Multiscale Analysis of Adaptive Population Dynamics

Dissertation

zur

Erlangung des Doktorgrades (Dr. rer. nat.)

der

Mathematisch-Naturwissenschaftlichen Fakultät

der

Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Shidong Wang

aus

Shandong, CHINA

Bonn, September 2011

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn.

1. Gutachter: Prof. Dr. Anton Bovier

2. Gutachter: Prof. Dr. Patrik Ferrari

Tag der Promotion: 20.12.2011

Erscheinungsjahr: 2011

Abstract

In this thesis we study a spatial population model based on a class of interacting locally regulated branching processes. The results consist of three parts which are independent of each other. The first part, which is the main part of this thesis as presented in Chapter 2 and 3, is concerned with a three-time-scale analysis of a spatially structured population specified with adaptive fitness landscape. More precisely, we obtain a new model, the so-called trait substitution tree (TST), in the limiting system by taking a rare mutation limit against a slow migration limit. These limits can be either simultaneous with a large population limit from a microscopic point of view (Chapter 3), or based on a deterministic approximation (Chapter 2). The TST process is a measure-valued Markov jump process with a well-described branching tree structure. The novelty of our work is that every phenotype, which may nearly die out on the migration time scale, has a chance to recover and further to be stabilized on the mutation time scale because of a change in the fitness landscape due to a new-entering mutant.

The second part (Chapter 4) deals with the neutral mutation case, i.e., the fixation probability of an advantageous mutant is of order $\frac{1}{K^\lambda}$ $(0 < \lambda \leq 1)$ in terms of a large population size K. We proceed by two cases. For $0 < \lambda < 1$ we consider the rescaling limit on a time scale of accumulated mutations and extend the trait substitution sequence model. For $\lambda = 1$ we obtain a rescaling limit in a weak sense, i.e., under conditioning on non-extinction up to observation time.

In the last part (Chapter 5) we study the fluctuation limit of the locally regulated population, and we obtain a limiting process as the solution both of a martingale problem and of a generalized Langevin equation. Under appropriate conditions we prove that the fluctuation limit and the long term limit are interchangeable.

Acknowledgements

First of all I would like to express my sincere gratitude to my advisor Prof. Anton Bovier. Without his consistent support this thesis would never come to the end. I extremely appreciate his patience and tolerance during my long term preparation to find my own track on research. While trustfully leaving me enough freedom to follow my interest, Anton can always kick me out of traps at every critical moment and push me to the right direction. I benefit not only from his deepest knowledge and inspiring intuition but also from his relaxing attitude towards life as a mathematician.

I gratefully acknowledge the hospitality from Prof. Amaury Lambert when I visited University Paris 06 in February 2011. I enjoy to discuss with him in a very pleasant atmosphere. The content of Chapter 4 is partially influenced by him. I acknowledge the hospitality of Prof. Steven Evans during my stay at UC Berkeley in the fall semester of 2011. I want to mention that two times discussion with Prof. Alison Etheridge brought some new ideas in this thesis. I am looking forward to working with her in the near future.

I would not have started my academic career if I didn't get involved in the Probability group at Beijing Normal University. I thank Prof. Zenghu Li to bring me into the mathematical world when I was nowhere.

I fell very lucky to spend three years with colleagues in the Probability group at Bonn, among whom I want to express special thanks to Prof. Sergio Albeverio, Prof. Patrik Ferrari, and Nicola, who encouraged me from time to time. I also thank Adela, Daniel, Evangelia, Giada, Martin, Nikolaus, and René, especially Mei-Ling who let me always feel at home. I often miss the time sharing with my ex-officemate and friend Giacomo.

The financial support of BIGS is acknowledged, which provided me many opportunities to attend workshops and visit other universities. I absolutely deem it as one of the best graduate schools in mathematics.

Finally I want to thank my parents, who haven't seen their son for more than two years, for their self-giving support. And also thank my brother who accompanies the family during my stay abroad.

Contents

1 Introduction

1.1 Biological background

One of the main purposes in the study of various population models (such as population genetics and branching population models) is to understand the complex pattern whereby the distribution of phenotype or genotype in a population changes over time. Moreover, it is of particular interest to understand what is the relative importance of mutation, selection, migration and population subdivision for the genetic diversity observed today. This origins can be traced back to Charles Darwin's pioneer work *On the Origin of Species* [14] published in 1859.

Apart from the discussions it brought about on a non-scientific level, there were also controversial debates where both biologists and mathematicians were involved. One of the points at issue was the claim that the driving force of evolution was *natural selection*, and among selectionists, there was disagreement about the nature of selectively induced evolutionary changes. Darwin adhered to the *gradualist* point of view, that changes in the nature of organisms in populations were gradual and incremental. Other biologists, like T. H. Huxley and F. Galton, were *saltationists*, believing that evolutionary changes usually occur in the form of *jumps* with considerable height. The main difference is that according to Darwin the evolution of well-adapted organisms depends on selection acting on a large number of slight variants of the same trait, whereas much of Mendel's work focuses on discontinuous changes in traits due to a mutation of a single gene. However, as we will see later, these two arguments do not contradict each other since they can be interpreted respectively on different time scales. Depending on the time scale which is chosen to measure the evolutionary changes, one can get different interpretation of the picture.

Up to the rediscovery of *Mendelian genetics* by Fisher, Haldane and Wright et al. around 1900, Darwin's theory of evolution driven by natural selection was finally reconciled with Mendelism, which was of course appealing to the saltationists. Indeed, the former relies crucially on the latter, and further it would be difficult to conceive a Mendelian system without any consideration of natural selection associated with it. To see why this should be the case, it is necessary to turn to some rigorous mathematical justification showing that natural selection can act extremely fast. In a series of papers starting from 1922, Wright introduced the conception of *adaptive landscape*, in which natural selection would drive a population towards a local maximum (in terms of fitness optimization), but a genetic drift could push the population away from such a peak paving the way for natural selection to push it towards another peak. During the last twenty years, due to further developments of mathematical tools, there has raised particular interest in reconsidering the consistency between natural selection and dramatically discontinuous changes in traits. We now give a short overview of the biological literature dealing with this question.

In 1997 Coyne et al. [13] published a perspective that questioned the validity and importance of Wright's *shifting balance theory*, which allows selection to act on alternative *adaptive peaks*, corresponding to different deterministic equilibria. Natural selection pushes populations towards one or another peak, which may then spread by various kinds of group selection. It is consistent with Darwin's theory of evolution that the main engine of adaption is natural selection. An obvious question from a mathematical point of view is whether we can interpret the shifting balance theory on the level of individual-based models? To discuss this issue, we first have to identify the evolutionary mechanisms that contribute to natural selection, such as reproduction, mortality, competition, migration and mutation, and how they fit into the spatial structure constraints of a population.

Back to one decade earlier, Newman et al. [41] investigated an evolutionary phenomenon called *punctuated equilibria* in a well-written biological paper that also caters mathematicians' taste. They emphasize random variation and natural selection as the two central elements of neo-Darwinian evolution. In Wright's view, these lead to random drift of mean population characters in fixed multiply peaked *adaptive landscape* with long periods spent near fitness peaks. They show that the transitions between peaks are rapid and undirectional even though random variation is small and transitions initially require movement against selection. They thus claim that punctuated equilibrium, the palaeontological pattern of rapid transitions between morphological equilibria, is a natural manifestation of standard Wrightian evolutionary theory. In their paper they employ a dynamical system with small random perturbations to illustrate a population with a random genetic variation. However, they mention the idea of different time scales only verbally without giving explicit mathematical justification.

Recently a so-called *trait substitution sequence* (TSS) model, proposed by Metz et al. [39] and justified by Champagnat [8], aims to describe the evolution as a Markov jump process at the population level in the space of phenotypic traits characterizing individuals. In this model, the population is *monomorphic* at any time (i.e., composed of individuals holding the same trait value). The evolution proceeds by a sequence of appearance of new mutant traits, which invade the population and replace, after a selective competition on a shorter time scale, the previous dominant trait. The TSS model belongs to the recent biological theory of evolution called *adaptive dynamics*. The theory of adaptive dynamics investigates the effects of the ecological aspects of population dynamics on the evolutionary process, and thus describes the population on the phenotypic level instead of genotypic level. The TSS model is one of the fundamental models for adaptive dynamics and it turns out to be a powerful tool for understanding various evolutionary phenomena, such as polymorphism (stable coexistence of different traits) and evolutionary branching (evolution from monomorphic population to a polymorphic one that may lead to speciation).

