}q_2 + \cdots + P_n^{(n)}q_n + \cdots,$$

$$\cdots.$$

Suppose we write the first n equations from above as follows:

$$q_1 = P_0^{(1)} + P_1^{(1)}q_1 + P_2^{(1)}q_2 + \cdots + P_n^{(1)}q_n + c_n^{(1)},$$

$$q_2 = P_0^{(2)} + P_1^{(2)}q_1 + P_2^{(2)}q_2 + \cdots + P_n^{(2)}q_n + c_n^{(2)},$$

$$\cdots,$$

$$q_n = P_0^{(n)} + P_1^{(n)}q_1 + P_2^{(n)}q_2 + \cdots + P_n^{(n)}q_n + c_n^{(n)}.$$

We write

$$\mathbf{A}_n = \begin{pmatrix} P_1^{(1)} & P_2^{(1)} & \cdots & P_n^{(1)} \\ P_1^{(2)} & P_2^{(2)} & \cdots & P_n^{(2)} \\ \cdots & \cdots & \cdots & \cdots \\ P_1^{(n)} & P_2^{(n)} & \cdots & P_n^{(n)} \end{pmatrix}.$$

Write $\mathbf{q}_n = (q_1, q_2, \cdots, q_n)^\top$, $\mathbf{p}_n = (P_0^{(1)}, P_0^{(2)}, \cdots, P_0^{(n)})^\top$, and $\mathbf{c}_n = (c_n^{(1)}, c_n^{(2)}, \cdots, c_n^{(n)})^\top$.
Then the above equation can be written

$$(\mathbf{I} - \mathbf{A}_n)\mathbf{q}_n = \mathbf{p}_n + \mathbf{c}_n, \tag{2.3.5}$$

where $\mathbf{I}$ is the identity matrix.

Theorem 2.3.3. *Suppose the environmental random variable $\{\zeta_n\}$ takes only finitely
many states. Further, suppose that the conditional mean matrix $\{M(\zeta_n)\}$ takes only
values in some space $\mathcal{M}(C, N)$. Suppose that the largest Lyapunov exponent satisfies
$\lambda > 0$. Suppose that the mpgf of conditional offspring distribution satisfies $\phi_{\zeta_n}(s) > \mathbf{0}$*

for some state of ζ_n. Let $\{\gamma_j\}_{j \in \mathbb{N}}$ be a decreasing sequence of real numbers tending to zero. Suppose that

$$q(\mathbf{a}) \leq \gamma_{\|\mathbf{a}\|_1}, \ \text{for } \mathbf{a} \in \mathbf{T}.$$

Then for each $n = 2 + 3 + \cdots + (\|\mathbf{a}\|_1 + 1)$, the vector $\mathbf{q}_n$ of extinction probabilities satisfies the inequality

$$(\boldsymbol{I} - \boldsymbol{A}_n)^{-1} \mathbf{p}_n \leq \mathbf{q}_n \leq (\boldsymbol{I} - \boldsymbol{A}_n)^{-1} \mathbf{p}_n + \gamma_{\|\mathbf{a}\|_1 + 1} \mathbf{1}. \tag{2.3.6}$$

Proof. We follow the proof of Theorem 2.1 in [Wil69]. Since $\phi_{\zeta_n}(\mathbf{s}) > \mathbf{0}$ for some state of ζ_n, we have $P_0^{(k)} > 0$ for all $k = 1, 2, \cdots, n$, and it follows that $\mathbf{A}_n$ is a non-negative matrix whose row sums are all less than 1. From the theory of non-negative matrices, it follows that $\mathbf{A}_n$ has a real non-negative eigenvalue ρ_n such that no eigenvalue of $\mathbf{A}_n$ has modulus exceeding ρ_n. Further, if s_n and S_n denote respectively the minimum and maximum row sums of $\mathbf{A}_n$, then

$$s_n \leq \rho_n \leq S_n,$$

and, since $S_n < 1$, we have $\rho_n < 1$.

Since $\rho_n < 1$, it follows that $|\mathbf{I} - \mathbf{A}_n| \neq 0$. Hence $(\mathbf{I} - \mathbf{A}_n)^{-1}$ exists for all n, and furthermore

$$(\rho \mathbf{I} - \mathbf{A}_n)^{-1} \geq \mathbf{0}, \ \text{for } \rho > \rho_n.$$

In particular, therefore, $(\mathbf{I} - \mathbf{A}_n)^{-1} \geq \mathbf{0}$. We thus obtain

$$\mathbf{q}_n = (\mathbf{I} - \mathbf{A}_n)^{-1} [\mathbf{p}_n + \mathbf{c}_n].$$

From

$$q(\mathbf{a}) \leq \gamma_{\|\mathbf{a}\|_1},$$

it follows that for $n = 2 + \cdots + \|\mathbf{a}\|_1$ and for $i = 1, \cdots, n$,

$$
\begin{aligned}
\mathbf{c}_n^{(i)} \ &\leq \ \gamma_{\|\mathbf{a}\|_1 + 1} \left(1 - \sum_{r=0}^{n} P_r^{(i)} \right) \\
&< \ \gamma_{\|\mathbf{a}\|_1 + 1} \left(1 - \sum_{r=1}^{n} P_r^{(i)} \right) \\
&= \ \gamma_{\|\mathbf{a}\|_1 + 1} r_n^{(i)}
\end{aligned}
$$

44

where $r_n^{(i)}$, is the ith row sum of $(\mathbf{I} - \mathbf{A}_n)$.

Since $(\mathbf{I} - \mathbf{A}_n)^{-1}$ is a non-negative matrix, and $\mathbf{c}_n$ is a non-negative vector,

$$(\mathbf{I} - \mathbf{A}_n)^{-1}\mathbf{p}_n \le \mathbf{q}_n \le (\mathbf{I} - \mathbf{A}_n)^{-1}[\mathbf{p}_n + \gamma_{\|\mathbf{a}\|_1 + 1}\mathbf{r}_n], \tag{2.3.7}$$

where $\mathbf{r}_n = (r_n^{(1)}, r_n^{(2)}, \cdots, r_n^{(n)})^\top$. But $\mathbf{r}_n = (\mathbf{I} - \mathbf{A}_n)\mathbf{1}$, so we get

$$(\mathbf{I} - \mathbf{A}_n)^{-1}\mathbf{p}_n \le \mathbf{q}_n \le (\mathbf{I} - \mathbf{A}_n)^{-1}\mathbf{p}_n + \gamma_{\|\mathbf{a}\|_1 + 1}\mathbf{1}. \tag{2.3.8}$$

$\square$

Corollary 2.3.1. *Suppose the mpgf $\phi_r^{(i)}(\mathbf{s})$, $i, r = 1, 2$ are all such that for some $\gamma < 1$,*

$$\phi_r^{(i)}(s, s) \le s, \ \text{for } s \ge \gamma. \tag{2.3.9}$$

Then for $n = 2 + 3 + \cdots + (\|\mathbf{a}\|_1 + 1)$,

$$(\boldsymbol{I} - \boldsymbol{A}_n)^{-1}\mathbf{p}_n \le \mathbf{q}_n \le (\boldsymbol{I} - \boldsymbol{A}_n)^{-1}\mathbf{p}_n + \gamma^{\|\mathbf{a}\|_1 + 1}\mathbf{1}. \tag{2.3.10}$$

Proof.

$$
\begin{aligned}
q(\mathbf{a}) &= E[\mathbf{X}^\mathbf{a}] \\
&= \lim_{n \to \infty} E[\mathbf{X}_n^\mathbf{a} \mid \mathbf{X}_0 = \mathbf{0}] \\
&\le \lim_{n \to \infty} E[\mathbf{X}_n^\mathbf{a} \mid \mathbf{X}_0 = (\gamma, \gamma)] \\
&\le \gamma^{\|\mathbf{a}\|_1}.
\end{aligned}
$$

$\square$

Remark: Bourgin and Cogburn deduced an error estimate of the approximate extinction probability in the case that the branching chain is one-dimensional and the environmental process is a two-state Markov chain, see (Theorem 7, [BC81]). A direct generalization of their result to two-dimensional branching chain is the following. Let $q(r, \mathbf{a})$ be the extinction probability of the process started at $(r, \mathbf{a})$ and $q^{(n)}(r, \mathbf{a})$ be the approximation of $q(r, \mathbf{a})$ for $n \in \{2 + 3 + \cdots + (\|\mathbf{a}\|_1 + 1) \mid \mathbf{a} \in \mathbf{T} \setminus \{\mathbf{0}\}\}$. Let $\phi_r^{(i)}(\mathbf{s})$ denote the probability generating function for the offspring distribution of a single particle of type i in environment r and $\gamma_r^{(i)}$ denote the smallest solution of $\phi_r^{(i)}(s, s) = s$.

45

Theorem 2.3.4. *If* $\gamma = \max\left\{\gamma_1^{(1)}, \gamma_1^{(2)}, \gamma_2^{(1)}, \gamma_2^{(2)}\right\} < 1$, *then* $q(r, \mathbf{a}) \leq \phi_r(\gamma, \gamma)^{\mathbf{a}} \equiv$ $\phi_r^{(1)}(\gamma, \gamma)^{a_1} \phi_r^{(2)}(\gamma, \gamma)^{a_2}$ *and* $\delta_n = \|q - q^{(n)}\| \leq \gamma^{\|\mathbf{a}\|_1 + 1}$, *for* $n = 2 + 3 + \cdots + (\|\mathbf{a}\|_1 + 1)$.

Example 8: Let the fitness matrix be given by $\begin{pmatrix} 1.8 & 1.5 \\ 1.5 & 1.6 \end{pmatrix}$ and the probability of environment 1 be given by $p = 0.28$. The mpgf are

$$
\begin{aligned}
\phi_1^{(1)}(\mathbf{s}) &= 0.1 + 0.9(s_1 x + s_2(1 - x))^2 \\
\phi_1^{(2)}(\mathbf{s}) &= 0.25 + 0.75(s_1 x + s_2(1 - x))^2 \\
\phi_2^{(1)}(\mathbf{s}) &= 0.25 + 0.75(s_1 x + s_2(1 - x))^2 \\
\phi_2^{(2)}(\mathbf{s}) &= 0.2 + 0.8(s_1 x + s_2(1 - x))^2.
\end{aligned}
$$

We want to find the optimal x that minimizes $(\mathbf{I} - \mathbf{A}_n)^{-1} \mathbf{p}_n$. By numerical computation, we find that $x_q^{opt} \approx 0.45$ (while $x_\lambda^{opt} \approx 0.88$).

Chapter 3

Variational characterization

In this chapter we develop a variational principle for models that are described by multitype Galton-Watson branching processes. We take the point of view of individual lineages through time, their ancestry and genealogy. First of all, the asymptotic properties of the branching process forward in time are, to a large extent, determined by the principle eigenvalue ρ of the mean matrix M, the associated right eigenvector u and left eigenvector v (Theorem 3.1.2). We introduce the retrospective process and identify it with the mutation process describing the type evolution along typical lineages of the branching process (Theorem 3.2.1). Furthermore, the asymptotic growth rate $\log \rho$ of the branching process satisfies a variational principle and the distribution of the retrospective process is the unique equilibrium state for the variational characterization (Theorem 3.3.1).

3.1   The asymptotic properties

In this section, we use the same notations for multitype Galton-Watson branching processes as in section 2.1. Let the type space be given by $S = \{1, \cdots, d\}$. Let $\{\mathbf{Z}_n = (Z_n^1, \cdots, Z_n^d)^\top\}$ denote a d-type Galton-Watson branching process where Z_n^i stands for the size of type i individuals in generation n with $i = 1, \cdots, d$. Define $M = (m_{ij})_{i,j \in S}$ to be the matrix of expected numbers of progeny. We will assume throughout that the mean matrix M of the branching process is positive (i.e. each

component satisfies $m_{ij} > 0$). Note that we make a stronger assumption on M than that in section 2.1. Perron-Frobenius theory then tells us that M has a principal eigenvalue ρ and associated positive right and left eigenvectors u and v which will be normalized so that $\langle u, v \rangle = 1$ and $\langle v, \mathbf{1} \rangle = 1$. We will further assume that $\rho > 1$, i.e., the branching process is supercritical. Write $\lambda = \log \rho$.

The Kesten-Stigum theorem ([KS66],[AN72]) says the following:

Theorem 3.1.1. *There is a scalar random variable W such that*

$$\lim_{n \to \infty} \frac{\boldsymbol{Z}_n}{\rho^n} = Wv \ a.s. \tag{3.1.1}$$

and $P\{W > 0\} > 0$ iff

$$E^i[\boldsymbol{Z}_1^j \log \boldsymbol{Z}_1^j] < \infty \ \text{for all} \ 1 \leq i, j \leq d. \tag{3.1.2}$$

When (3.1.2) holds,

$$E^i[W] = u_i, \ \ i = 1, \cdots, d. \tag{3.1.3}$$

Remark: The Kesten-Stigum theorem also holds for one-type Galton-Watson branching processes in random environments. However, a generalization to multi-type Galton-Watson branching processes in random environments is not known.

The asymptotic properties of the branching process forward in time are, to a large extent, determined by ρ, u and v. As is shown in the following theorem, $\log \rho$ is equal to the growth rate of the population size, u is a measure of the relative contribution made to the stationary population by each type, and v is the stationary composition of the population.

Theorem 3.1.2. *Assume $E^i[\boldsymbol{Z}_1^j \log \boldsymbol{Z}_1^j] < \infty$, for all $1 \leq i, j \leq d$. Then*

$$\lim_{n \to \infty} \frac{1}{n} \log \|\boldsymbol{Z}_n\|_1 = \log \rho = \lambda, \ \ a.s. \ conditioned \ on \ non\text{-}extinction, \tag{3.1.4}$$

$$\lim_{n \to \infty} \frac{E^i[\|\boldsymbol{Z}_n\|_1]}{\rho^n} = u_i, \ for \ all \ 1 \leq i, j \leq d, \ and \tag{3.1.5}$$

$$\lim_{n \to \infty} \frac{\boldsymbol{Z}_n}{\|\boldsymbol{Z}_n\|_1} = v, \ \ a.s. \ conditioned \ on \ non\text{-}extinction. \tag{3.1.6}$$

Proof. By the Kesten-Stigum theorem, conditioned on non-extinction,

$$\lim_{n\to\infty}\frac{1}{n}\log\|\mathbf{Z}_n\|_1$$
$$=\lim_{n\to\infty}\frac{1}{n}\log(\rho^n\langle v,\mathbf{1}\rangle W)$$
$$=\log\rho+\lim_{n\to\infty}\frac{1}{n}\log W$$
$$=\log\rho.$$

Similarly,

$$\lim_{n\to\infty}\frac{E^i[\|\mathbf{Z}_n\|_1]}{\rho^n}$$
$$=\langle v,\mathbf{1}\rangle E^i[W]$$
$$=u_i.$$

As is shown in Theorem 2 of [KLP97],

$$\lim_{n\to\infty}\frac{\mathbf{Z}_n}{\|\mathbf{Z}_n\|_1}=v.$$

$\square$

3.2   The retrospective process

We consider the type process along a representative line of the multitype Galton-Watson branching process with positive mean matrix M. Specially, the so-called retrospective process ([Jag89],[Jag92],[GB03]), defined below, plays an important role in the description of the equilibrium state for the variational characterization.

Definition 3.2.1. *The retrospective process $\{\eta_n\}$ is an S-valued Markov chain with transition matrix G whose ijth component is given by $g_{ij}=\frac{u_j m_{ij}}{\rho u_i}$ and corresponding stationary distribution α whose ith component is equal to $\alpha_i=u_i v_i$.*

The retrospective process was introduced by Jagers ([Jag89], p.195) and may be interpreted as the forward version of the backward Markov chain ([JN96], Proposition 1) that results from picking individuals randomly from the stationary type distribution v and following their lines of descent backward in time. As observed by Jagers

([Jag92], Corollary 1), and Jagers and Nerman ([JN96], Proposition 1), the probability vector α describes the distribution of ancestral types of an equilibrium population with type frequencies given by v.

Let $\Sigma = S^{\mathbb{Z}_+}$ be trajectory space of the stationary doubly infinite retrospective process. And let σ be a shift map on Σ. Then $\{\eta_n\}$ may be viewed as a stationary ergodic process with σ-invariant measure μ whose finite dimensional distribution is given by $\mu(\eta_j = i_0, \cdots, \eta_{j+n} = i_n) = \alpha_{i_0} \prod_{k=1}^{n} g_{i_{k-1} i_k}$. Our result below shed some light on the significance of the distribution μ of the doubly infinite retrospective process. To state this result we need some preparations.

We start with a description of the multitype Galton-Watson family tree according to ([Har63], Chapter VI) and [GB03]. Let $\mathbb{X} = \bigcup_{n \geq 0} \mathbb{X}_n$, where $\mathbb{X}_n$ describes the virtual nth generation. That is, $\mathbb{X}_0 = S$, and $i_0 \in \mathbb{X}_0$ specifies the type of the root, i.e. the founding ancestor. Next, $\mathbb{X}_1 = S \times \mathbb{N}$, and the element $x = (i_1; l_1) \in \mathbb{X}_1$ is the l_1th child of type i_1 of the root. Finally, for $n > 1$, $\mathbb{X}_n = S^n \times \mathbb{N}^n$, and $x = (i_1, \cdots, i_n; l_1, \cdots, l_n) \in \mathbb{X}_n$ is the l_nth child of type i_n of its parent $\tilde{x} = (i_1, \cdots, i_{n-1}; l_1, \cdots, l_{n-1})$. We write $t(x) = i_n$ for the type of $x \in \mathbb{X}_n$. With each $x \in \mathbb{X}$ we associate its random offspring $N_x = [N_{x,j}]_{j \in S} \in \mathbb{Z}_+^d$ with distribution $p^{t(x)}$ such that the family $\{N_x : x \in \mathbb{X}\}$ is independent. The random variables N_x indicate which of the virtual individuals $x \in \mathbb{X}$ are actually realized, namely those in the random set $X = \bigcup_{n \geq 0} X_n$ defined recursively by

$$X_0 = i_0, \ X_n = \{x = (\tilde{x}; i_n, l_n) \in \mathbb{X}_n : \tilde{x} \in X_{n-1}, l_n \leq N_{\tilde{x}, i_n}\},$$

where i_0 is the prescribed type of the root. The family tree is completely determined by the process $\{X_n\}_{n \geq 0}$. We write P^i for the distribution of $\{X_n\}_{n \geq 0}$ and E^i for the associated expectation, when the type of the root is $i_0 = i$. The branching process refers to the counting measures

$$\mathbf{Z}_n = \sum_{x \in X_n} \delta_{t(x)},$$

on S, where δ_i is the Dirac measure at i.