1.2 Motivation

The heuristics leading to the TSS model is based on the biological assumptions of large population and rare mutation, and on another assumption that no two different types of individuals can coexist on a long time scale: the selective competition eliminates one of them. Therefore, coexistence and diversity after entering of new mutants are not allowed due to the deficient spatial structure. On the other hand, natural selection is not only limited to competition mechanism but also is often combined with a survival strategy-migration mechanism. In spite of this heuristics, this model is still lack of a rigorous mathematical basis.

The adaptive-dynamics approach is controversially debated since it was criticized only feasible in the context of phenotypic approach. However, the link with its corresponding genetic picture is rarely developed (see Eshel [17]). As far as the natural selection is concerned, the population-genetics approach has dominated for many years mostly because it is proved powerful to model the sexual reproduction of diploid populations on the genetic level. For a fixed finite gene pool, most evolutionary mechanisms like recombination, selection, and inheritance are theoretically tractable even they can take a role in a very complicated way after recombination gets involved. In particular, the effect of recombination is far more complicated to characterize the long-term evolution since random shuffling of genes may create many genotypes for natural selection (e.g., see [12] for a three-genotype (combined by two genes) case). In contrast, adaptive-dynamics approach is mainly concerned with the long-term evolutionary property but usually ignoring the genetic complications. Is there one way to embody the features such as sexual production and the recombination mechanism arising from the genetic level but at the same time one can still analyze it on the quantitative trait level, i.e., taking advantage of adaptive-dynamics approach on a phenotypic level? This is the biological motivation of the main part of this thesis. In this thesis we propose a new model to justify the above arguments. We introduce a spatial migration mechanism among possible phenotypes, which can be viewed as the results of recombination in a fixed finite gene pool of a sexual population. After natural selection on a short-term evolution time scale, the population can attain an equilibrium configuration according to the fitness landscape. Every time there comes a new mutant gene into the gene pool, the phenotype (trait) space is enlarged due to recombination mechanism, and the spatial migration can be used to characterize the reshuffling procedure on the way to a new equilibrium configuration. Loosely speaking, the spatial movement is used to compensate the simplicity of genetic recombination in adaptive-dynamics approach. The critical point we need to take care is to distinguish these different time scales after introducing fitness spatial structure in the model.

The approach of this model differs from previous models in three key aspects. Firstly, no genetic information is lost on any time scale. Some phenotype containing a specific deleterious gene may be invisible due to its temporary low fitness on the migration time scale, but it can recover based on the reshuffling due to a new mutant gene on the longer mutation time scale. For example, some epidemic virus may become popular again periodically because of a change of its mutated genetic structure or a genetic change of its potential carrier. Secondly, thanks to the fitness spatial structure endowed on finite phenotype space, coexistence is allowed under the assumption of nearest-neighbor competition and migration. This distinguishes our model from the classical adaptive-dynamics system, which usually converges to a monomorphic equilibrium. What is more, we derive a well-defined branching tree structure in the limiting system, which is like a spatial version of the Galton-Watson branching process. Last but not least, the idea to introduce the spatial movement to interpret population genetics can build a bridge between adaptive-dynamics and its genetics counterpart. In particular, we believe that more can be done to quantify our model from the genetic side by mapping migration among phenotypes to recombination among genotypes.

As a reminder, we want to mention a genetic counterpart of the interacting branching population model described in next section, namely an interacting probability measure-valued population studied by Evans et al. [11, 21]. They study a continuous time dynamical system that models the evolving distribution of genotypes in an infinite-many or even a continuum of loci population where recombination acts on a faster time scale than mutation and selection. The intuition behind their asymptotic result is that the mutation preserves the Poisson property whereas selection and recombination respectively drive the population distribution away from and toward Poisson. If all three processes are operating together, we expect that the resulting system will preserve the Poisson property, and in the limit the detailed features of the recombination mechanism disappear. They call it *mutation-selection balance* model.

1.3 Mathematical framework

Most of the current research in biological mathematics is concerned with either phenotypic or genetic models, both of which trace back to the well-known Galton-Watson branching process or the Wright-Fisher model. Since the 1970s it has become very popular for mathematicians to study measure-valued stochastic processes (superprocesses) [15, 18, 37]. On the one hand, this research shows abundant spatial structures beyond the one-dimensional processes. On the other hand, it builds a beautiful bridge between probability theory and infinite-dimensional analysis. There are mainly two classes of superprocesses, *Dawson-Watanabe superprocess* and *Fleming-Viot superprocess*, which are essentially generalizations of resp. the Galton-Watson process and the Wright-Fisher process enriched by various spatial structures. From the biological point of view, they can be interpreted as resp. infinite-many type and infinite-many allele models endowed with spatial motions. We now start with a short survey on a recent framework proposed for spatially structured phenotypic models.

Since individuals can reproduce, mutate and die in varying rates according to their different spatial characteristics (phenotypes), one reasonable improvement one can make is to add spatial components to both branching and dispersal parameters. However, the spatial-dependent components destroy the independent relationship between branching and dispersal while bringing us abundant information from the biological point of view. Even though, the model is still deficient: such as in the finite-dimensional branching process model, a population either dies out or escapes to infinity, depending on the mean matrix of the offspring distribution. The model thus can not predict a non-trivial equilibrium which actually happens quite often in the biological world. Bolker and Pacala [4] propose a locally regulated model which attains the above two improved features. By employing the mechanism of the logistic branching random walks, they introduce a quadratic competition term in the density-dependent populations, which can help us to build some equilibria within subdivided populations. However, the loss of the branching property can also cause some new technical difficulties for the analysis of the model.

Law and Dieckmann [35] study this model in parallel with Bolker and Pacala [4]. We simply call it BPDL model. In recent years, this model has been extensively studied by Etheridge [19], Fournier and Méléard [22], Champagnat [8], Lambert [33], Dawson and Greven [16]. Etheridge [19] studies two diffusion limits, one is a *stepping stone version* of the BPDL model (interacting diffusions indexed by $\mathbb{Z}^d$) and another is a *superprocess version* of it. Also sufficient conditions are given for survival and local extinction. Fournier and Méléard [22] formulate a pathwise construction of the BPDL process in terms of a Poisson point process. Under a finiteness of third moment condition, they rigorously obtain a deterministic approximation (law of large numbers) of the BPDL processes. Our work is based on the formalization of Fournier and Méléard [22]. Champagnat [8], Champagnat and Méléard [10], Dawson and Greven [16] investigate long term behavior by the method of multiple time scale analysis in respective populations.

In this thesis a main goal is to to understand how the mechanisms of reproduction, competition, migration and mutation determine the manner in which the distribution of a spatial population changes over time. In particular, we prescribe relative strengths of these evolutionary mechanisms by constraints on corresponding parameters in order to study their relative importance on different time scales. In the first two chapters we study how the population starting from one single trait eventually colonizes the whole trait space comprised of finite/infinite-many traits as time evolves, meaning that a positive spatial density is reached and a local equilibrium situation arises where locally the process neither becomes extinct nor grows and becomes infinitely large as $t \to \infty$. This is in contrast to the behavior of classical branching models with their survival versus extinction dichotomy in a finite-many trait space and reflects limited resources in a given colony.

More precisely, the scaled population is characterized by a sequence of finite measure-valued processes with the following infinitesimal generator

$$\begin{aligned} L^{K,\epsilon}F(\nu) &= \int_{\mathcal{X}} \left[F(\nu + \frac{\delta_x}{K}) - F(\nu)\right] b(x)K\nu(dx) \\ &+ \int_{\mathcal{X}} \left[F(\nu - \frac{\delta_x}{K}) - F(\nu)\right] \left(d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu(dy)\right) K\nu(dx) \qquad (1.3.1) \\ &+ \epsilon \int_{\mathcal{X}}\int_{\mathcal{X}} \left[F(\nu + \frac{\delta_y}{K} - \frac{\delta_x}{K}) - F(\nu)\right] m(x,dy) 1_{y \in supp\{\nu\}} K\nu(dx), \end{aligned}$$

where we denote by $b(x)$ the birth rate of an individual with trait x, and resp. $d(x)$ its death rate, $\alpha(x,y)$ the competition pressure felt by one with trait x from another with trait y, $m(x,y)$ the migration kernel from x to y. Parameter ϵ is used to govern the migration strength in the population.

In Chapter 2 we firstly attain the large population approximation by letting $K\to\infty$. Then we identify the right time scale $O\left(\ln\frac{1}{\epsilon}\right)$ for fixation and describe the equilibrium configuration on a finite-many trait space $\mathcal{X}$. By adding rare mutation events driven by Poisson point processes, we enlarge the trait space every time there enters a new mutant and push the process to evolve as a well-defined *trait substitution tree* (TST) process, a measure-valued Markov jump process, on the mutation time scale.

In Chapter 3 we aim to justify the TST process by constraining migration and mutation strength to be sufficiently small simultaneously in terms of large population size. For the case on a finite geographic trait space without any mutation, under the condition

$$1 \ll K\epsilon \ll K, \tag{1.3.2}$$

we prove the right time scale for fixation is $O(\ln\frac{1}{\epsilon})$, and provide some specific paths on the way to their equilibria (see Fig. 1.1).

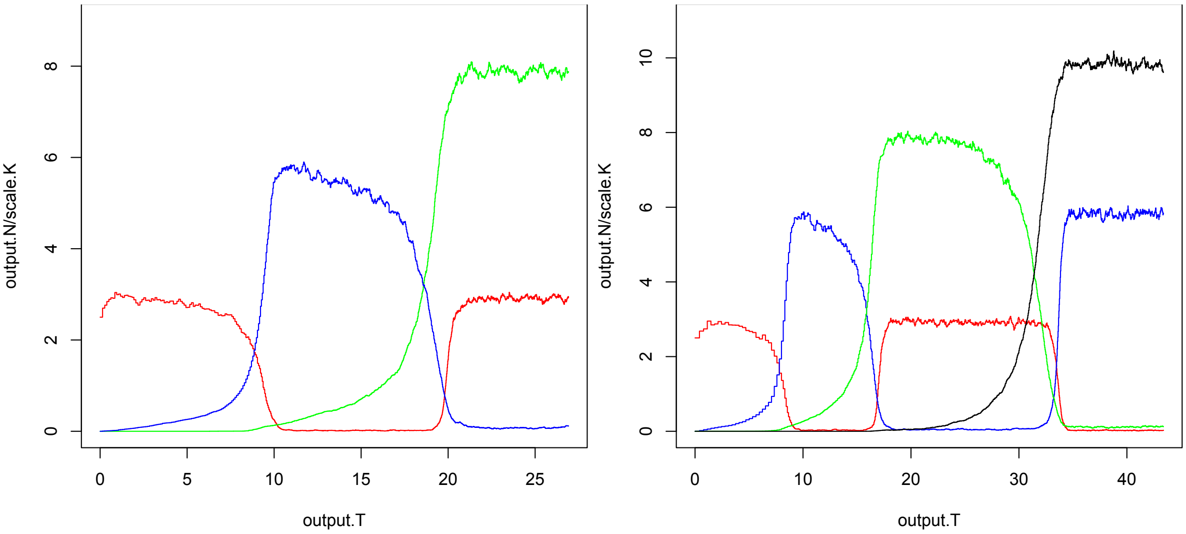

Figure 1.1: Simulations of TST on the migration time scale $O\left(\ln\frac{1}{\epsilon}\right)$

For the population on an infinite-many trait space with mutations, we specify the process with its infinitesimal generator

$$\begin{aligned} L^{K,\epsilon,\sigma}F(\nu) =& \tilde{L}^{K,\epsilon}F(\nu) \\ &+\sigma\int_{\mathcal{X}}\int_{\mathbb{R}^d}\left[F(\nu+\frac{\delta_{x+h}}{K})-F(\nu)\right]\mu(x)p(x,dh)K\nu(dx), \end{aligned} \tag{1.3.3}$$

where the last term represents mutant transition parameterized by σ, and the operator $\tilde{L}^{K,\epsilon}$ is a modification of $L^{K,\epsilon}$ with migration kernel's domain of integration $\mathrm{supp}\{\nu\}$

rather than $\mathcal{X}$. Combining with the constraint (1.3.2) for the migration rate, we impose one more constraint on the mutation rate

$$\ln\frac{1}{\epsilon} \ll \frac{1}{K\sigma} \ll e^{KC}. \tag{1.3.4}$$

By taking the large population limit, we attain the TST process (see Fig. 1.3), which is defined in Chapter 2 on the mutation time scale $O(\frac{1}{K\sigma})$. Note that the effect of migration and competition is indistinguishable in the limiting system, where at any time it takes a value as the equilibrium configuration on a corresponding finite-many trait space.

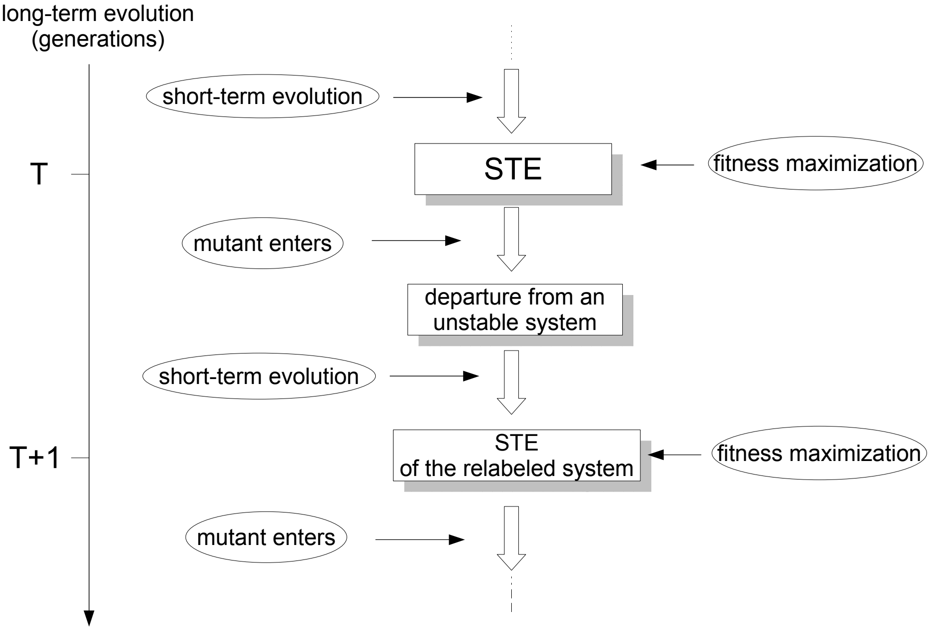

Figure 1.2: Long-term evolution from one STE at generation T to another at generation T+1

In all, we identify that there are three time scales which are related with different evolutionary mechanisms. The shortest one is of course the life cycle time scale of a single individual with no direct concern when studying the evolutionary processes. A middle level time scale is concerned with changes and interactions in a population from one generation to the next due to reproduction, competition and migration. We call it migration time scale. This is usually the topic of classical population dynamics or genetics, and we call it short-term evolution. On the longest one-a mutation time scale, we study the rapid transition from one equilibrium on the short-term scale to another due to the invasion of a new mutant type. More precisely, from one time step to the next on the long-term scale, we switch between three systems: one at the old short-term equilibrium (STE), one during short-term evolution (by one-mutant extension), and one at the new STE. We borrow a nice picture from Schneider [42] to illustrate the idea (see Fig. 1.2).

In Chapter 4 the goal is to consider the case when the fixation probability of the new mutant is nearly neutral (slightly advantageous). This idea can be realized by accelerating the branching rate (birth rate and death rate) in a suitable way, i.e., take individual birth rate $b_K = K^\lambda r + b$ and death rate $d_K = K^\lambda r + d$ for $0 < \lambda \leq 1$, where the parameter K is proportional to the initial population size. This procedure is highly reminiscent of the scaling method from branching particle systems to superprocesses (see [15, 37]). To prevent the population from rapid extinction, one either needs to make the assumption of large population, or to condition the process on not attaining 0 before the time of observation. We will show that, on a single mutation time scale, the fixation probability of the advantageous type is of order $\frac{1}{K^\lambda}$, which converges to 0 under the large population limit. Obviously, it is not the proper time scale to rescale the population process.

For $0 < \lambda < 1$ the large population limit of the population process (branching random walks system) is a deterministic measure-valued process. In particular, it satisfies an ODE which has a stable equilibrium (carrying capacity) for a one-type population without mutation. In order to find the suitable time scale to separate successively arising mutations, our argument proceeds by the way of an intermediate approximation, based on the trait substitute sequence approximation obtained in Champagnat [8]. Eventually, we justify the TSS model on an accumulation time scale of mutations. In other words, the invasion is not due to a single absolutely advantageous mutant, but due to relatively often occurring mutations.