We then introduce the time-averaged type evolution process of an individual in the

population tree. For $n > 0$ and $x \in X_n$, let $t(x)_{n, \ per} \in \Sigma$ be defined by

$$t(x)_{n, \ per}(m) = t(x(m_n)), \ m \in \mathbb{Z},$$

where m_n is the unique integer in $\{0, 1, \cdots, n-1\}$ with $m \equiv m_n \bmod n$, and $x(m_n)$ is the unique ancestor of x living at time m_n. The time-averaged type evolution of x is then described by the empirical type evolution process

$$R_n^x = \frac{1}{n} \sum_{m=0}^{n-1} \delta_{\sigma^m t(x)_{n, \ per}} \in \mathcal{P}_\sigma(\Sigma),$$

where $\mathcal{P}_\sigma(\Sigma)$ is the set of all probability measures on Σ which are invariant under the shift σ.

We are interested in the typical behavior of R_n^x when x is picked at random from the population X_n at time n. This is captured in the empirical distribution of the mapping $x \mapsto R_n^x$ from X_n to $\mathcal{P}_\sigma(\Sigma)$, i.e. the population average

$$\Gamma_n := \frac{1}{|X_n|} \sum_{x \in X_n} \delta_{R_n^x},$$

where $|X_n|$ denotes the total size of the population at time n and $|X_n| = \|\mathbf{Z}_n\|_1$. This definition requires that $X_n \neq \emptyset$. The population average Γ_n is a random element of the set $\mathcal{P}(\mathcal{P}_\sigma(\Sigma))$ of all probability measures on the space $\mathcal{P}_\sigma(\Sigma)$. Then the theorem below identifies the process describing the type evolution along typical lineages.

Theorem 3.2.1. *Let $\lambda > 0$ and $i \in S$. Then*

$$\lim_{n \to \infty} \Gamma_n = \delta_\mu, \ P^i\text{-almost surely conditioned on survival,}$$

where $\mu \in \mathcal{P}_\sigma(\Sigma)$ is the distribution of the stationary doubly infinite retrospective process $\{\eta_n\}$.

Remark: Georgii and Baake [GB03] proved this result for multitype Markov branching processes in continuous time. We show the discrete-time version here.

A crucial ingredient in the proof is a representation of the family tree in terms of a size-biased tree with trunk (with the retrospective process running along the trunk).

The size-biased tree $\{\hat{X}_n\}_{n\geq 0}$ was introduced by Lyons *et al.* [LPP95] and Kurtz *et al.* [KLP97], and it is constructed as follows: Set the size-biased offspring distribution

$$\hat{p}_{\mathbf{k}}^i := \frac{p_{\mathbf{k}}^i \langle \mathbf{k}, u \rangle}{\rho u_i},$$

where $\langle \mathbf{k}, u \rangle = \sum_j k_j u_j$. Start with one individual ξ_0 of type i. Generate offspring according to the distribution $\hat{p}_{\mathbf{k}}^i$. Pick one of these offspring ξ_1 as the successor on the trunk, with children being picked with probability proportional to u_j when their type is j. The successor ξ_1 gets an independent number of offspring according to the probability $\hat{p}_{\mathbf{k}}^j$ if its type is j, while the other children that do not belong to the trunk get ordinary descendant trees. Again, pick one of the offspring of ξ_1 as a further successor of the trunk, call it ξ_2, and give the others ordinary descendant trees, and so on. Define the measure $\hat{P}_*^i$ as the joint distribution of the random sized-biased tree $\{\hat{X}_n\}_{n\geq 0}$ and the random trunk $(\xi_0, \xi_1, \xi_2, \cdots)$, and $\hat{P}^i$ for its marginal on the space of trees. Note that the type process along the trunk, $\eta_n = t(\xi_n)$, is a copy of the retrospective process. The lemma below establishes the relationship between P^i, $\hat{P}_*^i$, and the retrospective process. We use the shorthand $y[0,n]$ for a path $(y_0, y_1, \cdots, y_n)$ and for any $x \in X_n$, the path $x[0,n]$ denotes the path starting from the root, not backtracking, through x. Denote $\mathcal{B}_f(\mathbb{X})$ for the set of all finite subset of $\mathbb{X}$.

Lemma 3.2.1. *Let $n > 0$ and $i \in S$, and let $F : \mathcal{B}_f(\mathbb{X})^{n+1} \times \mathbb{X}^{n+1} \to [0,\infty)$ be any measurable function. Then we have*

$$\hat{E}_*^i \left[F(\hat{X}[0,n], \xi[0,n]) \right] = \frac{1}{\rho^n u_i} E^i \left[\sum_{x \in X_n} F(X[0,n], x[0,n]) u_{t(x)} \right].$$

Proof. The result follows from the measure change in [KLP97]. $\square$

Another ingredient is the Donsker-Varadhan large deviation principle for the retrospective process [Var88][DS89]. For every $\nu \in \mathcal{P}_\sigma(\Sigma)$, let

$$H_G(\nu) = \sup_{n \geq 1} h(\nu_{[0,n]}; \nu_{[0,n-1]} \otimes_n G)$$

be the process-level large deviation rate function for the retrospective process with transition matrix G. In the above, $\nu_{[0,n]}$ is the restriction of ν to the sequence $[0,n]$,

$\nu_{[0,n-1]} \otimes_n G$ is the product of the measure $\nu_{[0,n-1]}$ and the transition kernel G, and $h(\nu_{[0,n]}; \nu_{[0,n-1]} \otimes_n G)$ is their relative entropy. See ([DS89], Equation 4.4.11).

Lemma 3.2.2. *For the empirical type evolution process of the retrospective process R_n^ξ, we have for $i \in S$ and closed $F \subset \mathcal{P}_\sigma(\Sigma)$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log \hat{P}_*^i \left\{ R_n^\xi \in F \right\} \leq - \inf_{\nu \in F} H_G(\nu),$$

while, for open $O \subset \mathcal{P}_\sigma(\Sigma)$,

$$\liminf_{n \to \infty} \frac{1}{n} \log \hat{P}_*^i \left\{ R_n^\xi \in O \right\} \geq - \inf_{\nu \in O} H_G(\nu).$$

Moreover, the function H_G is lower semicontinuous with compact level sets and attains its minimum 0 precisely at μ.

These two ingredients together imply a large deviation principle for the typical type evolution along the surviving lineages in the tree as follows.

Lemma 3.2.3. *For the empirical type evolution process R_n^x, we have for $i \in S$ and closed $F \subset \mathcal{P}_\sigma(\Sigma)$,*

$$\limsup_{n \to \infty} \frac{1}{n} \log E^i \left[\sum_{x \in X_n} 1_{\{R_n^x \in F\}} \right] \leq \lambda - \inf_{\nu \in F} H_G(\nu),$$

while, for open $O \subset \mathcal{P}_\sigma(\Sigma)$,

$$\liminf_{n \to \infty} \frac{1}{n} \log E^i \left[\sum_{x \in X_n} 1_{\{R_n^x \in O\}} \right] \geq \lambda - \inf_{\nu \in O} H_G(\nu).$$

Moreover, the function H_G is lower semicontinuous with compact level sets and attains its minimum 0 precisely at μ.

Proof. In view of Lemma 3.2.1, for every closed $F \subset \mathcal{P}_\sigma(\Sigma)$ we have

$$
\begin{aligned}
& E^i \left[\sum_{x \in X_n} 1_{\{R_n^x \in F\}} \right] \\
= \ & u_i \rho^n \hat{E}_*^i \left[1_{\{R_n^\xi \in F\}} u_{t(\xi_n)}^{-1} \right] \\
\leq \ & u_i \rho^n \max_{j \in S} u_j^{-1} \hat{P}_*^i \left\{ R_n^\xi \in F \right\}.
\end{aligned}
$$

Then applying Lemma 3.2.2, we get

$$\limsup_{n\to\infty} \frac{1}{n} \log E^i \left[\sum_{x\in X_n} 1_{\{R_n^x \in F\}} \right] \leq \lambda - \inf_{\nu\in F} H_G(\nu).$$

Similarly, for every open $O \subset \mathcal{P}_\sigma(\Sigma)$

$$\liminf_{n\to\infty} \frac{1}{n} \log E^i \left[\sum_{x\in X_n} 1_{\{R_n^x \in O\}} \right] \geq \lambda - \inf_{\nu\in O} H_G(\nu).$$

$\square$

We are now ready for the proof of Theorem 3.2.1.

Proof of Theorem 3.2.1. Let d be a metric on $\mathcal{P}_\sigma(\Sigma)$. For any fixed $\varepsilon > 0$, we consider the set $C = \{\nu \in \mathcal{P}_\sigma(\Sigma) : d(\nu, \mu) \geq \varepsilon\}$, which is the complement of the open ε-neighborhood of μ. Since C is closed and H_G has compact level sets and attains its minimum 0 only at μ, the infimum $c := \inf_{\nu\in C} H_G(\nu)$ is strictly positive. We can therefore choose a constant γ such that $\lambda - c < \log \gamma < \lambda$. Then we write

$$\frac{1}{|X_n|} \sum_{x\in X_n} 1_{\{R_n^x \in C\}}$$

$$= \left(\frac{\gamma^n}{|X_n|} \right) \left(\gamma^{-n} \sum_{x\in X_n} 1_{\{R_n^x \in C\}} \right)$$

and show that each factor tends to 0 almost surely conditioned on survival.

On the one hand, by Lemma 3 and 4 in [KLP97],

$$\lim_{n\to\infty} \frac{1}{Z_n^j} \sum_{x\in X_n^j} |X(x, n+1)| = E^j[|X_1|] \text{ a.s.,}$$

where $|X(x, n+1)|$ denotes the size of descendants of x living at time $n+1$. Then,

$$
\begin{aligned}
\frac{|X_{n+1}|}{|X_n|} &= \sum_{j\in S} \frac{Z_n^j}{\|\mathbf{Z}_n\|_1} \frac{1}{Z_n^j} \sum_{x\in X_n^j} |X(x, n+1)| \\
&\to \sum_{j\in S} v_j E^j[|X_1|] \\
&= \rho = e^\lambda,
\end{aligned}
$$

as $n \to \infty$ a.s. conditioned on survival. Hence $\frac{\gamma^n}{|X_n|} \sim (\frac{\gamma}{\rho})^n \to 0$ a.s. conditioned on survival.

On the other hand, we obtain for any $a > 0$ that

$$
\begin{aligned}
& \limsup_{n \to \infty} \frac{1}{n} \log P^i \left\{ \gamma^{-n} \sum_{x \in X_n} 1_{\{R_n^x \in C\}} \geq a \right\} \\
\leq \quad & \limsup_{n \to \infty} \frac{1}{n} \log E^i \left[\gamma^{-n} \sum_{x \in X_n} 1_{\{R_n^x \in C\}} \right] \\
\leq \quad & -\log \gamma + \lambda - c < 0.
\end{aligned}
$$

Then $\sum_n P^i \{ \gamma^{-n} \sum_{x \in X_n} 1_{\{R_n^x \in C\}} \geq a \} < \infty$ and Borel-Cantelli lemma [Kal01] shows that $\gamma^{-n} \sum_{x \in X_n} 1_{\{R_n^x \in C\}} \to 0$ as $n \to \infty$ a.s.

Hence, P^i-a.s. conditioned on survival, $\Gamma_n(F) \to 0$ for each closed $F \subset \mathcal{P}_\sigma(\Sigma)$ with $\mu \notin F$. According to the portmanteau theorem ([EK86], Theorem 3.1), $\lim_{n \to \infty} \Gamma_n = \delta_\mu$, P^i-a.s. conditioned on survival. $\square$

3.3 Variational characterization of the growth rate

Let $h_\nu(\sigma)$ denote the Kolmogorov-Sinai entropy for the shift σ with respect to $\nu \in \mathcal{P}_\sigma(\Sigma)$. The entropy of the stationary retrospective process can be computed as follows (cf. 4.27 in [Wal82] or example 5 on page 136 in [Jos05]).

Lemma 3.3.1. *The Kolmogorov-Sinai entropy of the retrospective process $\{\eta_n\}$ is*

$$
h_\mu(\sigma) = -\sum_{i,j \in S} \alpha_i g_{ij} \log g_{ij}.
$$

As is in [Dem83], the quantity $h_\mu(\sigma)$ describes the rate of increase of the number of typical genealogies. Roughly speaking $h_\mu(\sigma)$ corresponds to the total number of reproductives in the population. Let $[\eta_n]$ denote the sequence $\{\eta_0, \eta_1, \cdots, \eta_n\}$. Then $\mu([\eta_n]) = \alpha_{\eta_0} \prod_{k=0}^{n-1} g_{\eta_k \eta_{k+1}}$, a.s.

Lemma 3.3.2.

$$
\lim_{n \to \infty} -\frac{1}{n} \log \mu([\eta_n]) = h_\mu(\sigma), \;\; a.s.
$$

Proof. By the ergodic theorem [Kal01], we have

$$\lim_{n \to \infty} -\frac{1}{n} \log \mu([\eta_n])$$

$$= \lim_{n \to \infty} -\frac{1}{n} \left(\log \alpha_{\eta_0} + \sum_{k=0}^{n-1} \log g_{\eta_k \eta_{k+1}} \right)$$

$$= -E[\log g_{\eta_0 \eta_1}]$$

$$= -\sum_{i,j \in S} \alpha_i g_{ij} \log g_{ij}$$

$$= h_\mu(\sigma), \text{ a.s.}$$

$\square$

If the population is sufficiently large and will survives forever, then the law of large number theory tells us that the population can be approximated by the deterministic dynamics $E[\mathbf{Z}_{n+1}^\top] = E[\mathbf{Z}_n^\top]M$. Consequently, the shift on Σ, which is only concerned with the genealogical history of living individuals in the population, corresponds to the asymptotic dynamics defined by the matrix M. Motivated by this fact, we define a potential function $\varphi_M\left(\vec{i}\right) := \log m_{i_0 i_1}$, for any sequence $\vec{i} \in \Sigma$. Furthermore, define the reproductive potential with respect to $\nu \in \mathcal{P}_\sigma(\Sigma)$ as $\Phi_\nu(M) = \int_\Sigma \log m_{i_0 i_1} d\nu\left(\vec{i}\right)$.

Lemma 3.3.3. *The reproductive potential of $\{\eta_n\}$ is*

$$\Phi_\mu(M) = \sum_{i,j \in S} \alpha_i g_{ij} \log m_{ij}.$$

Proof. The result follows immediately from the finite dimensional distribution of μ.

$\square$

Then the growth rate $\lambda := \log \rho$ of the branching process, the entropy $h_\mu(\sigma)$ of the retrospective process and reproductive potential $\Phi_\mu(M)$ of the retrospective process satisfies the following relation.

Lemma 3.3.4.

$$\lambda = h_\mu(\sigma) + \Phi_\mu(M).$$

Proof. By Lemma 3.3.1 and Lemma 3.3.3, we have

$$
\begin{aligned}
& h_\mu(\sigma) + \Phi_\mu(M) \\
=\ & -\sum_{i,j \in S} \alpha_i g_{ij} \log g_{ij} + \sum_{i,j \in S} \alpha_i g_{ij} \log m_{ij} \\
=\ & -\sum_{i,j \in S} \alpha_i g_{ij} \log \frac{u_j m_{ij}}{\rho u_i} + \sum_{i,j \in S} \alpha_i g_{ij} \log m_{ij} \\
=\ & \log \rho + \sum_{i \in S} \alpha_i \log u_i \left(\sum_{j \in S} g_{ij} \right) - \sum_{j \in S} \left(\sum_{i \in S} \alpha_i g_{ij} \right) \log u_j \\
=\ & \lambda + \sum_{i \in S} \alpha_i \log u_i - \sum_{j \in S} \alpha_j \log u_j \\
=\ & \lambda.
\end{aligned}
$$

$\square$

By invoking the thermodynamic formalism described in [Rue78] and following the proof of Theorem 4.3 in [AGD94], we will show that the asymptotic growth rate of the branching process satisfies a variational principle which is formally analogous to the minimization of the free energy in statistical mechanics.

Theorem 3.3.1. *The growth rate λ of the branching process satisfies an extremal principle:*

$$
\lambda = \sup_{\nu \in \mathcal{P}_\sigma(\Sigma)} \{ h_\nu(\sigma) + \Phi_\nu(M) \}
$$

and the distribution μ of the retrospective process is the unique measure of $\mathcal{P}_\sigma(\Sigma)$ such that

$$
\lambda = h_\mu(\sigma) + \Phi_\mu(M).
$$

For the proof we will need the following simple information theory result.

Lemma 3.3.5. *If $(p_1, \cdots, p_d)$ and $(q_1, \cdots, q_d)$ are two probability vectors and $p_i > 0$ for $1 \le i \le d$, then*

$$
\sum_{i=1}^d q_i \log p_i - \sum_{i=1}^d q_i \log q_i \le 0
$$

with equality if and only if $p_i = q_i$ for $1 \le i \le d$.

57

Proof of Theorem 3.3.1. For $\nu \in \mathcal{P}_\sigma(\Sigma)$, choose

$$p_j\left(\vec{i}\right) = \frac{\alpha_j g_{j,i_1}}{\alpha_{i_1}}$$

and

$$q_j\left(\vec{i}\right) = E_\nu\left[\chi_{C_j} \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right]\left(\vec{i}\right),$$

where $E_\nu\left[\cdot \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right]$ denotes the conditional expectation w.r.t. $\sigma^{-1}(\mathcal{B}(\Sigma))$, $C_j :=$ $\left\{\vec{i} \in \Sigma \mid i_0 = j\right\}$ and χ_C denotes the characteristic function corresponding to a set C.