For $\lambda = 1$ the large population limit is a superprocess with a quadratic competition term. The randomness is generated by the accelerating birth and death events and is usually called "demographic stochasticity". In particular, as for the one dimensional case, the rescaled logistic branching process converges to a so called logistic Feller diffusion. As shown in [33], the diffusion will be absorbed at 0 with probability 1. Thus, to capture the long time behavior in some sense, we study the process conditioned on never attaining 0. This conditioned process , named Q-process, is studied in [7, 36]. It can be realized by compensating proper immigration to the system (see [34]). As for the fixation period, we use a classical result about the extinction probability for branching processes in random environments (see Kaplan [29]) to build up the non-coexistence condition, and further to give an implicit expression for the fixation probability of the advantageous type. Nevertheless, we can formulate the limiting processes both for $0 < \lambda < 1$ and $\lambda = 1$ in an unified regime in the context of random environments (see Remark 4.2.3).

In Chapter 5 based on the deterministic approximation of the BPDL processes studied by Fournier and Méléard [22], a fluctuation theorem is proved under a second order moment condition. The limiting process is justified to be an infinite-dimensional Gaussian process solving an inhomogeneous generalized Langevin equation. As an application, we study its properties in the one dimensional case. Finally, under some specific conditions we consider the stationary behavior of the BPDL processes and its fluctuation limit. We prove that the fluctuation limit and the long time limit are interchangeable.

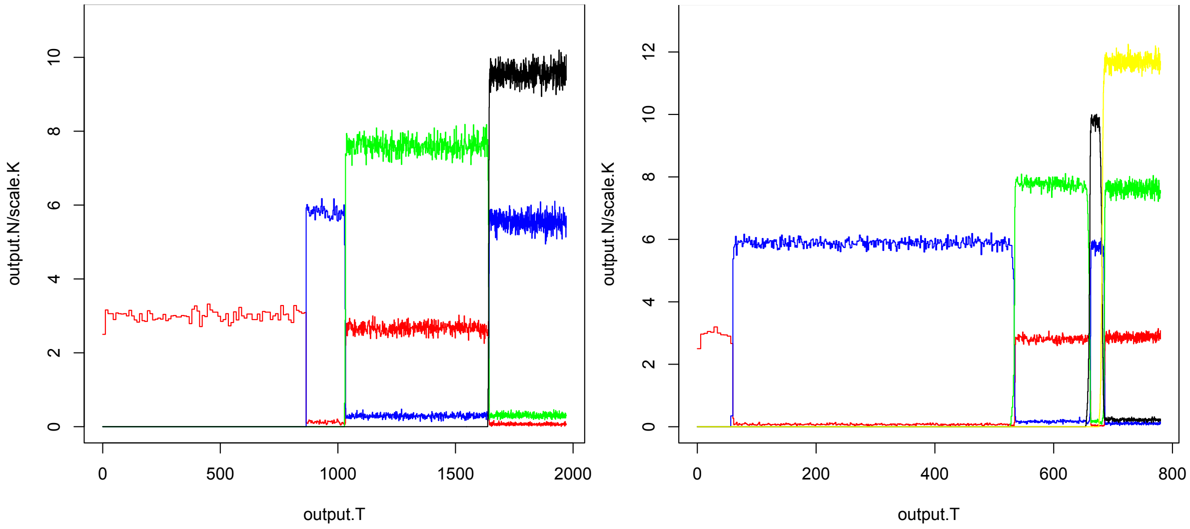

Figure 1.3: Simulations of TST on the mutation time scale $O(\frac{1}{K\sigma})$

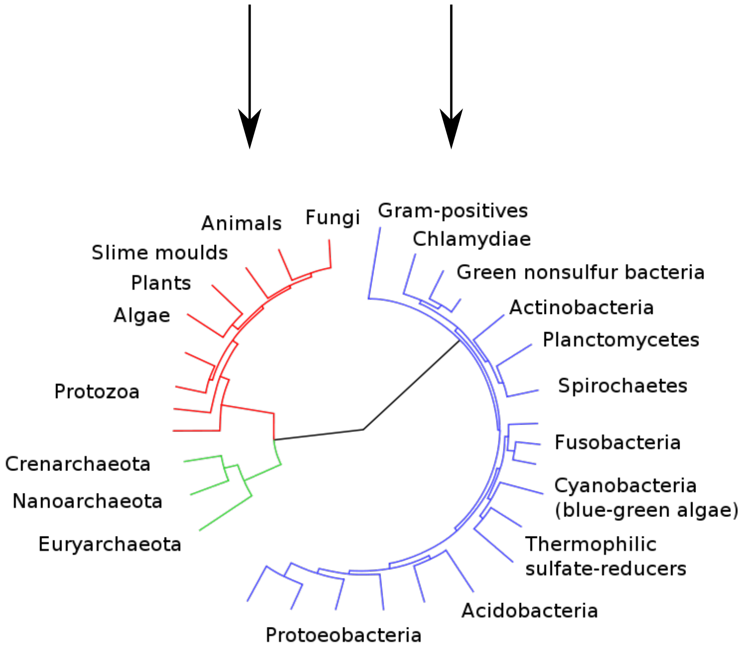

Figure 1.4: Evolution tree of all (from Wikipedia)

2 Trait substitution tree model based on a deterministic system

In this chapter we consider two continuous-mass population models as analogues of spatially structured branching random walks, one is supported on finite-many trait space and the other one is supported on infinite-many trait space. For the first model with nearest-neighbor competition and migration, we justify a well-described evolutionary path to the short-term equilibrium on a slow migration time scale. For the second one with an additional evolutionary mechanism-mutation, a Markov jump process-trait substitution tree model is established under a combination of rare mutation and slow migration limits. The transition rule of the tree highly depends on the relabeled trait sequence determined by the fitness landscape.

2.1 Introduction

In recent years a spatially structured population with migration (namely mutation in [8]) and local regulation proposed by Bolker and Pacala [4], Dieckmann and Law [35] (BPDL process) has attracted particular interest both from biologists and mathematicians. It has several advantages over the traditional branching processes, which make it more natural as population models: the quadratic competition term is used to prevent the population size from escaping to infinity and the migration term is used to transport the population mass from one colony to unoccupied colonies for survival, and further to get colonized. There are mainly two highlights of related papers. Etheridge [19], Fournier and Méléard [22], Hutzenthaler [26] , and Hutzenthaler and Wakolbinger [27] study the extinction and survival problems. Champagnat [8], Champagnat and Lambert [9], Méléard and Tran [38], Dawson and Greven [16] focus more on its long time behavior by multiscale analysis methods.

This paper is motivated partially by the time scales separation procedure used in [8, 9] and partially by the idea of virgin island model introduced in [26]. In [8], a so-called trait substitution sequence model is derived from an individual-based branching particle system under a combination of large population and rare migration (namely mutation there) limits. More precisely, rare migration rate is constrained by the large population parameter to guarantee that the fixation period is not visible on the migration time scale. Under a non-coexistence assumption, the unfit subpopulation can be killed off on the migration time scale since it evolves in the form of discrete-mass population, which is actually a subcritical branching process.

In this paper we will focus on an extreme case where we firstly let population size tend to infinity, and further consider a slow migration rescaling limit based on the continuous-mass population model. Thus, some subpopulations with unfit traits get a chance to recover due to the fact that the competition is only between nearest neighbors. For a finite-many trait space, some specific conditions can be imposed on the fitness and demographic parameters, and a well-described evolutionary path to the short-term equilibrium can be obtained on the slow migration time scale. We call it a trait substitution tree (TST) on the finite trait space. For any given sequence of traits, the equilibrium configuration is determined by their labeled order according to their fitness landscape.

Mutation, a key ingredient to determine the direction of evolution, always takes a significant role in the consideration of population dynamics. In the framework of continuous-mass populations, we want to characterize how the population starting with one single trait type spreads and eventually colonizes over the entire trait space. This leads to speciation and further diversity of population configurations. It can be realized by adding some Poisson-driven mutations to the finite TST. Then the trait space is enlarged every time when there enters a new mutant. Meanwhile, the trait sequence is relabeled according to their fitness values. By constraining a relative strength between mutation rate and migration rate, on the rare mutation time scale one obtains a jump-type TST process as time moves on. We call it a trait substitution tree on the infinite-many trait space.