Then we deduce from the Lemma 3.3.5 by integration that

$$\int_\Sigma \sum_j q_j\left(\vec{i}\right) \log p_j\left(\vec{i}\right) d\nu\left(\vec{i}\right) - \int_\Sigma \sum_j q_j\left(\vec{i}\right) \log q_j\left(\vec{i}\right) d\nu\left(\vec{i}\right) \leq 0.$$

On the one hand, it follows that

$$\int_\Sigma \sum_j q_j\left(\vec{i}\right) \log p_j\left(\vec{i}\right) d\nu\left(\vec{i}\right)$$

$$= \int_\Sigma \sum_j E_\nu\left[\chi_{C_j} \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right]\left(\vec{i}\right) \log \frac{\alpha_j g_{j,i_1}}{\alpha_{i_1}} d\nu\left(\vec{i}\right)$$

$$= \sum_j \int_\Sigma \chi_{C_j}\left(\vec{i}\right) \log \frac{\alpha_j g_{j,i_1}}{\alpha_{i_1}} d\nu\left(\vec{i}\right)$$

$$= \sum_j \int_{C_j} \log g_{j,i_1} d\nu\left(\vec{i}\right)$$

$$= \Phi_\nu(M) - \lambda.$$

On the other hand, we have

$$\int_\Sigma \sum_j q_j\left(\vec{i}\right) \log q_j\left(\vec{i}\right) d\nu\left(\vec{i}\right)$$

$$= \int_\Sigma \sum_j E_\nu\left[\chi_{C_j} \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right]\left(\vec{i}\right) \log E_\nu\left[\chi_{C_j} \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right] d\nu\left(\vec{i}\right)$$

$$= \int_\Sigma \sum_j \chi_{C_j}\left(\vec{i}\right) \log E_\nu\left[\chi_{C_j} \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right] d\nu\left(\vec{i}\right)$$

$$= -h_\nu(\sigma).$$

Hence, we obtain that

$$\lambda \geq h_\nu(\sigma) + \Phi_\nu(M). \tag{3.3.1}$$

The equality in (3.3.1) holds if and only if $p_j\left(\vec{i}\right) = q_j\left(\vec{i}\right)$ $\nu-$a.s. This condition is equivalent to

$$\int_{\Sigma} f\left(\vec{i}\right) \frac{\alpha_j g_{j,i_1}}{\alpha_{i_1}} d\nu\left(\vec{i}\right) = \int_{\Sigma} f\left(\vec{i}\right) E_\nu\left[\chi_{C_j} \mid \sigma^{-1}(\mathcal{B}(\Sigma))\right]\left(\vec{i}\right) d\nu\left(\vec{i}\right)$$

for all $\mathcal{B}(\Sigma)-$ measurable functions f. If we choose in particular $f = \chi_{C_{a_1,\cdots,a_n}}$, where $C_{a_1,\cdots,a_n} := \left\{\vec{i} \in \Sigma \mid i_1 = a_1, \cdots, i_n = a_n\right\}$, then we obtain the condition

$$\nu(C_{j,a_1,\cdots,a_n}) = \frac{\alpha_j g_{j,a_1}}{\alpha_{a_1}} \nu(C_{a_1,\cdots,a_n})$$

By iteration this yields that

$$\nu(C_{j,a_1,\cdots,a_n}) = \alpha_j g_{j,a_1} \cdots g_{a_{n-1}a_n}.$$

Hence the resulting measure is equal to μ. $\qquad\qquad\square$

Chapter 4

Continuous approximation

In this chapter, we firstly construct a multitype Markov branching process in a random environment (Proposition 4.1.1). When conditioned on the random environment, we deduce the Kolmogorov equations and the mean matrix for the conditioned branching process (Proposition 4.2.1-4.2.3). Then we introduce a parallel mutation-selection Markov branching process in a random environment and analyze its instability property (Theorem 4.3.2). Finally, we deduce a weak convergence result for a sequence of the parallel Markov branching processes in random environments (Theorem 4.4.1) and give examples for applications.

4.1 Multitype Markov branching processes in random environments

In the multitype Galton-Watson branching process the lifetime of each individual was one unit of time. A natural generalization is to allow these lifetimes to be random variables. Instead of the discrete time multitype Galton-Watson branching process in a random environment, now we consider a continuous time multitype Markov branching process in a random environment, denoted by $\{\mathbf{Z}(t)\}_{t \geq 0}$. We postulate that when conditioned on the random environment, $\{\mathbf{Z}(t)\}_{t \geq 0}$ behaves as a continuous time non-homogeneous vector-valued Markov branching process. Processes of this type are

discussed in Chap. 5 of [Har63] and the references therein. However, a generalization to multitype Markov branching processes in random environments is not known. We will give a description and construction of such processes.

Firstly, we give some notations. Let $(\Omega, \mathcal{F}, P, \{\mathcal{F}_t\}_{t \geq 0})$ be a complete filtered probability space. Let E denote the expectation operator for P. Let the type space be given by $S = \{1, \cdots, d\}$. Let $\{\mathbf{Z}(t) = (Z^1(t), \cdots, Z^d(t))^\top\}_{t \geq 0}$ denote a multitype Markov branching process in a random environment defined on $(\Omega, \mathcal{F}, P)$, where $Z^i(t)$ stands for the number of individuals of type i at time t for $i \in S$. Assume $\mathbf{Z}(0)$ is non-random. Let $\{\eta(t)\}_{t \geq 0}$ denote the environmental process of the Markov branching process defined on $(\Omega, \mathcal{F}, P)$.

Due to the random environment, both the rate of splitting $\lambda^{(i)}$ and offspring distribution $\{p_\gamma^{(i)}\}_{\gamma \in \mathbb{Z}_+^d}$ of an individual of type i are measurable functions of the environmental process $\{\eta(t)\}_{t \geq 0}$, where $\mathbb{Z}_+^d = \{\mathbf{n} = (n_1, \cdots, n_d)^\top \mid n_i \text{ nonnegative integer}\}$. Hence they are themselves stochastic processes on $(\Omega, \mathcal{F}, P)$. For each $t \geq 0$ and $\omega \in \Omega$, $\lambda^{(i)}(t, \omega)$ denotes the rate of splitting of an individual of type i at time t under a realization ω, i.e., an individual of type i has a probability $\lambda^{(i)}(t, \omega) \Delta t + o(\Delta t)$ of dying in the interval $(t, t + \Delta t)$ under a realization ω. For each $t \geq 0$ and $\omega \in \Omega$, $\{p_\gamma^{(i)}(t, \omega)\}_{\gamma \in \mathbb{Z}_+^d}$ denotes the offspring distribution of an individual of type i at time t under a realization ω. Define the splitting intensities of the branching process as $\Lambda_\gamma^{(i)}(t, \omega) = \lambda^{(i)}(t, \omega) p_\gamma^{(i)}(t, \omega)$, for $i \in S$ and $\gamma \in \mathbb{Z}_+^d$. Note that $\Lambda_\gamma^{(i)}$ for $i \in S$ and $\gamma \in \mathbb{Z}_+^d$ are themselves nonnegative stochastic processes defined on $(\Omega, \mathcal{F}, P)$. It is customary to only consider splitting intensities with $\Lambda_{e_i}^{(i)} = 0$ for each $i \in S$, i.e., we omit the possibility of a death of an individual followed by replacement by itself since this represents no change in the situation because of the lack of dependence on age.

Write $\tilde{\mathbb{Z}}_+^d = \bigcup_{i \in S} \{\mathbf{n} - e_i \mid \mathbf{n} \in \mathbb{Z}_+^d\}$, where e_i is a column vector with ith component equal to 1 and others 0. For each $\alpha \in \tilde{\mathbb{Z}}_+^d$, we write $\Lambda_\alpha = (\Lambda_{\alpha+e_1}^{(1)}, \cdots, \Lambda_{\alpha+e_d}^{(d)})^\top$, and $\Lambda_\alpha \cdot \mathbf{Z} = \sum_{j \in S} \Lambda_{\alpha+e_j}^{(j)} \mathbf{Z}^j$. Write $\Lambda = \{\Lambda_\alpha, \ \alpha \in \tilde{\mathbb{Z}}_+^d\}$. We want that the conditional transition probability of the process $\mathbf{Z}(t)$ in an instant time interval $(t, t + \Delta t)$ is

$$P\{\mathbf{Z}(t + \Delta t) - \mathbf{Z}(t) = \alpha \mid \mathcal{F}_t\} = E\left[\int_t^{t+\Delta t} \Lambda_\alpha(s) \cdot \mathbf{Z}(s) ds \ \Big| \ \mathcal{F}_t\right] + o(\Delta t). \quad (4.1.1)$$

That is, when conditioned on the environment, we want $\Lambda^{(i)}_{\alpha+e_i}(t)\Delta t$ to be the probability that a given individual of type i dies and is replaced by $\alpha + e_i$ offsprings in an instant time interval $(t, t + \Delta t)$. This implies the branching property which says that all individuals living at the same moment behave independently of one another when conditioned on the environment.

By employing the technique used in ([EK86], Chap. 6.4), the Markov branching process $\mathbf{Z}(t)$ is constructed as the solution of

$$\mathbf{Z}(t) = \mathbf{Z}(0) + \sum_{\alpha \in \tilde{\mathbb{Z}}^d_+} \alpha Y_\alpha \left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}(s) ds \right), \tag{4.1.2}$$

where $\{Y_\alpha, \ \alpha \in \tilde{\mathbb{Z}}^d_+\}$, are independent standard Poisson processes defined on $(\Omega, \mathcal{F}, P)$, independent of the Λ. It is shown in the following proposition that the solution of (4.1.2) exists and is unique and satisfies the transition property (4.1.1).

Proposition 4.1.1. *Assume*

$$\left\| \sum_{\alpha \in \tilde{\mathbb{Z}}^d_+} \alpha \int_0^t \Lambda^{(i)}_{\alpha+e_i}(s) ds \right\|_1 < \infty \ a.s. \ for \ all \ t \geq 0 \ and \ each \ i \in S. \tag{4.1.3}$$

Then

(i) The solution of (4.1.2) exists for all the time and is unique a.s.

(ii) The solution of (4.1.2) satisfies (4.1.1).

We will show the uniqueness of the solution of (4.1.2) by Gronwall's inequality. For convenience, we state Gronwall's inequality in the following lemma.

Lemma 4.1.1. *Let I denote an interval of the real line of the form $[a, \infty)$ or $[a, b]$ or $[a, b)$ with $a < b$. Let β and u be measurable functions defined on I and let μ be a locally finite measure on the Borel σ-algebra of I (we need $\mu([a, t]) < \infty$ for all t in I). Assume that u is integrable with respect to μ in the sense that*

$$\int_a^t |u(s)|\mu(ds) < \infty, \ t \in I,$$

and that u satisfies the integral inequality

$$u(t) \leq \beta(t) + \int_{[a,t)} u(s)\mu(ds), \ t \in I.$$

If, in addition,

63

- *the function β is non-negative*

- *or the function $t \to \mu([a,t])$ is continuous for t in I and the function β is integrable with respect to μ in the sense that*

$$\int_a^t |\beta(s)|\mu(ds) < \infty, \ t \in I,$$

then u satisfies Gronwall's inequality

$$u(t) \le \beta(t) + \int_{[a,t)} \beta(s)\exp(\mu(I_{s,t}))\mu(ds)$$

for all t in I, where $I_{s,t}$ denotes to open interval (s,t).

Remark: If β is the zero function and u is non-negative, then Gronwall's inequality implies that u is the zero function.

Proof of Proposition 4.1.1. (i) Suppose there exist two solutions $\mathbf{Z}$ and $\tilde{\mathbf{Z}}$ of the equation (4.1.2) with $\tilde{\mathbf{Z}}(0) = \mathbf{Z}(0)$. Then conditioned on Λ,

$$
\begin{aligned}
&E[\|\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\|_1 \mid \Lambda] \\
=\ & E\Big[\Big\| \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha\Big(Y_\alpha\Big(\int_0^t \Lambda_\alpha(s)\cdot\mathbf{Z}(s)ds\Big) - Y_\alpha\Big(\int_0^t \Lambda_\alpha(s)\cdot\tilde{\mathbf{Z}}(s)ds\Big)\Big)\Big\|_1 \,\Big|\, \Lambda\Big] \\
\le\ & \int_0^t \Big(\sum_{\alpha\in\tilde{\mathbb{Z}}_+^d} \|\alpha\|_1 \sum_i \Lambda^{(i)}_{\alpha+e_i}(s)\Big) E[\|\mathbf{Z}(s) - \tilde{\mathbf{Z}}(s)\|_1 \mid \Lambda]ds.
\end{aligned}
$$

Let $I = [0,\infty)$, $u(t) = E[\|\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\|_1 \mid \Lambda]$, $\mu(ds) = \Big(\sum_{\alpha\in\tilde{\mathbb{Z}}_+^d} \|\alpha\|_1 \sum_i \Lambda^{(i)}_{\alpha+e_i}(s)\Big)ds$ and $\beta \equiv 0$. By the assumption (4.1.3), μ is a locally finite measure on the Borel $\sigma-$algebra of I a.s. Since u is a step function a.s., u is measurable and integrable a.s. Then by Gronwall's inequality, $P\{E[\|\mathbf{Z}(t) - \tilde{\mathbf{Z}}(t)\|_1 \mid \Lambda] = 0, \text{ for } t \ge 0\} = 1$. Hence $P\{\tilde{\mathbf{Z}}(t) = \mathbf{Z}(t), \text{ for } t \ge 0\} = 1$, the uniqueness holds.

Next we show the existence of the solution. Construct the solution by iteration as follows:

$$
\begin{aligned}
\mathbf{Z}^{(0)}(t) &= \mathbf{Z}(0), \\
\mathbf{Z}^{(n)}(t) &= \mathbf{Z}(0) + \sum_{\alpha\in\tilde{\mathbb{Z}}_+^d} \alpha Y_\alpha\left(\int_0^t \Lambda_\alpha(s)\cdot\mathbf{Z}^{(n-1)}(s)ds\right).
\end{aligned}
$$

64

Define a sequence of jump times as follows:

$$\tau_0 \;=\; 0;$$

$$\tau_1 \;=\; \inf\left\{t > 0 \;\middle|\; \text{there exists an } \alpha, s.t., Y_\alpha\left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(0)}(s)ds\right) = 1,\right.$$

$$\left. Y_\beta\left(\int_0^t \Lambda_\beta(s) \cdot \mathbf{Z}^{(0)}(s)ds\right) = 0, \text{ for } \beta \neq \alpha\right\};$$

$$\tau_n \;=\; \inf\left\{t > \tau_{n-1} \;\middle|\; \text{there exists an } \alpha \text{ and a } j \in \{n-2, n-1\}, s.t.,\right.$$

$$Y_\alpha\left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(j)}(s)ds\right) - Y_\alpha\left(\int_0^{\tau_{n-1}} \Lambda_\alpha(s) \cdot \mathbf{Z}^{(j)}(s)ds\right) = 1,$$

$$Y_\beta\left(\int_0^t \Lambda_\beta(s) \cdot \mathbf{Z}^{(n-2)}(s)ds\right) - Y_\beta\left(\int_0^{\tau_{n-1}} \Lambda_\beta(s) \cdot \mathbf{Z}^{(n-2)}(s)ds\right) = 0,$$

$$\left. Y_\beta\left(\int_0^t \Lambda_\beta(s) \cdot \mathbf{Z}^{(n-1)}(s)ds\right) - Y_\beta\left(\int_0^{\tau_{n-1}} \Lambda_\beta(s) \cdot \mathbf{Z}^{(n-1)}(s)ds\right) = 0, \text{ for } \beta \neq \alpha\right\},$$

for $n > 1$. Then by induction we have

$$\mathbf{Z}^{(n)}(t) = \mathbf{Z}^{(n-1)}(t), \text{ for } t < \tau_n.$$

Indeed,

$$\begin{aligned}
\mathbf{Z}^{(1)}(t) &= \mathbf{Z}(0) + \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha Y_\alpha\left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(0)}(s)ds\right) \\
&= \mathbf{Z}^{(0)}(t), \text{ for } t < \tau_1.
\end{aligned}$$

Suppose that

$$\mathbf{Z}^{(n-1)}(t) = \mathbf{Z}^{(n-2)}(t), \text{ for } t < \tau_{n-1}.$$

Then

$$\begin{aligned}
\mathbf{Z}^{(n)}(t) &= \mathbf{Z}(0) + \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha Y_\alpha\left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(n-1)}(s)ds\right) \\
&= \mathbf{Z}(0) + \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha Y_\alpha\left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(n-2)}(s)ds\right) \\
&= \mathbf{Z}^{(n-1)}(t), \text{ for } t < \tau_{n-1};
\end{aligned}$$

$$
\begin{aligned}
\mathbf{Z}^{(n)}(t) &= \mathbf{Z}(0) + \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha Y_\alpha \left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(n-1)}(s)ds \right) \\
&= \mathbf{Z}(0) + \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha Y_\alpha \left(\int_0^{\tau_{n-1}} \Lambda_\alpha(s) \cdot \mathbf{Z}^{(n-2)}(s)ds \right) + \\
&\quad \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha \left[Y_\alpha \left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}^{(n-1)}(s)ds \right) - Y_\alpha \left(\int_0^{\tau_{n-1}} \Lambda_\alpha(s) \cdot \mathbf{Z}^{(n-1)}(s)ds \right) \right] \\
&= \mathbf{Z}^{(n-1)}(\tau_{n-1}) \\
&= \mathbf{Z}^{(n-1)}(t), \text{ for } \tau_{n-1} \le t < \tau_n.
\end{aligned}
$$

Therefore,

$$
\mathbf{Z}(t) = \lim_{n \to \infty} \mathbf{Z}^{(n)}(t), \ t < \infty,
$$

exists and $\mathbf{Z}(t)$ satisfies (4.1.2) since

$$
\mathbf{Z}(t) = \mathbf{Z}^{(n)}(t), \text{ for } t < \tau_n.
$$

The assumption (4.1.3) makes sure that $\lim_{n \to \infty} \tau_n = \infty$. Indeed,

$$
E[\|\mathbf{Z}(t)\|_1 \mid \Lambda] \le \|\mathbf{Z}(0)\|_1 + \int_0^t \left(\sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \|\alpha\|_1 \sum_i \Lambda_{\alpha+e_i}^i(s) \right) E[\|\mathbf{Z}(t)\|_1 \mid \Lambda]ds.
$$

By Gronwall's inequality, it can be deduced that $E[\|\mathbf{Z}(t)\|_1 \mid \Lambda] < \infty$ for $0 \le t < \infty$ a.s. Hence the existence of the solution of the equation (4.1.2) holds for all the time.

(ii) Let A denote the event

$$
\Big\{ Y_\alpha \left(\int_0^{t+\Delta t} \Lambda_\alpha(s) \cdot \mathbf{Z}(s)ds \right) - Y_\alpha \left(\int_0^t \Lambda_\alpha(s) \cdot \mathbf{Z}(s)ds \right) = 1,
$$
$$
Y_\beta \left(\int_0^{t+\Delta t} \Lambda_\beta(s) \cdot \mathbf{Z}(s)ds \right) - Y_\beta \left(\int_0^t \Lambda_\beta(s) \cdot \mathbf{Z}(s)ds \right) = 0, \text{ for } \beta \ne \alpha \Big\}.
$$

Then

$$
\begin{aligned}
&P\{\mathbf{Z}(t + \Delta t) - \mathbf{Z}(t) = \alpha \mid \mathcal{F}_t\} \\
&= P\{A \mid \mathcal{F}_t\} \\
&= E[E[\mathbf{1}_A \mid \mathcal{F}_{(t+\Delta t)-}] \mid \mathcal{F}_t] \\
&= E\left[\int_t^{t+\Delta t} \Lambda_\alpha(s) \cdot \mathbf{Z}(s)ds \ \Big| \ \mathcal{F}_t \right] + o(\Delta t).
\end{aligned}
$$

$\square$

4.2 Kolmogorov equations and the mean matrix

When conditioned on the environment, a multitype Markov branching process in a random environment is a time non-homogeneous vector-valued Markov chain in continuous time. The transition function of the non-homogeneous multitype Markov branching process is

$$P_{\alpha\beta}(u,t) = P\{\mathbf{Z}(t) = \beta \mid \mathbf{Z}(u) = \alpha, \Lambda\}, \ \alpha, \beta \in \mathbb{Z}_+^d, \ 0 \le u < t. \tag{4.2.1}$$

The corresponding generating function is

$$F(\alpha, \mathbf{s}; u, t) = \sum_{\beta \in \mathbb{Z}_+^d} P_{\alpha\beta}(u,t)\mathbf{s}^\beta, \tag{4.2.2}$$

where $\mathbf{s}^\beta$ denotes $\prod_{i=1}^d s_i^{\beta_i}$. Write

$$F(j, \mathbf{s}; u, t) = F(e_j, \mathbf{s}; u, t)$$

and

$$\mathbf{F}(\mathbf{s}; u, t) = (F(1, \mathbf{s}; u, t), \cdots, F(d, \mathbf{s}; u, t)).$$

For the offspring production probability and generating functions we write

$$\mathbf{p}_\gamma(t) = (p_\gamma^{(1)}(t), \cdots, p_\gamma^{(d)}(t)),$$

and

$$\mathbf{f}(\mathbf{s}; t) = (f^{(1)}(\mathbf{s}; t), \cdots, f^{(d)}(\mathbf{s}; t)),$$

where $f^{(j)}(\mathbf{s}; t) = \sum_{\gamma \in \mathbb{Z}_+^d} p_\gamma^{(j)}(t)\mathbf{s}^\gamma$. Let

$$u^{(j)}(\mathbf{s}; t) = \lambda^{(j)}(t)[f^{(j)}(\mathbf{s}; t) - s_j].$$

The backward equations of Kolmogorov for the generating functions of the non-homogeneous Markov branching process was formulated in [Har63][KM80] without a proof. Here we state the result and give a proof.

Proposition 4.2.1. *Assume that* (4.1.3) *holds. Then the backward equations of Kolmogorov for the generating functions of the Markov branching process conditioned on environment are*

$$\frac{\partial F(j, \mathbf{s}; u, t)}{\partial u} = -u^{(j)}(\mathbf{F}(\mathbf{s}; u, t); u), \tag{4.2.3}$$

with terminal conditions

$$F(j, \mathbf{s}; t, t) = s_j, \tag{4.2.4}$$

for $j = 1, \cdots, d$.

Proof. By Proposition 4.1.1, the assumption (4.1.3) makes sure that there exists a Markov branching process in a random environment and that the branching property holds conditioned on the environment, i.e. $F(\alpha, \mathbf{s}; u, t) = \mathbf{F}(\mathbf{s}; u, t)^{\alpha}$.

The transition function of the Markov branching process conditioned on environment satisfies the backward equation

$$\frac{\partial P_{\alpha\beta}(u, t)}{\partial u} = -\Big(\sum_{i=1}^{d} \sum_{\gamma-\alpha+e_i \geq 0, \gamma \in \mathbb{Z}_+^d} \lambda^{(i)}(u)\alpha_i p^{(i)}_{\gamma-\alpha+e_i}(u) P_{\gamma\beta}(u, t) - \sum_{i=1}^{d} \lambda^{(i)}(u)\alpha_i P_{\alpha\beta}(u, t) \Big),$$

with the terminal condition

$$P_{\alpha\beta}(t, t) = \delta_{\alpha\beta}.$$

Then for the corresponding generating function we have

$$\frac{\partial F(\alpha, \mathbf{s}; u, t)}{\partial u} = \sum_{\beta \in \mathbb{Z}_+^d} \frac{\partial P_{\alpha\beta}(u, t)}{\partial u} \mathbf{s}^{\beta}$$

$$= -\sum_{i=1}^{d} \lambda^{(i)}(u) \Big(\sum_{\beta \in \mathbb{Z}_+^d} \Big(\sum_{\gamma-\alpha+e_i \geq 0, \gamma \in \mathbb{Z}_+^d} \alpha_i p^{(i)}_{\gamma-\alpha+e_i}(u) P_{\gamma\beta}(u, t) - \alpha_i P_{\alpha\beta}(u, t) \Big) \mathbf{s}^{\beta} \Big).$$

68

Specially, when $\alpha = e_j$ for $j = 1, \cdots, d$, it follows that

$$
\begin{aligned}
\frac{\partial F(j, \mathbf{s}; u, t)}{\partial u} &= -\lambda^{(j)}(u)\Big(\sum_{\beta \in \mathbb{Z}_+^d}\Big(\sum_{\gamma \in \mathbb{Z}_+^d} p_\gamma^{(j)}(u)P_{\gamma\beta}(u,t) - P_{e_j\beta}(u,t)\Big)\mathbf{s}^\beta\Big) \\
&= -\lambda^{(j)}(u)\Big(\sum_{\gamma \in \mathbb{Z}_+^d} p_\gamma^{(j)}(u)\Big(\sum_{\beta \in \mathbb{Z}_+^d} P_{\gamma\beta}(u,t)\mathbf{s}^\beta\Big) - \Big(\sum_{\beta \in \mathbb{Z}_+^d} P_{e_j\beta}(u,t)\mathbf{s}^\beta\Big)\Big) \\
&= -\lambda^{(j)}(u)\Big(\sum_{\gamma \in \mathbb{Z}_+^d} p_\gamma^{(j)}(u)F(\gamma, \mathbf{s}; u, t) - F(j, \mathbf{s}; u, t)\Big) \\
&= -\lambda^{(j)}(u)\Big(\sum_{\gamma \in \mathbb{Z}_+^d} p_\gamma^{(j)}(u)\mathbf{F}(\mathbf{s}; u, t)^\gamma - F(j, \mathbf{s}; u, t)\Big) \\
&= -\lambda^{(j)}(u)\Big(f^{(j)}(\mathbf{F}(\mathbf{s}; u, t); u) - F(j, \mathbf{s}; u, t)\Big) \\
&= -u^{(j)}(\mathbf{F}(\mathbf{s}; u, t); u).
\end{aligned}
$$

The terminal condition of the generating function follows easily from that of the transition function. $\qquad\square$

The mean matrix of the non-homogeneous multitype Markov branching process is the $d \times d$ matrix $M(u, t) = (m_{ij}(u, t))$, where $m_{ij}(u, t) = E[Z_j(t) \mid Z(u) = e_i, \Lambda]$. Let $A^i(t) = \sum_{\alpha \in \tilde{\mathbb{Z}}_+^d} \alpha \Lambda_{\alpha + e_i}^{(i)}(t)$ and $A(t) = (A^1(t), \cdots, A^d(t))$.

Proposition 4.2.2. *Assume that (4.1.3) holds. Assume that*

$$
\Big(\int_u^t A^\top(s)ds\Big)A^\top(u) = A^\top(u)\Big(\int_u^t A^\top(s)ds\Big), \tag{4.2.5}
$$

i.e. for any $i, j \in S$ and $0 \le u < t$,

$$
\int_u^t \sum_{k \in S, \alpha \in \tilde{\mathbb{Z}}_+^d, \beta \in \tilde{\mathbb{Z}}_+^d} \alpha_i \beta_k \Lambda_{\alpha + e_k}^{(k)}(s)\Lambda_{\beta + e_j}^{(j)}(u)ds = \int_u^t \sum_{k \in S, \alpha \in \tilde{\mathbb{Z}}_+^d, \beta \in \tilde{\mathbb{Z}}_+^d} \alpha_i \beta_k \Lambda_{\alpha + e_k}^{(k)}(u)\Lambda_{\beta + e_j}^{(j)}(s)ds.
$$

Then

$$
M(u, t) = e^{\int_u^t A^\top(s)ds}.
$$

Proof. Note that for any $j, k \in S$, $m_{jk}(u, t) = \frac{\partial F(j, \mathbf{s}; u, t)}{\partial s_k}\big|_{\mathbf{s}=\mathbf{1}}$ and that the assumption (4.1.3) makes sure $\mathbf{F}(\mathbf{1}; u, t) = \mathbf{1}$. Differentiating (4.2.3) w.r.t. s_k and taking $\mathbf{s} = \mathbf{1}$,

69

we obtain

$$
\begin{aligned}
\frac{\partial m_{jk}(u,t)}{\partial u}
&= -\lambda^{(j)}(u)\Big(\sum_{l=1}^{d} \frac{\partial f^{(j)}(\mathbf{x};u)}{\partial x_l}\Big|_{\mathbf{x}=\mathbf{F}(\mathbf{1};u,t)} \cdot \frac{\partial F(l,\mathbf{s};u,t)}{\partial s_k}\Big|_{\mathbf{s}=\mathbf{1}} - \frac{\partial F(j,\mathbf{s};u,t)}{\partial s_k}\Big|_{\mathbf{s}=\mathbf{1}}\Big) \\
&= -\lambda^{(j)}(u)\Big(\sum_{l=1}^{d} \frac{\partial f^{(j)}(\mathbf{x};u)}{\partial x_l}\Big|_{\mathbf{x}=\mathbf{1}} \cdot m_{lk}(u,t) - m_{jk}(u,t)\Big) \\
&= -\lambda^{(j)}(u)\Big(\sum_{l=1}^{d} \sum_{\gamma \in \mathbb{Z}_+^d} \gamma_l p_\gamma^{(j)}(u) \cdot m_{lk}(u,t) - m_{jk}(u,t)\Big) \\
&= -\sum_{l=1}^{d} \lambda^{(j)}(u)\Big(\sum_{\gamma \in \mathbb{Z}_+^d} \gamma_l p_\gamma^{(j)}(u) - \delta_{jl}\Big) m_{lk}(u,t) \\
&= -\sum_{l=1}^{d} a_{lj}(u) m_{lk}(u,t),
\end{aligned}
$$

where $a_{lj}(u)$ is the ljth element of the matrix $A(u)$. The above equation can be expressed in the following matrix form:

$$
\frac{\partial M(u,t)}{\partial u} = -A^\top(u) M(u,t), \text{ for } 0 \le u < t, \tag{4.2.6}
$$

with the terminal condition

$$
M(t,t) = I, \tag{4.2.7}
$$

where I is a $d \times d$ identity matrix.

With the assumption (4.2.5) we imply that the solution of the equations (4.2.6), (4.2.7) is

$$
M(u,t) = e^{\int_u^t A^\top(s)ds}.
$$

$\square$

Similarly, we could deduce the forward equations of Kolmogorov [Har63][KM80] and the mean matrix. Here we only state the result since the idea of proof is the same as before.

Proposition 4.2.3. *Assume that (4.1.3) holds. Then the forward equations of Kolmogorov for the generating functions of the Markov branching process conditioned on the environment are*

$$\frac{\partial F(j, \mathbf{s}; u, t)}{\partial t} = \sum_{i \in S} u^{(i)}(\mathbf{s}; t) \frac{\partial F(j, \mathbf{s}; u, t)}{\partial s_i}, \tag{4.2.8}$$

with initial conditions

$$F(j, \mathbf{s}; u, u) = s_j, \tag{4.2.9}$$

for $j \in S$. Furthermore, assume that

$$\left(\int_u^t A^\top(s) ds \right) A^\top(t) = A^\top(t) \left(\int_u^t A^\top(s) ds \right), \tag{4.2.10}$$

i.e. for any $i, j \in S$ and $0 \le u < t$,

$$\int_u^t \sum_{k \in S, \alpha \in \tilde{\mathbb{Z}}_+^d, \beta \in \tilde{\mathbb{Z}}_+^d} \alpha_i \beta_k \Lambda_{\alpha+e_k}^{(k)}(s) \Lambda_{\beta+e_j}^{(j)}(t) ds = \int_u^t \sum_{k \in S, \alpha \in \tilde{\mathbb{Z}}_+^d, \beta \in \tilde{\mathbb{Z}}_+^d} \alpha_i \beta_k \Lambda_{\alpha+e_k}^{(k)}(t) \Lambda_{\beta+e_j}^{(j)}(s) ds.$$

Then

$$M(u, t) = e^{\int_u^t A^\top(s) ds}.$$

4.3 The parallel mutation-selection model in a random environment
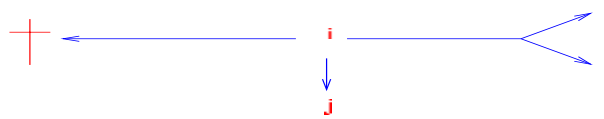

Fig. 1: The parallel mutation-reproduction model

We generalize a parallel mutation-selection model [BG07] with a random environment, which is a special multitype Markov branching process in a random environment, denoted by $\{\mathbf{Z}(t)\}_{t \ge 0}$. As depicted in Fig. 1, an i-particle may, at each instant in continuous time, do either of three things: It may produce a copy of itself (at rate $\Lambda_{2e_i}^{(i)}$), it may die (at rate $\Lambda_{\mathbf{0}}^{(i)}$), or it may mutate to type j ($j \ne i$) (at rate $\Lambda_{e_j}^{(i)}$). We want to show the instability property of $\{\mathbf{Z}(t)\}_{t \ge 0}$ as in [CT81].

For simplicity, we consider only two types, i.e. $S = \{1, 2\}$. However, a generalization to the multitype case is not difficult. Write $\mathcal{I} = \{e_1, -e_1, e_2 - e_1, e_2, -e_2, e_1 - e_2\}$. Assume $\Lambda^{(i)}_{\alpha + e_i}$ are controlled by the same environment stochastic process η for $i \in S$ and $\alpha \in \mathcal{I}$, i.e. $\Lambda^{(i)}_{\alpha + e_i}(\eta(t, \omega))$. Assume $\| \int_0^t \sum_{i \in S} \sum_{\alpha \in \mathcal{I}} \alpha \Lambda^{(i)}_{\alpha + e_i}(s) ds \|_1 < \infty$ a.s. for all $t \geq 0$.

Let $\{\eta(t)\}_{t \geq 0}$ be an irreducible, recurrent Markov chain in continuous time on a countable state space $\mathcal{Y}$, jump times $\tau_n \uparrow +\infty$, $n \geq 0$, (set $\tau_0 = 0$) and infinitesimal parameters $\{q_{xy}\}$, $x, y \in \mathcal{Y}$. Let $q_y = 1/E_y \tau_1$. We formulate the evolution of the process $\{\mathbf{Z}(t)\}_{t \geq 0}$ as follows: let $\eta(0) = y_0$. In the time span $w_n = \tau_n - \tau_{n-1}$, $n \geq 1$, $\{\mathbf{Z}(t)\}_{t \geq 0}$ evolves as a two-type parallel mutation-selection Markov branching process, and associated jump times $T_j(n)$, $1 \leq j \leq k_n$. Set $T_0(n) = \tau_{n-1}$, $n \geq 1$. So, the environmental process $\{\eta(t)\}_{t \geq 0}$ jumps at times $T_0(n)$, and the process $\{\mathbf{Z}(t)\}_{t \geq 0}$ jumps at times $T_j(n)$, $1 \leq j \leq k_n$. We thus have a sequence $\{T_j(n), \ 0 \leq j \leq k_n, \ n \geq 1\}$ of exponentially distributed random variables such that

$$0 = T_0(1) < T_1(1) < \cdots < T_{k_1}(1) < T_0(2) < T_1(2) < \cdots < T_{k_2}(2) < \cdots .$$

(Appropriate modifications in the above sequencing would be made in the event that $k_n = 0$.) For $j \geq 1$, let T_j denote the jth element of this increasing sequence so that for a fixed j, and for a given realization of $\{(\eta(t), \mathbf{Z}(t))\}_{t \geq 0}$, there exist unique n and i such that

$$T_j = T_i(n), \ \text{with } 0 \leq i \leq k_n. \tag{4.3.1}$$

Then $\{\mathbf{Z}(t)\}_{t \geq 0}$ is a two-type parallel mutation-selection Markov branching process in a random environment $\{\eta(t)\}_{t \geq 0}$.

Now we consider the embedded chain $\{(\eta_n, \mathbf{Z}_n)\}_{n \in \mathbb{N}_0}$ of the bivariate Markov process $\{(\eta(t), \mathbf{Z}(t))\}_{t \geq 0}$, defined by

$$\{(\eta_n, \mathbf{Z}_n)\}_{n \in \mathbb{N}_0} = \{(\eta(\tau_n), \mathbf{Z}(\tau_n))\}_{n \in \mathbb{N}_0}. \tag{4.3.2}$$

It is important to note that $\{(\eta_n, \mathbf{Z}_n)\}_{n \in \mathbb{N}_0}$ satisfy the following relation:

$$P_{(y, \mathbf{z})}\{\mathbf{Z}_{n+1} \in B \mid \eta_0, \eta_1, \cdots ; \mathbf{Z}_0, \mathbf{Z}_1, \cdots, \mathbf{Z}_n\} \tag{4.3.3}$$
$$= P_{(y, \mathbf{z})}\{\mathbf{Z}_{n+1} \in B \mid \eta_n, \mathbf{Z}_n\} \text{ for every } (y, \mathbf{z}) \in \mathcal{Y} \times \mathbb{Z}_+^2 \text{ and } B \subseteq \mathbb{Z}_+^2.$$

Then (4.3.3) together with the fact that $\{\eta_n\}_{n\in\mathbb{N}_0}$ is a Markov chain implies that $\{(\eta_n, \mathbf{Z}_n)\}_{n\in\mathbb{N}_0}$ is Markov.

Cogburn studied processes $\{(\eta_n, \mathbf{Z}_n)\}_{n\in\mathbb{N}_0}$ satisfying the relation (4.3.3) and called the marginal process $\{\mathbf{Z}_n\}_{n\in\mathbb{N}_0}$ a Markov chain in a random environment $\{\eta_n\}_{n\in\mathbb{N}_0}$ [Cog80]. An important and useful concept in the study of these processes is the notion of a proper Markov chain in a random environment. $\{\mathbf{Z}_n\}_{n\in\mathbb{N}_0}$ is called proper if whenever

$$\sup_{y\in\mathcal{Y}} P_{(y,\mathbf{z})}\{\mathbf{Z}_n = \mathbf{z} \text{ i.o.}\} > 0, \text{ for } \mathbf{z} \in \mathbb{Z}_+^2,$$

where $\{\mathbf{Z}_n = \mathbf{z} \text{ i.o.}\}$ means that the event $\{\mathbf{Z}_n = \mathbf{z}\}$ happens infinitely often, then $(y, \mathbf{z})$ is recurrent for some $y \in \mathcal{Y}$ for the Markov chain $\{(\eta_n, \mathbf{Z}_n)\}_{n\in\mathbb{N}_0}$.