Notice that here the term “rare” means phenomenon in the sense of stochastic whereas “slow” can be interpreted in the sense of deterministic dynamics.

The remainder of the paper is structured as follows. In Section 2.2, we briefly describe the model and give some preliminary results. In particular, we recall the law of large numbers of the BPDL processes. In Section 2.3, in a slow migration limit, for a finite-many trait space we retrieve a well-defined short-term evolution path to its TST configuration on the migration time scale $O\left(\ln\frac{1}{\epsilon}\right)$. In Section 2.4, under the rare mutation constraint we obtain a jump-type TST process on a longer evolutionary time scale-the mutation time scale. In Section 2.5, we provide proofs of the results in Section 2.3 and Section 2.4. Finally, for better understanding the TST process we provide a simulation algorithm in Section 2.6.

2.2 Microscopic model

2.2.1 Notations and description of the processes

Following [4], we assume the population at time t is composed of a finite number $I(t)$ of individuals characterized by their phenotypic traits $x_1(t), \cdots, x_{I(t)}(t)$ taking values in a compact subset $\mathcal{X}$ of $\mathbb{R}^d$.

We denote by $\mathcal{M}_F(\mathcal{X})$ the set of non-negative finite measures on $\mathcal{X}$. Let $\mathcal{M}_a(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$ be the set of atomic measures on $\mathcal{X}$:

$$\mathcal{M}_a(\mathcal{X}) = \left\{ \sum_{i=1}^{n} \delta_{x_i} : x_1, \cdots, x_n \in \mathcal{X}, n \in \mathbb{N} \right\}.$$

Then the population process can be represented as:

$$\nu_t = \sum_{i=1}^{I(t)} \delta_{X_i(t)}.$$

Let $B(\mathcal{X})$ denote the totality of functions on $\mathcal{X}$ that are bounded measurable. For $\nu \in \mathcal{M}_F(\mathcal{X})$ and $\phi \in \mathcal{M}_F(\mathcal{X})$, denote by $\langle \nu, \phi \rangle = \int \phi d\nu$.

Let's specify the population processes $(\nu_t^n)_{t>0}$ by introducing a sequence of biological parameters, for n$\in \mathbb{N}$:

- $b_n(x)$ is the rate of birth from an individual with trait x.
- $d_n(x)$ is the rate of death of an individual with trait x because of “aging”.
- $\alpha_n(x, y)$ is the competition kernel felt by some individual with trait x from another individual with trait y.
- $D_n(x, dy)$ is the children's dispersion law from its mother with trait x. In particular, it can be decomposed into two parts-local birth at location x and a small portion of migration based on birth, i.e.

  $$D_n(x, dy) = (1-\epsilon)1_{x=y} + \epsilon m_n(x, dy)1_{x \neq y}. \tag{2.2.1}$$

  Here, $m_n(x, dy)$ is the transition density of migration, which satisfies

  $$\int_{y \in \mathcal{X}} m_n(x, dy) = 1.$$

  We will omit the superscript ϵ in D_n in the sequel when this leads no ambiguity.

Fournier and Méléard [22] has formulated a pathwise construction of the BPDL process $\{(\nu_t^n)_{t \geq 0}; n \in \mathbb{N}\}$ in terms of Poisson random measures and justified its infinitesimal generator defined for any $\Phi \in B(\mathcal{M}_a(\mathcal{X}))$:

$$\begin{aligned} L_0^n \Phi(\nu) = &\int_{\mathcal{X}} \nu(dx) \int_{\mathbb{R}^d} [\Phi(\nu + \delta_y) - \Phi(\nu)]\, b_n(x) D_n(x, dy) \\ &+ \int_{\mathcal{X}} \nu(dx) [\Phi(\nu - \delta_x) - \Phi(\mu)] \left[d_n(x) + \int_{\mathcal{X}} \alpha_n(x, y) \nu(dy) \right]. \end{aligned} \tag{2.2.2}$$

The first term is used to model birth events, while the second term which is nonlinear is interpreted as natural death and competing death.

Instead of studying the original BPDL processes defined by (2.2.2), our goal is to study the rescaled processes

$$X_t^n := \frac{\nu_t^n}{n}, \qquad t \geq 0 \tag{2.2.3}$$

since it provides us a macroscopic approximation when we take the large population limits (we will see later, the initial population is proportional to n in some sense). The infinitesimal generator of the rescaled BPDL process has the following form, for any $\Phi \in B(\mathcal{M}_F(\mathcal{X}))$:

$$\begin{aligned} L^n\Phi(\nu) = &\int_{\mathcal{X}} n\nu(dx) \int_{\mathbb{R}^d} \left[\Phi(\nu + \frac{\delta_y}{n}) - \Phi(\nu)\right] b_n(x) D_n(x, dy) \\ &+ \int_{\mathcal{X}} n\nu(dx) \left[\Phi(\nu - \frac{\delta_x}{n}) - \Phi(\nu)\right] \left[d_n(x) + \int_{\mathcal{X}} \alpha_n(x, y) n\nu(dy)\right]. \end{aligned} \tag{2.2.4}$$

2.2.2 Preliminary results

Let's denote by (A) the following assumptions:

(A1) There exist $b(x)$, $d(x)$, $\bar{m}(x) \in B(\mathcal{X})$, $\alpha(x, y) \in B(\mathcal{X} \times \mathcal{X})$ with $\bar{m}(x)$ a probability density for $x, y \in \mathcal{X}$, $n \in \mathbb{N}$, such that

$$0 < b_n(x) \equiv b(x), \qquad 0 < d_n(x) \equiv d(x), \qquad m_n(x, y) \leq \bar{m}(y),$$
$$0 < \alpha_n(x, y) = \frac{\alpha(x, y)}{n}$$

(A2) $b(x) - d(x) > 0$.

The first assumption implies that there exist constants $\bar{b}$, $\bar{d}$, $\bar{\alpha}$ such that $b(x) \leq \bar{b}$, $d(x) \leq \bar{d}$, $\alpha(x, y) \leq \bar{\alpha}$. Furthermore, it guarantees the existence of the BPDL process (see [22]).

By neglecting the high order moment, Bolker and Pacala [4] use the "moment closure" procedure to approximate the stochastic population processes. As we can see from the generator formula (2.2.4), it should be enough to "close" the second order moment due to the quadratic nonlinear term. Actually, we indeed can improve the result of Fournier and Méléard [22] by giving a second moment condition $\sup_{n\geq 1} \mathbb{E}\langle X_0^n, 1\rangle^2 < \infty$ rather than the finiteness of the third moment condition. Since there is no essential difficulty in the improved proof, we only list the result here without repeating the proof.

Theorem 2.2.1 (Fournier and Méléard [22], convergence to an integrodifferenial equation)**.** *Under the assumption (A1), consider a sequence of processes* $(X_t^n)_{t\geq 0}$ *defined in* (2.2.3). *Suppose that* (X_0^n) *converges in law to some deterministic finite measure* $X_0 \in \mathcal{M}_F(\mathcal{X})$ *as* $n \to \infty$ *and satisfies* $\sup_{n\geq 1} \mathbb{E}\langle X_0^n, 1\rangle^2 < \infty$. *Then the sequence of processes* $(X_t^n)_{t\geq 0}$ *converges in law as* $n \to \infty$, *on* $D([0, \infty), \mathcal{M}_F(\mathcal{X}))$, *to a deterministic measure-valued process* $(X_t)_{t\geq 0} \in C([0, \infty), \mathcal{M}_F(\mathcal{X}))$, *where* $(X_t)_{t\geq 0}$ *is an unique solution satisfying*

$$\sup_{t\in[0,T]} \langle X_t, 1\rangle < \infty, \tag{2.2.5}$$

and for any $\phi \in B(\mathcal{X})$,

$$\begin{aligned} \langle X_t, \phi\rangle = &\langle X_0, \phi\rangle + \int_0^t ds \int_{\mathcal{X}} X_s(dx) b(x) \int_{\mathbb{R}^d} \phi(y) D(x, dy) \\ &- \int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi(x) \left[d(x) + \int_{\mathcal{X}} \alpha(x, y) X_s(dy)\right]. \end{aligned} \tag{2.2.6}$$

2.3 TST on finite trait space: without mutation

The trait substitution sequence (in short TSS) model is a powerful tool in understanding various evolutionary phenomena, such as evolutionary branching which may lead to speciation (see Bovier and Champagnat [5]). More precisely, it says that the population follows the "hill climbing" process on the fitness increasing landscape and always keeps monomorphic trait on a long time scale. This model is proposed by Metz [39] and mathematically studied in [8, 9, 38].