Note. If a Markov chain in a random environment, $\{\mathbf{Z}_n\}_{n\in\mathbb{N}_0}$, is proper, and states of $\mathcal{Y} \times (\mathbb{Z}_+^2\backslash\{\mathbf{0}\})$ communicate (two states communicate means that the former is accessible from the latter and the latter is also accessible from the former) and lead to $\mathcal{Y}\times\{\mathbf{0}\}$, which is closed, then necessarily $(y, \mathbf{z})$ is transient and $P_{(y,\mathbf{z})}\{\mathbf{Z}_n = \mathbf{z}' \text{ i.o.}\} = 0$ for all $(y, \mathbf{z}) \in \mathcal{Y} \times \mathbb{Z}_+^2$ and $\mathbf{z}' \neq \mathbf{0}$ and hence

$$P_{(y,\mathbf{z})}\{\lim_{n\to\infty} \|\mathbf{Z}_n\|_1 = 0 \text{ or } \infty\} = 1, \text{ for every } (y, \mathbf{z}) \in \mathcal{Y} \times \mathbb{Z}_+^2.$$

In this section, the main result (Theorem 4.3.2) follows from an application of a result (Theorem 4.3.1) due to Cogburn [Cog80]. Theorem 4.3.1 will allow us to assert that the Markov chain in a random environment, $\{\mathbf{Z}(\tau_n)\}_{n\in\mathbb{N}_0}$, is proper. For the definition of uniform $\varphi-$recurrence, see e.g. [Tor78].

Theorem 4.3.1 (Cogburn). *Let $\{\mathbf{Z}_n\}_{n\in\mathbb{N}_0}$ be a Markov chain in a random environment $\{\eta_n\}_{n\in\mathbb{N}_0}$. Let $P^{(y)}\{\mathbf{z}, \mathbf{z}'\}$ denote the transition probability of $\{\mathbf{Z}_n\}_{n\in\mathbb{N}_0}$ in the yth environment. Suppose that*

(a) *for each $\mathbf{z} \in \mathbb{Z}_+^2$ there exists a finite set $B_{\mathbf{z}} \subseteq \mathbb{Z}_+^2$ such that*
$\inf_{y\in\mathcal{Y}} P^{(y)}\{\mathbf{z}, B_{\mathbf{z}}\} > 0;$

(b) *the Markov chain $\{\eta_n\}_{n\in\mathbb{N}_0}$ is uniformly $\varphi-$recurrent.*

Then $\{\mathbf{Z}_n\}_{n\in\mathbb{N}_0}$ is a proper Markov chain in a random environment.

Theorem 4.3.2. *Let* $\{\mathbf{Z}(t)\}_{t \geq 0}$ *be a two-type parallel mutation-selection Markov branching process in a random environment* $\{\eta(t)\}_{t \geq 0}$. *Suppose that*

(a) *for each* $\mathbf{z} \in \mathbb{Z}_+^2$ *there exists a positive integer* $n_\mathbf{z} > \|\mathbf{z}\|_1$ *such that* $\inf_{y \in \mathcal{Y}} m_\mathbf{z}^{(y)}/q_y > 0$, *where* $m_\mathbf{z}^{(y)} = \sum_{k=\|\mathbf{z}\|_1}^{n_\mathbf{z}} \sigma_k^{(y)}$ *with* $\sigma_k^{(y)} = \sum_{\{(z_1,z_2)^\top \in \mathbb{Z}_+^2 : z_1 + z_2 = k\}} (z_1 \Lambda_{(2,0)^\top}^{(1)}(y) + z_2 \Lambda_{(0,2)^\top}^{(2)}(y))$;

(b) *the embedded chain* $\{\eta(\tau_n)\}_{n \in \mathbb{N}_0}$ *is uniformly* φ−*recurrent;*

(c) *states of* $\mathcal{Y} \times (\mathbb{Z}_+^2 \backslash \{\mathbf{0}\})$ *communicate and lead to* $\mathcal{Y} \times \{\mathbf{0}\}$.

Then $P_{(y,\mathbf{z})}\{\lim_{t \to \infty} \|\mathbf{Z}(t)\|_1 = 0 \text{ or } \infty\} = 1$, *for every* $(y, \mathbf{z}) \in \mathcal{Y} \times \mathbb{Z}_+^2$.

Proof. The proof follows from Theorem 2.2 in [CT81] with a generalization to two-dimensional state space. Consider the set $B_\mathbf{z} = \{\mathbf{z} \in \mathbb{Z}_+^2 : \|\mathbf{z}\|_1 \leq n_\mathbf{z}\}$. We will show

$$\inf_{y \in \mathcal{Y}} P^{(y)}\{\mathbf{z}, B_\mathbf{z}\} > 0. \tag{4.3.4}$$

Verification of (4.3.4) is tantamount to showing that if T is the first-passage time to a state whose norm is $n_\mathbf{z} + 1$, then $\inf_{y \in \mathcal{Y}} P_{(y,\mathbf{z})}\{T > \tau_y\} > 0$ where τ_y is the first jump time of $\{\eta(t)\}_{t \geq 0}$ starting at y. Let $N = n_\mathbf{z} + 1 - \|\mathbf{z}\|_1$.

Now

$$
\begin{aligned}
P_{(y,\mathbf{z})}\{T > \tau_y\} &= \int_0^\infty P_{(y,\mathbf{z})}\{T > t\} q_y e^{-q_y t} dt \tag{4.3.5} \\
&\geq q_y \int_0^\infty \exp\{-(q_y + (m_\mathbf{z}^{(y)}/N))t\} dt \\
&= N q_y / (N q_y + m_\mathbf{z}^{(y)}).
\end{aligned}
$$

The inequality in (4.3.5) follows by noting that T is stochastically larger than the sum of the independent first-passage times of $\mathbf{Z}(t)$ to level $k+1$ starting at a state in level k, $\|\mathbf{z}\|_1 \leq k \leq n_\mathbf{z}$. By assumption (a), $\inf_{y \in \mathcal{Y}} m_\mathbf{z}^{(y)}/q_y > 0$ so that $\inf_{y \in \mathcal{Y}} P_{(y,\mathbf{z})}\{T > \tau_y\} > 0$ and (4.3.4) follows.

An application of Theorem 4.3.1 will enable us to conclude that the process $\{\mathbf{Z}(\tau_n)\}_{n \in \mathbb{N}_0}$ is proper. By assumption (c) and the fact that $\mathbf{0}$ is an absorbing state in each environment, we may conclude from the note that the following holds:

$$P_{(y,\mathbf{z})}\{\lim_{n \to \infty} \|\mathbf{Z}(\tau_n)\|_1 = 0 \text{ or } \infty\} = 1, \text{ for every } (y, \mathbf{z}) \in \mathcal{Y} \times \mathbb{Z}_+^2. \tag{4.3.6}$$

74

We claim that the conclusion above holds for $\{\mathbf{Z}(t)\}_{t \geq 0}$. By assumption (a), we get that

$$\inf_{y \in \mathcal{Y}} P_{(y,\mathbf{z})}\{\|\mathbf{Z}(\tau_1)\|_1 \leq n_{\mathbf{z}}\} = \inf_{y \in \mathcal{Y}} P_{(y,\mathbf{z})}\{T > \tau_y\} > 0.$$

Then there exists $\varepsilon > 0$ such that $\inf_{y \in \mathcal{Y}} P_{(y,\mathbf{z})}\{\|\mathbf{Z}(\tau_1)\|_1 \leq n_{\mathbf{z}}\} \geq \varepsilon$. By the strong Markov property, given $\mathbf{z}' \in \mathbb{Z}_+^2 \backslash \{\mathbf{0}\}$,

$$P_{(y,\mathbf{z})}\{\|\mathbf{Z}(\tau_n)\|_1 \leq n_{\mathbf{z}'} \mid \eta(T_j(n)) = y', \mathbf{Z}(T_j(n)) = \mathbf{z}'\} \geq \varepsilon, \text{ for } y' \in \mathcal{Y}.$$

Hence by the law of total probability and property of conditional probability,

$$P_{(y,\mathbf{z})}\{\|\mathbf{Z}(\tau_n)\|_1 \leq n_{\mathbf{z}'} \mid \mathbf{Z}(T_j(n)) = \mathbf{z}'\} \geq \varepsilon.$$

Recall the relation of the $T_j(n)'$s and T_n's. Then it follows that

$$P_{(y,\mathbf{z})}\{\|\mathbf{Z}(\tau_n)\|_1 \leq n_{\mathbf{z}'} \text{ i.o.}\} \geq \varepsilon P_{(y,\mathbf{z})}\{\mathbf{Z}(T_n) = \mathbf{z}' \text{ i.o.}\} \qquad (4.3.7)$$

Since the asymptotic behavior of $\{\mathbf{Z}(t)\}_{t \geq 0}$ is the same as $\{\mathbf{Z}(T_n)\}_{n \in \mathbb{N}}$ and noting that

$$P_{(y,\mathbf{z})}\{\lim_{n \to \infty} \|\mathbf{Z}(\tau_n)\|_1 = 0\} = P_{(y,\mathbf{z})}\{\lim_{n \to \infty} \|\mathbf{Z}(T_n)\|_1 = 0\}, \text{ for every } (y, \mathbf{z}) \in \mathcal{Y} \times \mathbb{Z}_+^2,$$

(4.3.6) and (4.3.7) give us the conclusion of the theorem. $\qquad \square$

We give an example. Let $\eta(t)$ be a standard Poisson process. For a fixed n, let $\Lambda_{(0,0)^\top}^{(1)}(t) \equiv 1$, $\Lambda_{(2,0)^\top}^{(1)}(t) = 1 + n^{-1/2}(-1)^{\eta(t)}$, $\Lambda_{(0,1)^\top}^{(1)}(t) = n^{-1}$, $\Lambda_{(0,0)^\top}^{(2)}(t) \equiv 1$, $\Lambda_{(0,2)^\top}^{(2)}(t) = 1 + n^{-1/2}(-1)^{\eta(t)}$, $\Lambda_{(1,0)^\top}^{(2)}(t) = (2n)^{-1}$, and else 0. Note that the example satisfies the assumptions of Theorem 4.3.2, therefore the instability holds.

4.4 Continuous approximation for the parallel processes

4.4.1 Main result

Consider a sequence of two-type parallel mutation-selection Markov branching processes $\{\mathbf{Z}^{(n)}(t)\}_{t \geq 0}$ in random environments $\{\eta^{(n)}(t)\}_{t \geq 0}$. Throughout this section, assume that the corresponding intensity processes $\Lambda^{(n)} \equiv \{\Lambda_{\alpha+e_i}^{(n)i}(\eta^{(n)}), i \in S \text{ and } \alpha \in$

$\mathcal{I}\}$ satisfy the following conditions denoted by (C1):

(C1.1) For each n, $\| \int_0^t \sum_{i\in S} \sum_{\alpha\in\mathcal{I}} \alpha \Lambda_{\alpha+e_i}^{(n)i}(s)ds \|_1 < \infty$, a.s. for all $t \geq 0$.

(C1.2) For each n and for any $i, j \in S$, $\int_0^t \sum_{k\in S, \alpha\in\mathcal{I}, \beta\in\mathcal{I}} \alpha_i \beta_k \Lambda_{\alpha+e_k}^{(n)k}(ns)\Lambda_{\beta+e_j}^{(n)j}(nt)ds = \int_0^t \sum_{k\in S, \alpha\in\mathcal{I}, \beta\in\mathcal{I}} \alpha_i \beta_k \Lambda_{\alpha+e_k}^{(n)k}(nt)\Lambda_{\beta+e_j}^{(n)j}(ns)ds$, a.s. for all $t \geq 0$.

(C1.3) $\Lambda_{(0,1)^\top}^{(n)1}(t) = O\left(\frac{1}{n}\right)$, $\Lambda_{(1,0)^\top}^{(n)2}(t) = O\left(\frac{1}{n}\right)$, $\lim_{n\to\infty}\left(\Lambda_{(2,0)^\top}^{(n)1} - \Lambda_{(0,0)^\top}^{(n)1}\right)(t) = 0$, and $\lim_{n\to\infty}\left(\Lambda_{(0,2)^\top}^{(n)2} - \Lambda_{(0,0)^\top}^{(n)2}\right)(t) = 0$, a.s. for all $t \geq 0$.

(C1.4) For each n, $\{\eta^{(n)}(t)\}_{t\geq 0}$ is an irreducible, recurrent Markov chain in continuous time on a countable state space $\mathcal{Y}^{(n)}$, jump times $\tau_m^{(n)} \uparrow +\infty$, $m \geq 0$, (set $\tau_0^{(n)} = 0$) and infinitesimal parameters $\left\{q_{xy}^{(n)}, \ x, y \in \mathcal{Y}^{(n)}\right\}$. Let $q_y^{(n)} = 1/E_y\tau_1^{(n)}$.

(C1.5) For each n sufficiently large and for each $\mathbf{z} \in \mathbb{Z}_+^2$ there exists a positive integer $l_{\mathbf{z}}^{(n)} > \|\mathbf{z}\|_1$ such that $\inf_{y\in\mathcal{Y}^{(n)}} m_{\mathbf{z}}^{(y)(n)}/q_y^{(n)} > 0$, where $m_{\mathbf{z}}^{(y)(n)} = \sum_{k=\|\mathbf{z}\|_1}^{l_{\mathbf{z}}^{(n)}} \sigma_k^{(y)(n)}$ with $\sigma_k^{(y)(n)} = \sum_{\{(z_1,z_2)^\top\in\mathbb{Z}_+^2 : z_1+z_2=k\}} \left(z_1\Lambda_{(2,0)^\top}^{(n)1}(y) + z_2\Lambda_{(0,2)^\top}^{(n)2}(y)\right)$.

(C1.6) For each n sufficiently large, the embedded chain $\left\{\eta^{(n)}\left(\tau_m^{(n)}\right)\right\}_{m\in\mathbb{N}_0}$ is uniformly $\varphi-$recurrent.

(C1.7) For each n sufficiently large, states of $\mathcal{Y}^{(n)} \times (\mathbb{Z}_+^2\backslash\{\mathbf{0}\})$ communicate and lead to $\mathcal{Y}^{(n)} \times \{\mathbf{0}\}$.

Remark: 1. The condition (C1.1) makes sure that for each n, there exists a two-type parallel mutation-selection Markov branching process in a random environment, denoted by $\{(\mathbf{Z}^{(n)}(t), \eta^{(n)}(t))\}_{t\geq 0}$.

2. The condition (C1.2) will be used in deducing the commutative property of matrices in (4.4.1).

3. The condition (C1.3) means that the limiting process of $\{(\mathbf{Z}^{(n)}(t), \eta^{(n)}(t))\}_{t\geq 0}$ is a critical Markov branching process in a random environment.

4. Conditions (C1.4)-(C1.7) imply that for n sufficiently large, the instability property holds for $\{\mathbf{Z}^{(n)}(t)\}_{t\geq 0}$, i.e. $P_{(y,\mathbf{z})}\{\lim_{t\to\infty} \|\mathbf{Z}^{(n)}(t)\|_1 = 0 \text{ or } \infty\} = 1$, for every $(y, \mathbf{z}) \in \mathcal{Y}^{(n)} \times \mathbb{Z}_+^2$.

Now we apply the rescaling to the Markov branching process $\{\mathbf{Z}^{(n)}(t)\}_{t\geq 0}$. Define $\mathbf{X}_n(t) = \mathbf{Z}^{(n)}(nt)/n$. Assuming $\mathbf{Z}^{(n)}(0) = n\mathbf{x}$ and defining $A_n^i(t) = \sum_{\alpha\in\mathcal{I}} \alpha\Lambda_{\alpha+e_i}^{(n)i}(t)$,

we get

$$
\begin{aligned}
\mathbf{X}_n(t) &= \mathbf{x} + \sum_{\alpha \in \mathcal{I}} \alpha n^{-1} Y_\alpha \Big(n^2 \int_0^t \Big(\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns)\mathbf{X}_n^i(s)\Big) ds \Big) \\
&= \mathbf{x} + \sum_{\alpha \in \mathcal{I}} \alpha n^{-1} \tilde{Y}_\alpha \Big(n^2 \int_0^t \Big(\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns)\mathbf{X}_n^i(s)\Big) ds \Big) \\
&\quad + \int_0^t n \Big(\sum_i A_n^i(ns)\mathbf{X}_n^i(s)\Big) ds \\
&\equiv U_n(t) + \int_0^t n \Big(\sum_i A_n^i(ns)\mathbf{X}_n^i(s)\Big) ds,
\end{aligned}
$$

where

$$
\tilde{Y}_\alpha(t) = Y_\alpha(t) - t
$$

and

$$
U_n(t) = \mathbf{x} + \sum_{\alpha \in \mathcal{I}} \alpha n^{-1} \tilde{Y}_\alpha \Big(n^2 \int_0^t \Big(\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns)\mathbf{X}_n^i(s)\Big) ds \Big).
$$

Set $B_n^i(t) = \int_0^t n A_n^i(ns)ds$ and $B_n(t) = (B_n^1(t), B_n^2(t))$. The condition $(C1.2)$ implies that the following commutative property of matrices holds:

$$
B_n(t)\dot{B}_n(t) = \dot{B}_n(t)B_n(t), \tag{4.4.1}
$$

where $\dot{B}_n(t)$ denote the derivative of the matrix $B_n(t)$ w.r.t. t. Then we have

$$
\mathbf{X}_n(t) = e^{B_n(t)} \Big(\mathbf{x} + \int_0^t e^{-B_n(s)} dU_n(s) \Big). \tag{4.4.2}
$$

Set $D_n(t) = (D_n^1(t), D_n^2(t))$, where $D_n^i(t)$ is a 2×2 matrix whose jkth entry equals $\sum_{\alpha \in \mathcal{I}} \alpha^j \alpha^k \Lambda_{\alpha+e_i}^{(n)i}(nt)$. Let $\langle D_n(t), \mathbf{X}_n(t)\rangle$ denote $\sum_i \mathbf{X}_n^i(t)D_n^i(t)$, and $\sqrt{\langle D_n(s), \mathbf{X}_n(s)\rangle}$ denote the square root of the matrix $\langle D_n(t), \mathbf{X}_n(t)\rangle$. Let $(B_n, D_n) \Rightarrow (B, D)$ denote that (B_n, D_n) converge in distribution to (B, D).

When the population is large and the time scale is fast, we can deduce a continuous-state approximation for a sequence of discrete-state Markov branching processes in random environments. Our main result is as follows:

Theorem 4.4.1. *Assume that the conditions (C.1) hold and that $(B_n, D_n) \Rightarrow (B, D)$ where D has continuous paths. Let W be a two-dimensional standard Brownian mo-*

tion independent of B and D. If there exists a global solution $\boldsymbol{X}$ of

$$\boldsymbol{X}(t) = e^{B(t)}\Big(\boldsymbol{x} + \int_0^t e^{-B(s)}\sqrt{\langle D(s), \boldsymbol{X}(s)\rangle}dW(s)\Big), \qquad (4.4.3)$$

and if weak local uniqueness holds for (4.4.3), then $\boldsymbol{X}_n$ converges in distribution to $\boldsymbol{X}$.

Remark: 1. By Proposition 4.2.3, we obtain $E[\mathbf{Z}^{(n)}(nt) \mid \{\Lambda^{(n)}(s)\}_{s\geq 0}] = e^{B_n(t)}\mathbf{Z}^{(n)}(0)$. So B_n acts like the generator of the first moment of the branching process $\mathbf{Z}^{(n)}$. For each $i \in S$, D_n^i equals to the production of the rate of splitting $\lambda^{(n)i}$ and the second moment of increment of offspring distribution $\{p_{\alpha+e_i}^{(n)i}\}_{\alpha\in\mathcal{I}}$. So the assumption that $(B_n, D_n) \Rightarrow (B, D)$ is classical for showing the relative compactness of the processes $\mathbf{X}_n$.

2. Theorem 4.4.1 also holds for the weak convergence of a sequence of one-dimensional birth and death processes in random environments. Such case is also included in Theorem 3.1 of Chap. 9 of [EK86].

3. Theorem 4.4.1 also holds for the weak convergence of a sequence of two-type parallel mutation-selection Markov branching processes processes in constant environment. Such case is also included in Theorem 4.4.2 of [JM86].

4.4.2 Proofs

Firstly, we begin by treating the case of nonrandom, but time dependent environments, i.e. $\Lambda_{\alpha+e_i}^{(n)i}(t)$, $t \geq 0$, is a realization of the stochastic process $\Lambda_{\alpha+e_i}^{(n)i}$ satisfying conditions $(C1)$, for each $i \in S$ and $\alpha \in \mathcal{I}$. Then $B_n(t)$ and $D_n(t)$ are nonrandom functions of t.

Let $V_n(t) \equiv \int_0^t e^{-B_n(s)}dU_n(s)$. The following lemma gives the properties of V_n.

Lemma 4.4.1. *$V_n(t)$ is a locally square integrable martingale and the predictable quadratic variation of $V_n(t)$ is*

$$\langle V_n\rangle(t) = \int_0^t \Big(e^{-B_n(s)}\sqrt{\langle D_n(s), \boldsymbol{X}_n(s)\rangle}\Big)\Big(e^{-B_n(s)}\sqrt{\langle D_n(s), \boldsymbol{X}_n(s)\rangle}\Big)^\top ds. \quad (4.4.4)$$

Proof. Since B_n is continuous and U_n has bounded variation, the stochastic integral V_n is well-defined. In order to show that V_n is a locally square integrable martingale,

it is enough to show that U_n is a locally square integrable martingale since $\{e^{-B_n(t)}\}$ is a locally bounded, predictable process. Recall that

$$U_n(t) = \mathbf{x} + \sum_{\alpha \in \mathcal{I}} \alpha n^{-1} \tilde{Y}_\alpha \left(n^2 \int_0^t \left(\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns) \mathbf{X}_n^i(s) \right) ds \right).$$

Then for each $T > 0$ and $t \leq T$ we get

$$
\begin{aligned}
E\left[\|U_n(t)\|^2\right] &\leq 2\|\mathbf{x}\|^2 + 2E\left[\sum_{\alpha \in \mathcal{I}} \|\alpha\|^2 n^{-2} \tilde{Y}_\alpha^2 \left(n^2 \int_0^t \left(\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns) \mathbf{X}_n^i(s) \right) ds \right)\right] \\
&= 2\|\mathbf{x}\|^2 + 2\sum_{\alpha \in \mathcal{I}} \|\alpha\|^2 n^{-2} E\left[\tilde{Y}_\alpha^2 \left(n^2 \int_0^t (\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns) \mathbf{X}_n^i(s)) ds \right)\right] \\
&= 2\|\mathbf{x}\|^2 + 2\int_0^t \sum_{\alpha \in \mathcal{I}} \|\alpha\|^2 \left(\sum_i \Lambda_{\alpha+e_i}^{(n)i}(ns) E\mathbf{X}_n^i(s) \right) ds \\
&\leq 2\|\mathbf{x}\|^2 + 2\sum_i \left[\int_0^t \left(\sum_{\alpha \in \mathcal{I}} \|\alpha\|^2 \Lambda_{\alpha+e_i}^{(n)i}(ns) \right)^2 ds \right]^{1/2} \left[\int_0^t (E[\mathbf{X}_n^i(s)])^2 ds \right]^{1/2} \\
&< \infty,
\end{aligned}
$$

where $\|U_n(t)\|$ denotes $\sqrt{\sum_{i \in S}(U_n^i(t))^2}$. Hence U_n is a locally square integrable martingale.

It is easy to calculate that the predictable quadratic variation of $V_n(t)$ is

$$\langle V_n \rangle(t) = \int_0^t e^{-B_n(s)} \langle D_n(s), \mathbf{X}_n(s) \rangle \left(e^{-B_n(s)} \right)^\top ds.$$

Since the matrix $\langle D_n(s), \mathbf{X}_n(s) \rangle$ is positive semi-definite, its square root $\sqrt{\langle D_n(s), \mathbf{X}_n(s) \rangle}$ is well-defined. Then

$$\langle V_n \rangle(t) = \int_0^t \left(e^{-B_n(s)} \sqrt{\langle D_n(s), \mathbf{X}_n(s) \rangle} \right) \left(e^{-B_n(s)} \sqrt{\langle D_n(s), \mathbf{X}_n(s) \rangle} \right)^\top ds.$$

$\square$

Following the proof of Doob integral theorem [Kal01], we deduce an alternative form of $V_n(t)$ in Lemma 4.4.2.

Lemma 4.4.2. *Define*

$$W_n(t) = \int_0^t \left(e^{-B_n(s-)} \sqrt{\langle D_n(s-), \mathbf{X}_n(s-) \rangle} \right)^{-1} \pi_{R_{s-}^n} dV_n(s) + \int_0^t \pi_{N_{s-}^n} d\xi(s), \quad (4.4.5)$$

79

where $\left(e^{-B_n(s)}\sqrt{\langle D_n(s), \boldsymbol{X}_n(s)\rangle}\right)^{-1}$ denotes the inverse mapping from R_s^n to $N_s^{n\perp}$ since $e^{-B_n(s)}\sqrt{\langle D_n(s), \boldsymbol{X}_n(s)\rangle}$ is a bijection from $N_s^{n\perp}$ to R_s^n, $\pi_{N_s^n}$ and $\pi_{R_s^n}$ denote the corresponding orthogonal projections of the 2×2 matrix $e^{-B_n(s)}\sqrt{\langle D_n(s), \boldsymbol{X}_n(s)\rangle}$ to its null and range spaces and ξ is a two-dimensional standard Brownian motion. Then

$$V_n(t) = \int_0^t e^{-B_n(s-)}\sqrt{\langle D_n(s-), \boldsymbol{X}_n(s-)\rangle}dW_n(s). \qquad (4.4.6)$$

Proof. By the definition of W_n, we get

$$
\begin{aligned}
&\int_0^t e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}dW_n(s) \\
=\ &\int_0^t \left(e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\right)\left(e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\right)^{-1}\pi_{R_{s-}^n}dV_n(s) \\
&+\int_0^t e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\pi_{N_{s-}^n}d\xi(s) \\
=\ &\int_0^t \pi_{R_{s-}^n}dV_n(s) \\
=\ &\int_0^t \left(\pi_{R_{s-}^n} + \pi_{R_{s-}^n\perp}\right)dV_n(s) \\
=\ &V_n(t).
\end{aligned}
$$

Here the third equality holds since $\int_0^t \pi_{R_{s-}^n\perp}dV_n(s)$ is a right-continuous locally square integrable martingale and its predictable quadratic variation is 0 a.s. $\qquad\square$

Then, (4.4.2) can be written in the following form

$$\mathbf{X}_n(t) = e^{B_n(t)}\left(\mathbf{x} + \int_0^t e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}dW_n(s)\right). \qquad (4.4.7)$$

That is, we characterize $\mathbf{X}_n$, $n \in \mathbb{N}$, as solutions of the above stochastic equations. A convergence theory can be developed using this characterization. If the equations "converge", then one hopes the solutions converge.

The following lemma shows the weak convergence of W_n. Let $\|\cdot\|$ denote a norm of a matrix or a vector.

Lemma 4.4.3. *Assume that $\Lambda_{\alpha+e_i}^{(n)i}(t)$ is a realization of the stochastic process $\Lambda_{\alpha+e_i}^{(n)i}$ satisfying conditions $(C1)$, for each $i \in S$ and $\alpha \in \mathcal{I}$. Assume that for each $T > 0$,*

$\lim_{n\to\infty} \sup_{t\leq T} \|B_n(t) - B(t)\| = 0$, *and* $\lim_{n\to\infty} \sup_{t\leq T} \|D_n(t) - D(t)\| = 0$. *Then* W_n *converges in distribution to a two-dimensional Brownian motion* W *in Skorohod topology, denoted by*

$$W_n(s) \Rightarrow W(s). \tag{4.4.8}$$

Proof. By the martingale central limit theorem (see [EK86]), it is enough to show the following two conditions: For each $T > 0$,

(a) The predictable quadratic variation of W_n is

$$\langle W_n \rangle(t) = tI, \ t \leq T, \tag{4.4.9}$$

where I is a 2×2 identity matrix.

(b)

$$\lim_{n\to\infty} E\left[\sup_{t\leq T} \|W_n(t) - W_n(t-)\|^2\right] = 0. \tag{4.4.10}$$

Proof of (a):

$$
\begin{aligned}
\langle W_n \rangle(t) &= \int_0^t \left(\left(e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\right)^{-1}\pi_{R_{s-}^n}\right)\left(e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\right)\\
&\quad \left(e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\right)^\top\left(\left(e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}\right)^{-1}\pi_{R_{s-}^n}\right)^\top ds\\
&\quad + \int_0^t \pi_{N_{s-}^n}\left(\pi_{N_{s-}^n}\right)^\top ds\\
&= \int_0^t \pi_{N_{s-}^n\perp}\left(\pi_{N_{s-}^n\perp}\right)^\top ds + \int_0^t \pi_{N_{s-}^n}\left(\pi_{N_{s-}^n}\right)^\top ds\\
&= \int_0^t \left(\pi_{N_{s-}^n\perp} + \pi_{N_{s-}^n}\right)ds\\
&= tI.
\end{aligned}
$$

Proof of (b): Firstly, we want to show that

$$\lim_{n\to\infty} P\left\{\inf_{t\leq T} \|\mathbf{X}_n(t)\|_1 = 1/n\right\} = 0.$$

It is equivalent to show that

$$\lim_{n\to\infty} P\left\{\inf_{t\leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1\right\} = 0.$$

For a fixed n, we define

$$\alpha_n = \left\{\omega \in \Omega : \lim_{t\to\infty} \|\mathbf{Z}^{(n)}(t,\omega)\|_1 = 0\right\},$$

and

$$\beta_n = \left\{ \omega \in \Omega : \lim_{t \to \infty} \|\mathbf{Z}^{(n)}(t, \omega)\|_1 = \infty \right\}.$$

Then by the instability property of $\mathbf{Z}^{(n)}$ (see the remark for conditions $(C1.4) -$ $(C1.7)$), we obtain that

$$P\{\alpha_n\} + P\{\beta_n\} = 1.$$

Then

$$
\begin{aligned}
& P\Big\{ \inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \Big\} \\
= \; & P\Big\{ \alpha_n \bigcap \big\{ \inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \big\} \Big\} + P\Big\{ \beta_n \bigcap \big\{ \inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \big\} \Big\} \\
\leq \; & P\Big\{ \alpha_n \bigcap \big\{ \inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \big\} \Big\} + P\Big\{ \beta_n \bigcap \big\{ \|\mathbf{Z}^{(n)}(u)\|_1 = 1, \text{ for some } u \leq nT \big\} \Big\} \\
= \; & P\{\alpha_n\} P\Big\{ \inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \,\Big|\, \alpha_n \Big\} + \\
& P\Big\{ \|\mathbf{Z}^{(n)}(u)\|_1 = 1, \text{ for some } u \leq nT \Big\} P\Big\{ \beta_n \,\Big|\, \|\mathbf{Z}^{(n)}(u)\|_1 = 1, \text{ for some } u \leq nT \Big\}.
\end{aligned}
$$

By the condition $(C1.3)$, the limiting process corresponding to $\mathbf{Z}^{(n)}$, $n \in \mathbb{N}$, is a critical two-type Markov branching process in a random environment. Then $\lim_{n \to \infty} P\{\beta_n \mid \|\mathbf{Z}^{(n)}(u)\|_1 = 1, \text{ for some } u \leq nT\} = 0$. It is obvious that $\lim_{n \to \infty} P\{\inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \mid \alpha_n\} = 0$. Hence

$$\lim_{n \to \infty} P\Big\{ \inf_{t \leq T} \|\mathbf{Z}^{(n)}(nt)\|_1 = 1 \Big\} = 0.$$

Secondly, since for each $T > 0$, $\lim_{n \to \infty} \sup_{t \leq T} \|B_n(t) - B(t)\| = 0$, and $\lim_{n \to \infty} \sup_{t \leq T} \|D_n(t) -$

82

$D(t)\| = 0$, we have

$$E\left[\sup_{t\leq T}\|W_n(t) - W_n(t-)\|^2\right]$$

$$= E\left[\sup_{t\leq T}\left\|\left(e^{-B_n(t-)}\sqrt{\langle D_n(t-), \mathbf{X}_n(t-)\rangle}\right)^{-1}\pi_{R_{t-}^n}(V_n(t) - V_n(t-))\right\|^2\right]$$

$$= E\left[\sup_{t\leq T}\left\|\left(e^{-B_n(t-)}\sqrt{\langle D_n(t-), \mathbf{X}_n(t-)\rangle}\right)^{-1}\pi_{R_{t-}^n}(V_n(t) - V_n(t-))\right\|^2\right]$$

$$= E\left[\sup_{t\leq T}\left\|\left(e^{-B_n(t-)}\sqrt{\langle D_n(t-), \mathbf{X}_n(t-)\rangle}\right)^{-1}\pi_{R_{t-}^n}e^{-B_n(t-)}(U_n(t) - U_n(t-))\right\|^2\right]$$

$$\leq \sum_{k=1}^{\infty}C\cdot\sup_{t\leq T}\frac{n}{k\cdot\min_i\lambda_n^i(t)}\cdot P\left\{\inf_{t\leq T}\|\mathbf{X}_n(t)\|_1 = k/n\right\}\cdot\frac{1}{n^2}\cdot\sup_{\alpha\in\mathcal{I}}\|\alpha\|^2\cdot$$

$$\sup_{t\leq T}\sum_{\alpha\in\mathcal{I}}E\left[Y_\alpha\left(n^2\int_0^t\left(\sum_i\Lambda_{\alpha+e_i}^{(n)i}(ns)\mathbf{X}_n^i(s)\right)ds\right) - Y_\alpha\left(\left(n^2\int_0^t\left(\sum_i\Lambda_{\alpha+e_i}^{(n)i}(ns)\mathbf{X}_n^i(s)\right)ds\right)-\right)\right]$$

$$\to\quad 0,\ \text{as } n\to\infty,$$

where $\lambda_n^i(t)$ is the minimal nonzero eigenvalue of $D_n^i(t)$. $\qquad\square$

The conclusion of the following lemma is an important condition to show the weak convergence of stochastic integrals, denoted by $(C2)$.

Lemma 4.4.4. *Assume that $W_n(s)\Rightarrow W(s)$. Then for each $a > 0$, there exist stopping times τ_n^a such that $P\{\tau_n^a\leq a\}\leq 1/a$ and $\sup_n E[[W_n]_{ij}(t\wedge\tau_n^a)] < \infty$.*

Proof. For $c > 0$, define

$$\tau_n^c = \inf\{t\geq 0\mid \|W_n(t)\|\vee\|W_n(t-)\|\geq c\}.$$

Since convergence in distribution of W_n in the Skorohod topology implies $\sup_{t\leq a}\|W_n(t)\|$ is stochastically bounded in n for each a and hence there exists c_a so that

$$P\{\tau_n^{c_a}\leq a\}\leq 1/a.$$

With $\tau_n^a = \tau_n^{c_a}$, we get

$$P\{\tau_n^a\leq a\}\leq 1/a.$$

Since W_n is a locally square integrable martingale, we get

$$\sup_n E[[W_n]_{ij}(t\wedge\tau_n^a)] = \sup_n E[\langle W_n\rangle_{ij}(t\wedge\tau_n^a)] = E[\delta_{ij}(t\wedge\tau_n^a)] < \infty.$$

$\qquad\square$

For $(y, t) \in D_{\mathbb{R}^2_+}[0, \infty) \times [0, \infty)$, define

$$F_n(y, t) = G_n(y(t), t) = e^{-B_n(t)} \sqrt{\langle D_n(t), y(t) \rangle},$$

and

$$F(y, t) = G(y(t), t) = e^{-B(t)} \sqrt{\langle D(t), y(t) \rangle}.$$

The next lemma gives the properties of $F_n(y)$ and $F(y)$ under transformations of the time scale.

Lemma 4.4.5. *Assume that for each $T > 0$, $\lim_{n \to \infty} \sup_{t \leq T} \|B_n(t) - B(t)\| = 0$, and $\lim_{n \to \infty} \sup_{t \leq T} \|D_n(t) - D(t)\| = 0$ where D is a continuous function of t. Then F_n and F satisfy the following conditions denoted by $(C3)$:*

$(C3)(i)$ For each compact subset $\mathcal{H} \subset D_{\mathbb{R}^2_+}[0, \infty) \times A_1[0, \infty)$ and $T > 0$,

$$\sup_{(y, \lambda) \in \mathcal{H}} \sup_{t \leq T} \|G_n(y(\lambda(t)), \lambda(t)) - G(y(\lambda(t)), \lambda(t))\| \to 0,$$

where A_1 denotes the collection of nondecreasing mappings λ of $[0, \infty)$ onto $[0, \infty)$ (in particular, $\lambda(0) = 0$) such that $\lambda(t + h) - \lambda(t) \leq h$ for all $t, h \geq 0$.