Notice that the dispersal kernel $D(x, dz)$ implicitly depends on a parameter ϵ (see (2.2.1)). Rather than taking large population and rare migration limits simultaneously as in [8], we justify a so-called trait substitution tree (in short TST) here from a macroscopic point of view. More precisely, we first consider the large population limit to attain the macroscopic approximation of the individual-based model (see Theorem 2.2.1). Then, we consider the slow migration limit by a rescaling procedure based on the macroscopic limit. In contrast to the model in Champagnat [8], the migration rate here doesn't need to be restricted by the demographic parameter (population size) in the microscopic model.

Here, the so-called TST process arises under the slow migration limit when we assume the nearest-neighbor competition. Note that a variety of short-term evolution paths can be attained by specifying different competition strengths. In other words, the order of invasion and recovery has no special significance even here we restrict the picture by forward invasion into the fitter direction and backward recovery into the unfit direction along the fitness landscape. However, these different paths are indistinguishable on a longer scale-the mutation time scale in next section. Nevertheless, besides the interesting tree structure TST model also brings us some insights into speciation phenomena - evolution from a monomorphic ancestor to diverse species.

Denote by (C) the following assumptions:

(C1) Monomorphic initial trait: $X_0^n = \frac{N_0^n}{n}\delta_{x_0}$, and $\frac{N_0^n}{n} \xrightarrow{\text{law}} \bar{\xi}(x_0)$,

(C2) Nearest-neighbor competition and migration: $\alpha(x_i, x_j) = m(x_i, x_j) \equiv 0$ for $|i-j| > 1$, and non-coexistence condition: $f_{i,i-1} > 0, f_{i-1,i} < 0$ for any $1 \leq i \leq L$, where fitness function $f_{i,j} := b(x_i) - d(x_i) - \alpha(x_i, x_j)\bar{\xi}(x_j)$, and $\bar{\xi}(x_j) := \frac{b(x_j)-d(x_j)}{\alpha(x_j,x_j)}$.

(C3) For any $i \geq 2$,

$$\frac{i}{b(x_i) - d(x_i)} \geq \frac{1}{f_{i,i-1}} + \frac{1}{f_{i-1,i-2}} + \cdots + \frac{1}{f_{1,0}}. \tag{2.3.1}$$

(C4) For any $i \geq 0$, $\frac{|f_{i,i+1}|}{f_{i+2,i+1}} < 1$, and

$$\frac{|f_{i,i+1}|}{f_{i+2,i+1}(b(x_i) - d(x_i))} - \frac{1}{f_{i+3,i+2}} > \frac{|f_{i+1,i+2}|}{f_{i+3,i+2}(b(x_{i+1}) - d(x_{i+1}))} \tag{2.3.2}$$

which implies that the recovery time of trait x_i is later than that of type x_{i+1}.

Notice that (C3-C4) are just technical assumption for results in this section but not necessary for results in next section.

Firstly, we consider the macroscopic limit (2.2.6) which involves parameter ϵ and rewrite it in another form, for any $\phi \in B(\mathcal{X})$,

$$\begin{aligned}\langle X_t^\epsilon, \phi\rangle =& \langle X_0, \phi\rangle + \epsilon \int_0^t ds \int_{\mathcal{X}} X_s^\epsilon(dx) b(x) \int_{\mathcal{X}} [\phi(y) - \phi(x)]\, m(x, dy) \\ &+ \int_0^t ds \int_{\mathcal{X}} X_s^\epsilon(dx)\phi(x) \left[b(x) - d(x) - \int_{\mathcal{X}} \alpha(x,y) X_s^\epsilon(dy)\right].\end{aligned} \tag{2.3.3}$$

Suppose that the current measure is supported on a finite-many trait space

$$\mathcal{X} = \{x_0, x_1, \cdots, x_L\},$$

and it only has nearest-neighbor range competition and migration as in assumption (C2). From the above generator form, it leads to a system which satisfies the following equations

$$\begin{aligned}\xi_t(x_i) =& \xi_0(x_i) + \int_0^t \left[b(x_i) - d(x_i) - \sum_{j=i\pm1,i} \alpha(x_i, x_j)\xi_s(x_i)\right] \xi_s(x_j) ds \\ &+ \epsilon \int_0^t \sum_{j=i\pm1} [\xi_s(x_j) m(x_j, x_i) - \xi_s(x_i) m(x_i, x_j)]\, ds, \qquad 1 \le i \le L.\end{aligned} \tag{2.3.4}$$

The existence and uniqueness of the processes is implied straightforwardly from Theorem 2.2.1. The quadratic regulation term prevents the population from escaping to infinity.

In the following theorem, we build a trait substitution tree model based on the above macroscopic approximation by rescaling on another time scale as migration rate ϵ tends to 0.

Theorem 2.3.1. *Admit assumptions (A) and (C), consider the deterministic measure-valued processes* $(X_t^\epsilon)_{t\ge0}$ *defined by* (2.3.4) *on the trait space* $\mathcal{X} = \{x_0, x_1, x_2, \cdots, x_L\}$, *for any* $L \in \mathbb{N}$. *Then the sequence of rescaled processes* $\left(X^\epsilon_{t\cdot\ln\frac{1}{\epsilon}}\right)_{t\ge0}$ *converges, as* $\epsilon \to 0$, *to* $(U_t)_{t\ge0}$ *which has the following forms depending on the integer* L *is even or odd.*

(i) When $L = 2l$ *for some* $l \in \mathbb{N} \cup 0$,

$$U_t \equiv \begin{cases} \bar{\xi}(x_0)\delta_{x_0} & \text{for } 0 \le t \le I_1, \\ \bar{\xi}(x_k)\delta_{x_k} & \text{for } I_k < t \le I_{k+1},\ k = 1, \cdots, L-1, \\ \bar{\xi}(x_L)\delta_{x_L} & \text{for } I_L < t \le I_L + S_{L-2}, \\ \sum_{i=j}^{l} \bar{\xi}(x_{2i})\delta_{x_{2i}} & \text{for } I_{2j+2} + S_{2j} < t \le I_{2j} + S_{2j-2},\ j = l-1, \cdots, 1, \\ \sum_{i=0}^{l} \bar{\xi}(x_{2i})\delta_{x_{2i}} & \text{for } t > I_2 + S_0. \end{cases} \tag{2.3.5}$$

where $I_k = \sum_{i=1}^{k} \frac{1}{f_{i,i-1}}$, *and* $S_k = \frac{|f_{k,k+1}|}{f_{k+2,k+1}(b(x_k) - d(x_k))}$.

(ii) When $L = 2l + 1$ *for some* $l \in \mathbb{N} \cup 0$,

$$U_t \equiv \begin{cases} \bar{\xi}(x_0)\delta_{x_0} & \textit{for } 0 \le t \le I_1, \\ \bar{\xi}(x_k)\delta_{x_k} & \textit{for } I_k < t \le I_{k+1},\ k = 1, \cdots, L-1, \\ \bar{\xi}(x_L)\delta_{x_L} & \textit{for } I_L < t \le I_L + S_{L-2}, \\ \sum\limits_{i=j}^{l+1} \bar{\xi}(x_{2i-1})\delta_{x_{2i-1}} & \textit{for } I_{2j+1} + S_{2j-1} < t \le I_{2j-1} + S_{2j-3},\ j = l, \cdots 2, \\ \sum\limits_{i=1}^{l+1} \bar{\xi}(x_{2i-1})\delta_{x_{2i-1}} & \textit{for } t > I_3 + S_1. \end{cases} \tag{2.3.6}$$