$(C3)(ii)$ For $\{(y_n, \lambda_n)\} \in D_{\mathbb{R}^2_+}[0, \infty) \times A_1[0, \infty)$, $\sup_{t \leq T} \|y_n(\lambda_n(t)) - y(\lambda(t))\| \to 0$ and $\sup_{t \leq T} |\lambda_n(t) - \lambda(t)| \to 0$ for each $T > 0$ implies

$$\sup_{t \leq T} \|G(y_n(\lambda_n(t)), \lambda_n(t)) - G(y(\lambda(t)), \lambda(t))\| \to 0.$$

Proof. (1) Let $\mathcal{H}_1$ be the sectional set of $\mathcal{H}$.

$$
\begin{aligned}
&\sup_{(y, \lambda) \in \mathcal{H}} \sup_{t \leq T} \|G_n(y(\lambda(t)), \lambda(t)) - G(y(\lambda(t)), \lambda(t))\| \\
\leq \quad &\sup_{y \in \mathcal{H}_1} \sup_{t \leq T} \|G_n(y(t), t) - G(y(t), t)\| \\
= \quad &\sup_{y \in \mathcal{H}_1} \sup_{t \leq T} \left\| e^{-B_n(t)} \sqrt{\langle D_n(t), y(t) \rangle} - e^{-B(t)} \sqrt{\langle D(t), y(t) \rangle} \right\| \\
\to \quad &0,
\end{aligned}
$$

since $\lim_{n \to \infty} \sup_{t \leq T} \|B_n(t) - B(t)\| = 0$, $\lim_{n \to \infty} \sup_{t \leq T} \|D_n(t) - D(t)\| = 0$ for each $T > 0$.

84

(2) Since D has continuous paths, the function G is continuous. Therefore,

$$\sup_{t \le T} \|G(y_n(\lambda_n(t)), \lambda_n(t)) - G(y(\lambda(t)), \lambda(t))\| \to 0$$

when $\sup_{t \le T} \|y_n(\lambda_n(t)) - y(\lambda(t))\| \to 0$ and $\sup_{t \le T} |\lambda_n(t) - \lambda(t)| \to 0$ for each $T > 0$. $\qquad \square$

Now we are ready to show the weak convergence of the sequence $\mathbf{X}_n$ in the case of nonrandom environment.

Theorem 4.4.2. *Assume that $\Lambda_{\alpha+e_i}^{(n)i}(t)$ is a realization of the stochastic process $\Lambda_{\alpha+e_i}^{(n)i}$ satisfying conditions (C1), for each $i \in S$ and $\alpha \in \mathcal{I}$. Assume that for each $T > 0$, $\lim_{n \to \infty} \sup_{t \le T} \|B_n(t) - B(t)\| = 0$, and $\lim_{n \to \infty} \sup_{t \le T} \|D_n(t) - D(t)\| = 0$, where D is a continuous function of t. Let W be a two-dimensional standard Brownian motion. If there exists a global solution $\boldsymbol{X}$ of*

$$\boldsymbol{X}(t) = e^{B(t)}\Big(\boldsymbol{x} + \int_0^t e^{-B(s)}\sqrt{\langle D(s), \boldsymbol{X}(s)\rangle}dW(s)\Big), \qquad (4.4.11)$$

and if weak local uniqueness holds for (4.4.11), then $\boldsymbol{X}_n$ converges in distribution to $\boldsymbol{X}$.

Proof. The proof follows from Theorem 5.4 in [KP91]. Firstly, by Lemma 4.4.1 and 4.4.2, we get

$$
\begin{aligned}
\mathbf{X}_n(t) &= e^{B_n(t)}\Big(\mathbf{x} + \int_0^t e^{-B_n(s-)}\sqrt{\langle D_n(s-), \mathbf{X}_n(s-)\rangle}dW_n(s)\Big) \\
&= e^{B_n(t)}\Big(\mathbf{x} + \int_0^t F_n(\mathbf{X}_n, s-)dW_n(s)\Big).
\end{aligned}
$$

By Lemma 4.4.3, we obtain

$$W_n \Rightarrow W,$$

where W is a two-dimensional standard Brownian motion.

Secondly, for $b > 0$, define

$$\zeta_n^b = \inf\{t \mid \|F_n(\mathbf{X}_n, t)\| \vee \|F_n(\mathbf{X}_n, t-)\| \ge b\}$$

and let $\mathbf{X}_n^b$ denote the solution of

$$\mathbf{X}_n^b(t) = e^{B_n(t)}\left(\mathbf{x} + \int_0^t \chi_{[0,\zeta_n^b)}(s-)F_n(\mathbf{X}_n^b, s-)dW_n(s)\right) \tag{4.4.12}$$

that agrees with $\mathbf{X}_n$ on $[0, \zeta_n^b)$. We want to show that $\{(\mathbf{X}_n^b, W_n)\}$ is relatively compact and any limit point $(\mathbf{X}^b, W)$ gives a local solution $(\mathbf{X}^b, \tau)$ of the equation (4.4.11) with $\tau = \zeta^c \equiv \inf\{t \mid \|F(\mathbf{X}^b, t)\| \vee \|F(\mathbf{X}^b, t-)\| \geq c\}$ for any $c < b$.

Let $H_n(t) = \sup_{s \leq t} \|\chi_{[0,\zeta_n^b)}(s)F_n(\mathbf{X}_n^b, s)\|$. Then $\{H_n(t)\}$ is stochastically bounded for each t. Define

$$\mathbf{Y}_n^b(t) = \int_0^t \chi_{[0,\zeta_n^b)}(s-)F_n(\mathbf{X}_n^b, s-)dW_n(s).$$

By Lemma 4.4.4 we have that W_n satisfies the condition $(C2)$. Then by Lemma 4.1 in [KP91], there exist nondecreasing, adapted processes γ_n, with $\gamma_n(0) = 0$ and $\gamma_n(t+h) - \gamma_n(t) \leq h$ for all $t, h \geq 0$, such that, $\{(W_n \circ \gamma_n, \mathbf{Y}_n^b \circ \gamma_n, \gamma_n)\}$ is relatively compact in $D_{\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}}[0, \infty)$. By the continuity and convergence of B_n, it is obvious that $\{(W_n \circ \gamma_n, \mathbf{X}_n^b \circ \gamma_n, \gamma_n)\}$ is relatively compact in $D_{\mathbb{R}^2 \times \mathbb{R}^2 \times \mathbb{R}}[0, \infty)$. $\{\chi_{[0,\zeta_n^b)} \circ \gamma_n F_n(\mathbf{X}_n^b) \circ \gamma_n\} = \{\chi_{[0,\zeta_n^b)} \circ \gamma_n G_n(\mathbf{X}_n^b \circ \gamma_n, \gamma_n)\}$ is relatively compact in $D_{\mathbb{R}^{2 \times 2}}[0, \infty)$ since $\{F_n\}$ and F satisfy the conditions $(C3)$ by Lemma 4.4.5. Applying Corollary 4.5 and Proposition 4.3 in [KP91], we get the relative compactness of $\{(\mathbf{X}_n^b, W_n)\}$ in $D_{\mathbb{R}^2 \times \mathbb{R}^2}[0, \infty)$. The sequence $\{(\mathbf{X}_n^b, W_n, \zeta_n^b)\}$ is relatively compact in $D_{\mathbb{R}^2 \times \mathbb{R}^2}[0, \infty) \times [0, \infty)$. Let $(\mathbf{X}^b, W, \zeta_0^b)$ denote a weak limit point. To simplify notation, assume that the original sequence converges and (with reference to the Skorohod representation theorem) assume that the convergence is almost sure rather than in distribution. Note that $\zeta^b \equiv \inf\{t \mid \|F(\mathbf{X}^b, t)\| \vee \|F(\mathbf{X}^b, t-)\| \geq b\} \leq \zeta_0^b$. It follows that

$$e^{B_n}\left(\mathbf{x} + \int F_n(\mathbf{X}_n^b)dW_n\right) \to e^{B}\left(\mathbf{x} + \int F(\mathbf{X}^b)dW\right).$$

Since

$$\mathbf{X}_n^b(t) = e^{B_n(t)}\left(\mathbf{x} + \int_0^t F_n(\mathbf{X}_n^b, s-)dW_n(s)\right) \tag{4.4.13}$$

for $t \leq \zeta_n^b$, we get

$$\mathbf{X}^b(t) = e^{B(t)}\left(\mathbf{x} + \int_0^t F(\mathbf{X}^b, s-)dW(s)\right) \tag{4.4.14}$$

86

for $t < \zeta_0^b$. Let $c < b$. If $\zeta^c < \zeta^b$, then (4.4.14) holds for $t \leq \zeta^c$. If $\zeta^c = \zeta^b$, then $F(\mathbf{X}^b)$ has a discontinuity at ζ^c with $\|F(\mathbf{X}^b, \zeta^c-)\| \leq c$ and $\|F(\mathbf{X}^b, \zeta^c)\| \geq b$. It follows that for $c < d < b$, $(\mathbf{X}_n^b(\zeta_n^d), W_n(\zeta_n^d), W_n(\zeta_n^d-), F_n(\mathbf{X}_n^b, \zeta_n^d), F_n(\mathbf{X}_n^b, \zeta_n^d-), \zeta_n^d)$ converges to $(\mathbf{X}^b(\zeta^d), W(\zeta^d), W(\zeta^d-), F(\mathbf{X}^b, \zeta^d), F(\mathbf{X}^b, \zeta^d-), \zeta^d)$ and

$$\mathbf{X}^b(\zeta^d) = e^{B(\zeta^d)}\Big(\mathbf{x} + \int_0^{\zeta^d} F(\mathbf{X}^b, s-) dW(s)\Big), \tag{4.4.15}$$

so that (4.4.14) holds for $t \leq \zeta^c (= \zeta^d)$. Consequently, $(\mathbf{X}^b, \zeta^c)$ is a local solution of the equation (4.4.11).

Finally, we will show that $\{(\mathbf{X}_n, W_n)\}$ converge in distribution to $(\mathbf{X}, W)$ if there exists a global weak solution $\mathbf{X}$ for (4.4.11) and weak local uniqueness holds.

Note that ζ^c is a measurable function of $\mathbf{X}^b$ (say $h_c(\mathbf{X}^b)$). Consequently, if there exists a global weak solution $\mathbf{X}$ for (4.4.11) and weak local uniqueness holds, then $(\mathbf{X}^b, \zeta^c)$ must have the same distribution as $(\mathbf{X}, h_c(\mathbf{X}))$ for all c and b with $c < b$. Since $\mathbf{X}$ is a global solution, $h_c(\mathbf{X}) \to \infty$ as $c \to \infty$. Convergence in distribution of $(\mathbf{X}_n, W_n)$ follows. $\qquad\square$

Now we could easily deduce the corresponding conclusion in the case of random environments by the Skorohod representation theorem and the technique of conditioning on the intensity processes. Recall that the intensity processes $\Lambda^{(n)}$ satisfy the assumptions (C1).

Proof of Theorem 4.4.1. Since $(B_n, D_n) \Rightarrow (B, D)$ with (B, D) having continuous paths, we may assume by the Skorohod representation theorem that

$$\lim_{n \to \infty} \sup_{s \leq t} \|B_n(s) - B(s)\| = 0 \ a.s., \tag{4.4.16}$$

$$\lim_{n \to \infty} \sup_{s \leq t} \|D_n(s) - D(s)\| = 0 \ a.s., \tag{4.4.17}$$

for all $t > 0$.

Let $F_\Lambda^{(n)} = \sigma\{\Lambda_{\alpha+e_i}^{(n)i}(s), \ s \geq 0, \ i \in S, \ \alpha \in \mathcal{I}\}$. From Lemma 4.4.3, one can conclude that for every bounded continuous functional f on $D_{\mathbb{R}^2}[0, \infty)$

$$\lim_{n \to \infty} E\Big[f(W_n) \ \Big| \ F_\Lambda^{(n)}\Big] = E[f(W)] \ a.s. \tag{4.4.18}$$

Consequently, it follows that

$$
\begin{aligned}
\lim_{n\to\infty} E[f(W_n)g(B_n, D_n)] &= \lim_{n\to\infty} E\Big[E\Big[f(W_n)g(B_n, D_n) \,\Big|\, F_\Lambda^{(n)}\Big]\Big] \\
&= \lim_{n\to\infty} E\Big[g(B_n, D_n)E\Big[f(W_n) \mid F_\Lambda^{(n)}\Big]\Big] \\
&= E[g(B, D)E[f(W)]] \\
&= E[f(W)]E[g(B, D)],
\end{aligned}
$$

for all bounded continuous functionals f on $D_{\mathbb{R}^2}[0, \infty)$ and g on $D_{\mathbb{R}^{2\times 2}\times\mathbb{R}^{2\times 4}}[0, \infty)$. The weak convergence $(W_n, B_n, D_n) \Rightarrow (W, B, D)$ follows. Meanwhile, we get the independence of W and (B, D).

Let $F_\Lambda^{(\infty)} = \sigma\{B(s), D(s), s \geq 0\}$. By Theorem 4.4.2, for all bounded continuous functional h on $D_{\mathbb{R}_+^2}[0, \infty)$, we get

$$
\lim_{n\to\infty} E\Big[h(\mathbf{X}_n) \,\Big|\, F_\Lambda^{(n)}\Big] = E\Big[h(\mathbf{X}) \,\Big|\, F_\Lambda^{(\infty)}\Big] \quad a.s., \tag{4.4.19}
$$

where $\mathbf{X}$ is the solution of

$$
\mathbf{X}(t) = e^{B(t)}\Big(\mathbf{x} + \int_0^t e^{-B(s)}\sqrt{\langle D(s), \mathbf{X}(s)\rangle}dW(s)\Big). \tag{4.4.20}
$$

Therefore,

$$
\begin{aligned}
\lim_{n\to\infty} E[h(\mathbf{X}_n)] &= \lim_{n\to\infty} E\Big[E\Big[h(\mathbf{X}_n) \,\Big|\, F_\Lambda^{(n)}\Big]\Big] \\
&= E\Big[E\Big[h(\mathbf{X}) \,\Big|\, F_\Lambda^{(\infty)}\Big]\Big] \\
&= E[h(\mathbf{X})].
\end{aligned}
$$

Hence the conclusion of Theorem 4.4.1 holds. $\qquad\square$

4.5 Examples

Example 1. One-dimension and random environment

Let $\mathbf{Z}^{(n)}(0) = nx_0$. Let $\eta(t)$ be a standard Poisson process. Let $\Lambda_0^{(n)}(t) = 1$, $\Lambda_2^{(n)}(t) = 1 + n^{-1/2}(-1)^{\eta(t)}$, and $\Lambda_k^{(n)}(t) = 0$ for $k \neq 0, 2$. This gives

$$
B_n(t) = \int_0^t n^{1/2}(-1)^{\eta(ns)}ds \tag{4.5.1}
$$

and

$$D_n(t) = 2 + n^{-1/2}(-1)^{\eta(nt)}. \tag{4.5.2}$$

Then $(B_n, D_n) \Rightarrow (B, D)$ where B is a standard one-dimensional Brownian motion and $D(t) = 2$. It is easy to verify the other conditions in Theorem 4.4.1 in one dimensional case. Then $\mathbf{X}_n \Rightarrow \mathbf{X}$, where $\mathbf{X}$ is the solution of

$$\mathbf{X}(t) = e^{B(t)}\left(x_0 + \int_0^t e^{-B(s)}\sqrt{2\mathbf{X}(s)}dW(s)\right). \tag{4.5.3}$$

Note that B and $M(\cdot) = \int_0^\cdot e^{-B(s)}\sqrt{2\mathbf{X}(s)}dW(s)$ are martingales with $\langle B \rangle_t = t$, $\langle B, M \rangle_t = 0$, and $\langle M \rangle_t = \int_0^t 2e^{-2B(s)}\mathbf{X}(s)ds$. Then by Itô's formula, we see that $\mathbf{X}$ is a solution of the martingale problem for L with

$$Lf(x) = \frac{1}{2}xf'(x) + \left(x + \frac{1}{2}x^2\right)f''(x), \tag{4.5.4}$$

for $f \in C_c^2(\mathbb{R}_+)$.

It is well known [KT81] that the extinction probability of the diffusion process starting from state x_0 is

$$\begin{aligned}
q(x_0) &= P\left\{\lim_{t\to\infty}\mathbf{X}(t) = 0 \,\Big|\, \mathbf{X}(0) = x_0\right\} \\
&= (1 + x_0/2)^{-1}.
\end{aligned}$$

Example 2. Two-dimension and constant environment

Let $\Lambda^{(n)1}_{(0,0)^\top} = 1/2 - 1/4n$, $\Lambda^{(n)1}_{(2,0)^\top} = 1/2$, $\Lambda^{(n)1}_{(0,1)^\top} = 1/4n$, $\Lambda^{(n)2}_{(0,0)^\top} = 1/2 - 1/2n$, $\Lambda^{(n)2}_{(0,2)^\top} = 1/2$, $\Lambda^{(n)2}_{(1,0)^\top} = 1/2n$, and else 0.

Then

$$B_n(t) = \begin{pmatrix} 0 & t/2 \\ t/4 & 0 \end{pmatrix} \Rightarrow B(t) = \begin{pmatrix} 0 & t/2 \\ t/4 & 0 \end{pmatrix},$$

$$D_n^1(t) = \begin{pmatrix} 1 & -1/4n \\ -1/4n & 1/4n \end{pmatrix} \Rightarrow D^1(t) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$D_n^2(t) = \begin{pmatrix} 1/2n & -1/2n \\ -1/2n & 1 \end{pmatrix} \Rightarrow D^2(t) = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

The limit $\mathbf{X}$ then satisfies

$$\mathbf{X}(t) = e^{B(t)} \left(\mathbf{x} + \int_0^t e^{-B(s)} \sqrt{\begin{pmatrix} X_1(s) & 0 \\ 0 & X_2(s) \end{pmatrix}} \, dW(s) \right). \tag{4.5.5}$$

By Itô's formula, we see that $\mathbf{X}$ is a solution of the martingale problem in $\mathbb{R}_+^2$ associated with the diffusion operator $L = \sum_i a_i \frac{\partial}{\partial x_i} + \sum_{i,j} b_{ij} \frac{\partial^2}{\partial x_i x_j}$ and with initial condition $\mathbf{x}$, where $a_1 = x_2/2$, $a_2 = x_1/4$, $b_{11} = x_1/2$, $b_{12} = b_{21} = 0$, $b_{22} = x_2/2$.

The uniqueness of the solution of the martingale problem in $\mathbb{R}_+^2$ associated with the diffusion operator $L = \sum_i a_i \frac{\partial}{\partial x_i} + \sum_{i,j} b_{ij} \frac{\partial^2}{\partial x_i x_j}$ and with initial condition $\mathbf{x}$ is easily seen in the following way [JM86]. It is a standard result in the study of martingale problems that, if the solutions Φ_x of the martingale problem $(L, C_0, \mathbf{x})$ for all $\mathbf{x} \in \mathbb{R}_+^2$ are such that for any $t > 0$ the law $X_t^{-1} \circ \tilde{P}_{\mathbf{x}}$ is uniquely determined, then the marginals $(X_{t_1}^{-1} \circ \tilde{P}_{\mathbf{x}}, \cdots, X_{t_n}^{-1} \circ \tilde{P}_{\mathbf{x}})$ are uniquely determined for all $\mathbf{x}$ and any finite family $t_0 < t_1 < \cdots < t_n$. Therefore $\tilde{P}_{\mathbf{x}}$ is unique for any initial condition $\mathbf{x}$. If $\tilde{P}$ is any limit law of the laws $\tilde{P}_n$ of the processes X_n, we need only to check that the moments of the X_t under $\tilde{P}$ are uniquely determined. But if ψ is a monomial of degree k, the particular form of L gives

$$E[\psi(X_t)] - \psi(\mathbf{x}) - \int_0^t \sum_i \alpha_i^k E[\psi^{k,i}(X_u)] du - \int_0^t \sum_i \beta_i^k E[\psi^{k-1,i}(X_u)] du = 0,$$

where $\{\psi^{k,i}\}_i$ is the family of all monomials of degree k, and α_i^k and β_i^k are constants bounded by $Kk(k-1)$. The moments of order k are therefore recursively determined.

Hence by Theorem 4.4.1, $\mathbf{X}_n \Rightarrow \mathbf{X}$, where $\mathbf{X}$ is the unique solution of the equation (4.5.5).

It is interesting to study the extinction probability of the diffusion process $\mathbf{X}$. From the differential form of the stochastic equation of $\mathbf{X}$, it is easy to know that $\mathbf{0}$ is an absorbing state of the process $\mathbf{X}$ and that the process returns to the interior of the domain $\mathbb{R}_+^2$ once it hits the boundary of the domain $\mathbb{R}_+^2$ except for $\mathbf{0}$. We will show the difficulty to calculate extinction probability which also appears in the next example in the next subsection. Let $q(\mathbf{x})$ be the extinction probability of the diffusion process

$\mathbf{X}$, i.e.

$$q(\mathbf{x}) = P\Big\{ \lim_{t\to\infty} \mathbf{X}(t) = \mathbf{0} \ \Big| \ \mathbf{X}(0) = \mathbf{x} \Big\}.$$

Let $u_{a,b}(\mathbf{x}) = P\{\tau_a < \tau_b \mid \mathbf{X}(0) = \mathbf{x}\}$, for $0 < a < b < \infty$, where $\tau_a = \inf\{t \geq 0 \mid \|\mathbf{X}(t)\| = a\}$ and $\tau_b = \inf\{t \geq 0 \mid \|\mathbf{X}(t)\| = b\}$. Then $q(\mathbf{x}) = \lim_{a\downarrow 0, b\uparrow \infty} u_{a,b}(\mathbf{x})$. $u_{a,b}(\mathbf{x})$ satisfies the equation

$$L u_{a,b}(\mathbf{x}) = 0, \text{ for } a < \|\mathbf{x}\| < b, \tag{4.5.6}$$

with $u_{a,b}(\mathbf{x}) = 1$ for $\|\mathbf{x}\| = a$ and $u_{a,b}(\mathbf{x}) = 0$ for $\|\mathbf{x}\| = b$. So the difficulty lies in: on the one hand, what kind of boundary conditions is reasonable?; on the other hand, what about the existence and uniqueness of the solution with a degenerate elliptic operator?

Example 3. Two-dimension and random environment

Let $\eta(t)$ be a standard Poisson process. Let $\Lambda^{(n)1}_{(0,0)^\top}(t) \equiv 1, \Lambda^{(n)1}_{(2,0)^\top}(t) = 1+n^{-1/2}(-1)^{\eta(t)}$, $\Lambda^{(n)1}_{(0,1)^\top}(t) = n^{-1}, \Lambda^{(n)2}_{(0,0)^\top}(t) \equiv 1, \Lambda^{(n)2}_{(0,2)^\top}(t) = 1 + n^{-1/2}(-1)^{\eta(t)}, \Lambda^{(n)2}_{(1,0)^\top}(t) = (2n)^{-1}$, and else 0.

By the definition of B_n, we have

$$B_n(t) = \begin{pmatrix} \int_0^t n^{1/2}(-1)^{\eta(ns)}\,ds - t & t/2 \\ t & \int_0^t n^{1/2}(-1)^{\eta(ns)}\,ds - t/2 \end{pmatrix}$$

Then the commutative property holds, i.e.

$$B_n(t)\dot{B}_n(t) = \dot{B}_n(t)B_n(t). \tag{4.5.7}$$

Remark: Condition (4.5.7) helps to make sure that the two intensity stochastic processes are driven by the same underlying environmental process. For example, if there are two independent standard Poisson processes $\eta_1(t)$ and $\eta_2(t)$ such that $\Lambda^{(n)1}_{(0,0)^\top}(t) \equiv 1, \Lambda^{(n)1}_{(2,0)^\top}(t) = 1 + n^{-1/2}(-1)^{\eta_1(t)}, \Lambda^{(n)1}_{(0,1)^\top}(t) = n^{-1}, \Lambda^{(n)2}_{(0,0)^\top}(t) \equiv 1, \Lambda^{(n)2}_{(0,2)^\top}(t) = 1 + n^{-1/2}(-1)^{\eta_2(t)}, \Lambda^{(n)2}_{(1,0)^\top}(t) = (2n)^{-1}$, and else 0, then Condition (4.5.7) holds if and only if $\eta_1(t) = \eta_2(t)$.

Similarly, we have

$$D_n^1(t) = \begin{pmatrix} 2 + n^{-1/2}(-1)^{\eta(nt)} + n^{-1} & -n^{-1} \\ -n^{-1} & n^{-1} \end{pmatrix},$$

and

$$D_n^2(t) = \begin{pmatrix} (2n)^{-1} & -(2n)^{-1} \\ -(2n)^{-1} & 2 + n^{-1/2}(-1)^{\eta(nt)} + (2n)^{-1} \end{pmatrix}.$$

Let ξ denote a one-dimensional standard Brownian motion. Then we get the convergence of (B_n, D_n) as follows:

$$B_n(t) \Rightarrow B(t) = \begin{pmatrix} \xi(t) - t & t/2 \\ t & \xi(t) - t/2 \end{pmatrix},$$

$$D_n^1(t) \Rightarrow D^1(t) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix},$$

and

$$D_n^2(t) \Rightarrow D^2(t) = \begin{pmatrix} 0 & 0 \\ 0 & 2 \end{pmatrix}.$$

Therefore the limit $\mathbf{X}$ satisfies

$$\mathbf{X}(t) = e^{B(t)} \left(\mathbf{x} + \int_0^t e^{-B(s)} \sqrt{\begin{pmatrix} 2X_1(s) & 0 \\ 0 & 2X_2(s) \end{pmatrix}} \, dW(s) \right). \tag{4.5.8}$$

By Itô's formula, we see that $\mathbf{X}$ is a solution of the martingale problem in $\mathbb{R}_+^2$ associated with the operator $L = \sum_i a_i \frac{\partial}{\partial x_i} + \sum_{i,j} b_{ij} \frac{\partial^2}{\partial x_i x_j}$ and with initial condition $\mathbf{x}$, where $a_1 = x_2/2$, $a_2 = x_1 + x_2/2$, $b_{11} = x_1 + x_1^2/2$, $b_{12} = b_{21} = x_1 x_2/2$, $b_{22} = x_2 + x_2^2/2$. The uniqueness of the solution of the martingale problem follows the same argument as in the last example. Then by Theorem 4.4.1, $\mathbf{X}_n \Rightarrow \mathbf{X}$, where $\mathbf{X}$ is the unique solution of the equation (4.5.8).

Bibliography

[AGD94] L. Arnold, V.M. Gundlach and L. Demetrius. *Evolutionary formalism for Products of Positive Random Matrices.* The Annals of Applied Probability, Vol. 4, No. 3. (Aug., 1994), pp. 859-901.

[AK70] K.B. Athreya and S. Karlin. *Branching processes in random environments.* PAMS 76, 865-870 (1970).

[AK72a] K.B. Athreya and S. Karlin. *On branching processes in random environments: I. Extinction probability.* AMS 42, 1499-1520 (1972).

[AK72b] K.B. Athreya and S. Karlin. *On branching processes in random environments: II. Limit theorems.* AMS 42, 1843-1858 (1972).

[AN72] K.B. Athreya and P.E. Ney. *Branching Processes.* Berlin: Springer-Verlag, 1972.

[BC81] R.D. Bourgin and R. Cogburn. *On Determining Absorption Probabilities for Markov Chains in Random Environments.* Advances in Applied Probability 13, 369-387, 1981.

[Ber05] M.P. Bertsekas. *Dynamic Programming and Optimal Control, 3rd edn.* Athena Scientific, 2005.

[BG07] E. Baake and H.O. Georgii. *Mutation, selection, and ancestry in branching models: a variational approach.* J. Math. Biol. 54: 257-303 (2007).

[BL04]   C.T. Bergstrom and M. Lachmann. *Shannon Information and Biological Fitness.* Information Theory Workshop, IEEE, pp. 50-54, 2004.

[Bor02]   K. Borovkov. *A Note on Diffusion-Type Approximation to Branching Processes in Random Environments.* Theory Probab. Appl., Vol. 47, No. 1 (2002).

[Cog80]   R. Cogburn. *Markov Chains in Random Environments: The Case of Markovian Environments.* The Annals of Probability, Vol. 8, No. 5 (Oct., 1980), pp. 908-916.

[CT81]   R. Cogburn and W.C. Torrez. *Birth and Death Processes with Random Environments in Continuous Time.* Journal of Applied Probability, Vol. 18, No. 1 (Mar., 1981), pp. 19-30.

[Dem83]   L. Demetrius. *Statistical Mechanics and Population Biology.* Journal of Statistical Physics, Vol. 30, No. 3, 1983.

[Dem97]   L. Demetrius. *Directionality principles in thermodynamics and evolution.* Proc. Natl. Acad. Sci. USA, Vol. 94, pp. 3491-3498, April 1997.

[DMB09] C. Dombry, C. Mazza and V. Bansaye. *Phenotypic diversity and population growth in fluctuating environment: a MBPRE approach.* arXiv: 0912.1194v1 [math.PR] Dec.7 2009.

[DS89]   J.D. Deuschel and D.W. Stroock. *Large deviations.* Academic Press, Boston, MA, 1989. Reprinted: American Mathematical Society, Providence, RI, 2001.

[EK86]   S.N. Ethier and T.G. Kurtz. *Markov Processes: Characterization and convergence.* John wiley and sons (1986).

[Fel50]   W. Feller. *Diffusion processes in genetics.* Proc. Second Berkeley Symp. Math. Statist Prob. Univ. of Calofornia Press, 227-246, 1950.

94

[Fel71]   W. Feller. *An Introduction to Probability Theory and Its Applications, 2nd Edition.* Wiley Series in Probability and Mathematical Statistics, 1971.

[FK60]   H. Furstenberg and H. Kesten. *Products of random matrices.* Ann. Math. Statist. 31, 457-469, 1960.

[GB03]   H.O. Georgii and E. Baake. *Multitype branching processes: the ancestral types of typical individuals.* Adv. Appl. Prob. 35, 1090C1110 (2003).

[Gre80]   D.R. Grey. *Minimisation of extinction probabilities in reproducing populations.* Theoretical Population Biology 18, 430-443 (1980).

[Hal82]   P. Hall. *Rates of Convergence in the Central Limit Theorem.* In Research Notes in Mathematics. (Vol. 62, p. 6). Pitman Advanced Publishing Program (1982).

[Har63]   T.E. Harris. *The theory of branching procresses.* Springer, 1963.

[Hel81]   I.S. Helland. *Minimal Conditions for Weak Convergence to a Diffusion Process on the Line.* The Annals of Probability, Vol. 9, No. 3 (Jun., 1981), pp. 429-452.

[HI95]   P. Haccou and Y. Iwasa. *Optimal mixed strageties in stochastic environments.* Theoretical Population Biology, 47, 212-243, 1995.

[HJV05]   P. Haccou, P. Jagers and V.A. Vatutin. *Branching Processes: Variation, Growth, and Extinction of Populations.* Cambride (2005).

[Jag89]   P. Jagers. *General branching processes as Markov fields.* Stoch. Proc. Appl. 32, 183C242 (1989).

[Jag92]   P. Jagers. *Stabilities and instabilities in population dynamics.* J. Appl. Prob. 29, 770C780 (1992).

[Jir69]   M. Jiřina. *On Feller's branching diffusion processes.* Časopis pro pěstování matematiky, roč. 94 (1969), Praha.

[JL08]   P. Jagers and A.N. Lagerås. *General branching processes conditioned on extinction are still branching processes.* Elect. Comm. in Probab. 13 (2008), 540-547.

[JM86]   A. Joffe and M. Metivier. *Weak convergence of sequences of semimartingales with applications to multitype branching processes.* Advances in Applied Probability, Vol. 18, No. 1 (Mar., 1986), pp. 20-65.

[JN96]   P. Jagers and O. Nerman. *The asymptotic composition of supercritical multitype branching populations.* In Séminaire de Probabilités XXX (Lecture Notes Math. 1626), eds J. Azéma, M. Emery, and M. Yor, Springer, Berlin, pp. 40-54.

[Jos05]   J. Jost. *Dynamical Systems. Examples of Complex Behaviour.* Springer, 2005.

[JP08]   J. Jost and J. Pepper. *Individual optimization efforts and population dynamics: a mathematical model for the evolution of resource allocation strategies, with applications to reproductive and mating systems.* Theory Biosci, 127(1): 31-43, Mar. 2008.

[KA02]   M. Kimmel and D.E. Axelrod. *Branching Processes in Biology.* Springer, 2002.

[Kal01]   O. Kallenberg. *Foundation of modern probability.* Springer, 2001.

[Kei75]   N. Keiding. *Extinction and Exponential Growth in Random Environments.* Theor. Population Biology **8**, 49-63 (1975).

[KLP97]   T. Kurtz, R. Lyons, R. Pemantle and Y. Peres. *Aconceptual proof of the Kesten-Stigum theorem for multitype branching processes.* In: K. B. Athreya, P. Jagers (eds.) Classical and Modern Branching Processes, pp. 181C185. Springer, New York (1997).

[KM80]   B. Klein and P.D.M. Macdonald. *The multitype continuous-time Markov branching process in a periodic environment.* Advances in Applied Probability, Vol. 12, No. 1, 81-93, 1980.

[KP91]   T.G. Kurtz and P. Protter. *Weak Limit Theorems for Stochastic Integrals and Stochastic Differential Equations.* The Annals of Probability, Vol. 19, No. 3 (Jul., 1991), pp. 1035-1070.

[KS66]   H. Kesten and B.P. Stigum. *A limit theorem for multidimensional Galton-Watson processes.* Ann. Math. Statist. 37, 1211C1233 (1966).

[KT81]   K.S. Karlin and H.M. Taylor. *A second course in stochastic processes, 2nd edn.* Academic Press, San Diego (1975).

[Kur78] T.G. Kurtz. *Diffusion approximaitons for branching processes.* Branching processes, Adv. Probab. Related Topics 5, A. Joffe and P. Ney, eds., Marcel Dekker, Inc., New York, 1978, pp. 445-450.

[LO88]   R. Lande and S.H. Orzack. *Extinction dynamics of age-structured populations in a fluctuating environment.* Proc. Natl. Acad. Sci. USA, Population Biology, Vol. 85, pp. 7418-7421, October 1988.

[LPP95] R. Lyons, R. Pemantle, and Y. Peres. *Conceptual proofs of L log L criteria for mean behavior of branching processes.* Ann. Probab. 23, 1125-1138, 1995.

[MHC01] J.M. McNamara, A.I. Houston, and E.J. Collins. *Optimality Models in Behavioral Biology.* Siam Review, Vol. 43, No. 3, pp. 413-466, 2001.

[Rue78] D. Ruelle. *Thermodynamic formalism.* Encyclopedia of mathematics and its application, 5, Addison-Wesley, Reading, MA.

[Sut05] W.J. Sutherland. *The best solution.* Nature, Vol. 435, 2, June 2005.

[Tan81] D. Tanny. *On multitype Branching processes in random environments.* Adv. Appl. Prob., 13, 464-497, 1981.

[TO80]   S.D. Tuljapurkar and S.H. Orzack. *Population dynamics in variable environments, I. Long-run growth rates and extinction.* Theor. Popul. Biol. 18, 314-342 (1980).

[Tor78]   W.C. Torrez. *The birth and death chain in a random environment: instability and extinction theorems.* Ann. Prob. 6, 1026-1043 (1978).

[Var88]   S.R.S Varadhan. *Large deviations.* Lectures Notes Math. 1362, ed. P.L. Hennequin, Springer, Berlin, pp. 1-49, 1988.

[Wal82]   P. Walters. *An Introduction to Ergodic Theory.* Springer, 1982.

[Wil69]   W.E. Wilkinson. *On calculating extinction probabilities for branching processes in random environments.* J. Appl. Prob. 6, 478-492, 1969.

Wissenschaftlicher Werdegang des Autors

- Sep. 2000 – Jul. 2004 Studium der Mathematik an der Universität Shandong, Shandong, V. R. China, Bachelor of Science

- Sep. 2004 – Sep. 2006 Studium der Mathematik an der Academy of Mathematics and Systems Science, Chinesische Akademie der Wissenschaften, Peking, V. R. China

- seit Okt. 2006 Promotionsstudium am Max Planck Institut für Mathematik in den Naturwissenschaften, Leipzig, mit Stipenium des CAS joint doctoral promotion program und der International Max Planck Research School

- Sep. 2007 – Feb. 2008 Semesterprogramm ,,Complex Stochastic Systems: Discrete vs. Continuous," Hausdorff Research Institute for Mathematics, Bonn

Bibiliographische Daten

Selbstständigkeitserklärung

Hiermit erkläre ich, die vorliegende Dissertation selbständig und ohne unzulässige fremde Hilfe angefertigt zu haben. Ich habe keine anderen als die angeführten Quellen und Hilfsmittel benutzt und sämtliche Textstellen, die wörtlich oder sinngemäß aus veröffentlichen oder unveröffentlichten Schriften entnommen wurden, und alle Angaben, die auf mündlichen Auskünften beruhen, als solche kenntlich gemacht. Ebenfalls sind alle von anderen Personen bereitgestellten Materialien oder erbrachten Dienstleistungen als solche gekennzeichnet.

Leipzig, 05.07.2010

....

(Ying Wang)