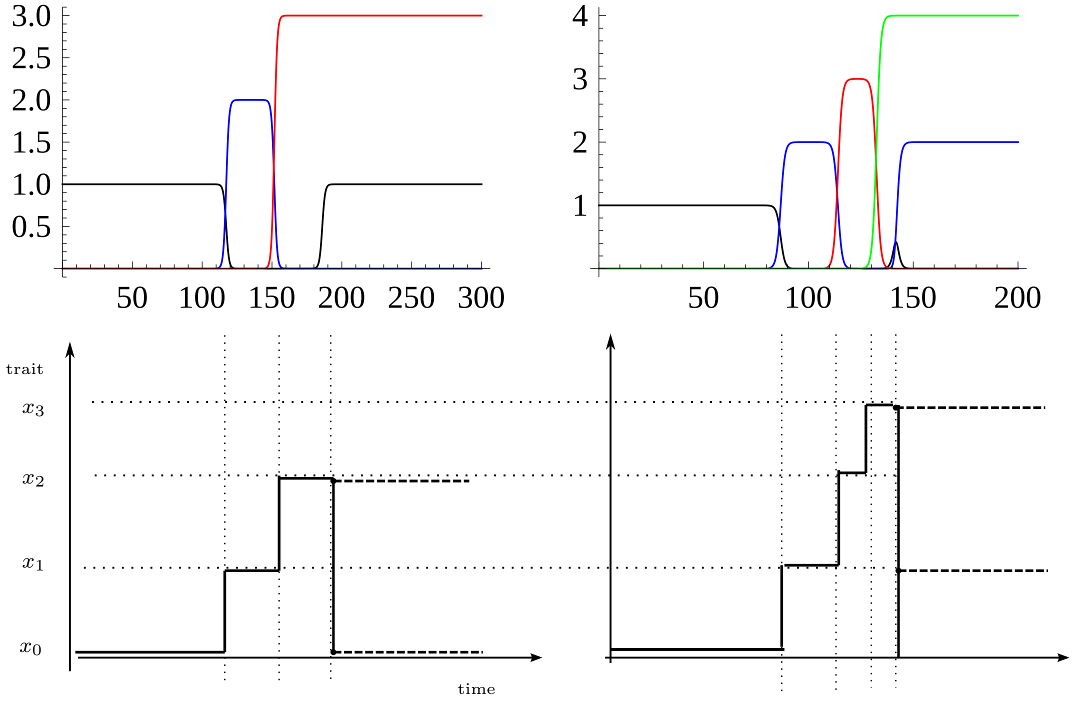

Figure 2.1: Numerical simulations of evolution of a dynamical system with monomorphic initial type and finite trait space (the upper left one has $\mathcal{X} = \{x_0, x_1, x_2\}$ while the upper right one has $\mathcal{X} = \{x_0, x_1, x_2, x_3\}$). Curves describing $\xi_t(x_0), \xi_t(x_1), \xi_t(x_2), \xi_t(x_3)$ are colored black, blue, red, green, resp.. The equilibrium configuration for the first case is $\delta_{x_0} + 3\delta_{x_2}$ and is $2\delta_{x_1} + 4\delta_{x_3}$ for the second one. The lower panel gives their corresponding "trait substitution tree" structure.

Remark 2.3.2. *(1) As time passes on, in the beginning the limiting process* U_t *keeps monomorphic substitution up to the domination of the fittest trait. Afterwards, the relatively unfit traits start to recover along the fitness decreasing direction. From the fittest trait back to the initial one every second one appears in the limit. For instance, when* $\mathcal{X} = \{x_0, x_1, x_2\}$, *the stable configuration has support* $\{x_0, x_2\}$*; when* $\mathcal{X} = \{x_0, x_1, x_2, x_3\}$, *the stable configuration has support* $\{x_1, x_3\}$ *(refer to Figure 2.1).*

This happens because the competition is only between the nearest neighbors and the trait on the RHS is always fitter than the traits on the LHS.

(2) The TST process indexed by $L+2$ can be constructed from the TST process indexed by L by adding a three-type subtree on top of it. For instance, in Figure 2.2, it is showed that the TST ($L = 4$) can be constructed from a smaller TST ($L = 2$) by connecting another excursion consisting of traits $\{x_2, x_3, x_4\}$.

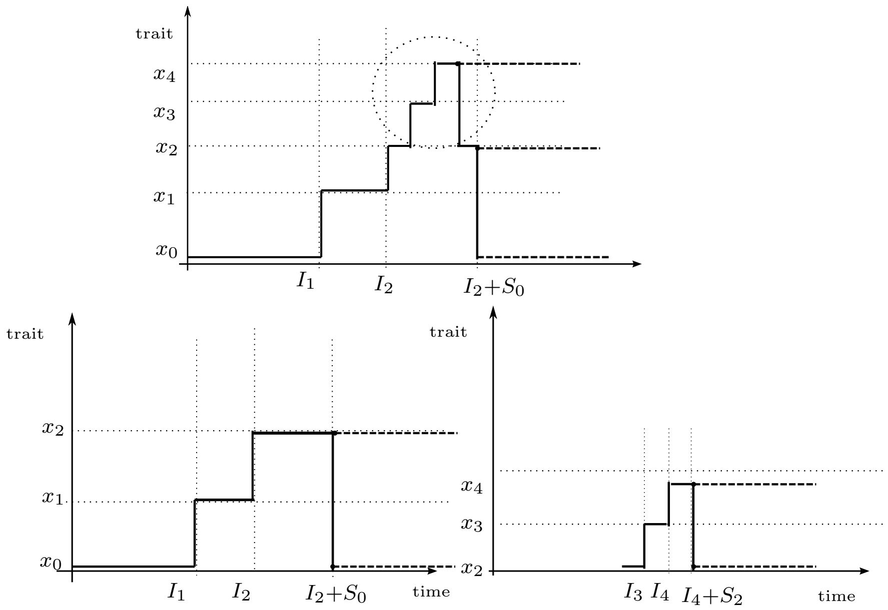

Figure 2.2: Trait substitution tree constructed by embedding excursions.

We postpone the proof of the above result in Section 2.5.1.

2.4 TST on infinite trait space: with mutation

In Section 2.3 we analyze a continuous-mass population on a finite-many trait space defined by equation (2.3.4). On the way to approach its equilibrium configuration, under some specific conditions, a deterministic evolutionary picture is obtained on the slow migration time scale $O\left(\ln \frac{1}{\epsilon}\right)$. An extension to an infinite trait space is natural for our further consideration to complete the whole picture.

In order to enlarge the finite trait space, we introduce another evolutionary mechanism-mutation with a mutant variation density $p(x, dh)$. More precisely, we specify the new model with the following infinitesimal generator, for any $\epsilon, \sigma \geq 0$ and proper testing functions F and ϕ,

$$\begin{aligned} L^{\epsilon,\sigma}F(\nu) = &\int_{\mathcal{X}} \left[b(x) - d(x) - \int_{\mathcal{X}} \alpha(x,y)\nu(dy) \right] \frac{\delta F(\nu;x)}{\delta \nu} \nu(dx) \\ &+ \epsilon \int_{\mathcal{X}} A\Big(\frac{\delta F(\nu;x)}{\delta \nu}\Big)\nu(dx) \\ &+ \sigma \int_{\mathcal{X}} \int_{\mathbb{R}^d} \left[F(\nu+\delta_{x+h}) - F(\nu)\right] \mu(x)p(x,dh)\nu(dx), \end{aligned} \tag{2.4.1}$$

where the differentiation of F is defined by

$$\frac{\delta F(\nu;x)}{\delta \nu} = \lim_{\varepsilon \to 0+} \frac{F(\nu + \varepsilon\delta_x) - F(\nu)}{\varepsilon} \tag{2.4.2}$$

and the operator A coincides with the migration term in (2.3.4)

$$A\phi(x) = \int_{\mathcal{X}} \big[\phi(y) - \phi(x)\big] 1_{\{y \in \mathrm{supp}\{\nu\}\}} m(x,dy). \tag{2.4.3}$$

The first term of the generator describes the local regulation of the population dynamics. The second term describes the migration among support sites. Notice that migration here is not based on birth anymore as in (2.3.4) since it doesn't make any difference if we view on a longer time scale. The last term brings new mutant trait in the form of discontinuous mass to current population dynamics. Non-negative function $\mu(x)$ describes the mutation rate from the resident trait x. Parameters ϵ and σ are used to rescale the strength of migration and mutation of the population. For more discussion on discontinuous superprocesses with a general branching mechanism, one can refer to [37].

We list the following assumptions to guarantee that the limiting process is well-defined.

(D1) For given distinct traits $\{x_0, x_1, \cdots, x_n\} \subset \mathcal{X}, n \in \mathbb{N}$, there exists a total order relation

$$x_{i_0} \prec x_{i_1} \prec \cdots \prec x_{i_{n-1}} \prec x_{i_n}, \tag{2.4.4}$$

where $x \prec y$ means that the fitness functions satisfy $f(x,y) := b(x) - d(x) - \alpha(x,y)\bar{\xi}(y) < 0$, and $f(y,x) := b(y) - d(y) - \alpha(y,x)\bar{\xi}(x) > 0$.

For simplicity of notation, we always assume $x_0^{(n)} \prec x_1^{(n)} \prec \cdots \prec x_n^{(n)}$ with $x_i^{(n)} = x_i$. Every time there enters a new trait x whose fitness is between $x_j^{(n)}$ and $x_{j+1}^{(n)}$ for some $0 \le j \le n$, we relabel new traits by order as following

$$x_0^{(n+1)} \prec x_1^{(n+1)} \prec \cdots \prec x_n^{(n+1)} \prec x_{n+1}^{(n+1)}, \tag{2.4.5}$$

where $x_i^{(n+1)} = x_i^{(n)}$ for $0 \le i \le j$, $x_{j+1}^{(n+1)} = x$ and $x_i^{(n+1)} = x_{i-1}^{(n)}$ for $j+2 \le i \le n+1$.

(D2) Competition and migration only occurs between nearest neighbors, i.e., for totally ordered traits in (D1), we have $m(x_i^{(n)}, x_j^{(n)}) = \alpha(x_i^{(n)}, x_j^{(n)}) \equiv 0$ for $\mid i-j \mid > 1$.

Notice that assumption (C3-C4) is not necessary for the following results, where we will give the rescaling limit on a longer time scale than $O\left(\ln\frac{1}{\epsilon}\right)$. On the migration time scale, there are a variety of different paths to approach the equilibrium configuration by giving different parameters. Under the restrictive condition (C3-C4), it can provide us one specific clear-described evolutionary path on the migration time scale. However, the equilibrium configuration of system (2.3.4) is always the same up to the ordered sequence determined as in assumption (D1) and the convergence time scale is always of order $O(\ln\frac{1}{\epsilon})$.

Definition 2.4.1. *A Markov jump process $\{\Gamma_t : t \geq 0\}$ characterized as following is called a trait substitution tree with the ancestor $\Gamma_0 = \bar{\xi}(x_0)\delta_{x_0}$.*

(i) For any nonnegative integer l, it jumps from $\Gamma^{(2l)} := \sum_{i=0}^{l} \bar{\xi}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$ to $\Gamma^{(2l+1)}$ with transition rate $\bar{\xi}(x_{2k}^{(2l)})\mu(x_{2k}^{(2l)})p(x_{2k}^{(2l)}, dh)$ for any $0 \leq k \leq l$, where

- $\Gamma^{(2l+1)} = \sum_{i=1}^{j} \bar{\xi}(x_{2i-1}^{(2l)})\delta_{x_{2i-1}^{(2l)}} + \bar{\xi}(x_{2k}^{(2l)} + h)\delta_{x_{2k}^{(2l)}+h} + \sum_{i=j+1}^{l} \bar{\xi}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$
  if there exists $0 \leq j \leq l$ s.t. $x_{2j}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j+1}^{(2l)}$,

- $\Gamma^{(2l+1)} = \sum_{i=1}^{j} \bar{\xi}(x_{2i-1}^{(2l)})\delta_{x_{2i-1}^{(2l)}} + \sum_{i=j}^{l} \bar{\xi}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$
  if there exists $0 \leq j \leq l$ s.t. $x_{2j-1}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j}^{(2l)}$.

Then, we relabel the trait squence according to the total order relation as in (D1):

$$x_0^{(2l+1)} \prec x_1^{(2l+1)} \prec \cdots \prec x_{2l}^{(2l+1)} \prec x_{2l+1}^{(2l+1)}, \tag{2.4.6}$$

where in associate with the first case

$$x_i^{(2l+1)} := x_i^{(2l)} \text{ for } 0 \leq i \leq 2j, \qquad x_{2j+1}^{(2l+1)} := x_{2k}^{(2l)} + h,$$
$$x_i^{(2l+1)} := x_{i-1}^{(2l)} \text{ for } 2j+2 \leq i \leq 2l+1,$$

and in associate with the second case

$$x_i^{(2l+1)} := x_i^{(2l)} \text{ for } 0 \leq i \leq 2j-1, \quad x_{2j}^{(2l+1)} := x_{2k}^{(2l)} + h,$$
$$x_i^{(2l+1)} := x_{i-1}^{(2l)} \text{ for } 2j+1 \leq i \leq 2l+1.$$

(ii) For nonnegative integer l, it jumps from $\Gamma^{(2l+1)} := \sum_{i=1}^{l+1} \bar{\xi}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$ to $\Gamma^{(2l+2)}$ with transition rate $\bar{\xi}(x_{2k-1}^{(2l+1)})\mu(x_{2k-1}^{(2l+1)})p(x_{2k-1}^{(2l+1)}, dh)$ for any $1 \leq k \leq l+1$, where

- $\Gamma^{(2l+2)} = \sum_{i=1}^{j} \bar{\xi}(x_{2(i-1)}^{(2l+1)})\delta_{x_{2(i-1)}^{(2l+1)}} + \bar{\xi}(x_{2k-1}^{(2l+1)} + h)\delta_{x_{2k-1}^{(2l+1)}+h} + \sum_{i=j+1}^{l+1} \bar{\xi}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$
  if there exists $1 \leq j \leq l+1$ s.t. $x_{2j-1}^{(2l+1)} \prec x_{2k-1}^{(2l+1)} + h \prec x_{2j}^{(2l+1)}$,

- $\Gamma^{(2l+1)} = \sum_{i=1}^{j} \bar{\xi}(x^{(2l+1)}_{2(i-1)})\delta_{x^{(2l+1)}_{2(i-1)}} + \sum_{i=j}^{l+1} \bar{\xi}(x^{(2l+1)}_{2i-1})\delta_{x^{(2l+1)}_{2i-1}}$
  if there exists $1 \leq j \leq l+1$ *s.t.* $x^{(2l+1)}_{2j-2} \prec x^{(2l+1)}_{2k-1} + h \prec x^{(2l+1)}_{2j-1}$.

Then, we relabel the trait sequence according to the total order relation as in (D1):

$$x_0^{(2l+2)} \prec x_1^{(2l+2)} \prec \cdots \prec x_{2l+1}^{(2l+2)} \prec x_{2l+2}^{(2l+2)}, \tag{2.4.7}$$

where in associate with the first case

$$\begin{aligned} &x_i^{(2l+2)} := x_i^{(2l+1)} \text{ for } 0 \leq i \leq 2j-1, \quad x_{2j}^{(2l+2)} := x_{2k-1}^{(2l+1)} + h, \\ &x_i^{(2l+2)} := x_{i-1}^{(2l+1)} \text{ for } 2j+1 \leq i \leq 2l+2, \end{aligned}$$

and in associate with the second case

$$\begin{aligned} &x_i^{(2l+2)} := x_i^{(2l+1)} \text{ for } 0 \leq i \leq 2j-2, \quad x_{2j-1}^{(2l+2)} := x_{2k-1}^{(2l+1)} + h, \\ &x_i^{(2l+2)} := x_{i-1}^{(2l+1)} \text{ for } 2j \leq i \leq 2l+2. \end{aligned}$$

Theorem 2.4.2. *Admit assumption (A) and (D), consider processes* $\{X_t^{\epsilon,\sigma}, t \geq 0\}$ *described by the generator* (2.4.1). *Suppose that* $X_0^{\epsilon,\sigma} = \xi^\epsilon(x_0)\delta_{x_0}$ *and* $\xi^\epsilon(x_0) \to \bar{\xi}(x_0)$ *in law, as* $\epsilon \to 0$. *If it holds that*

$$\frac{1}{\sigma} \gg \ln \frac{1}{\epsilon}, \tag{2.4.8}$$

$(X^{\epsilon,\sigma}_{\frac{t}{\sigma}})_{t\geq 0}$ *converges to the trait substitution tree* $(\Gamma_t)_{t\geq 0}$ *defined in Definition 2.4.1 in the sense of f.d.d. as* ϵ *tends to 0.*

We postpone the proof of the above result in Section 2.5.2.

2.5 Outline of proofs

2.5.1 Proof of Theorem 2.3.1

In this section, we provide rigorous proofs of results in Section 2.3. The main idea behind the proof is that the migration spreads linearly and the nearest neighbor competitive growth is in exponential speed. Before proving Theorem 2.3.1, let's give some preliminary results which are key ingredients for the proof of Theorem 2.3.1.

The following lemma provides the non-coexistence condition for a dimorphic Lotka-Volterra system. We will include the proof in Appendix A.

Lemma 2.5.1. *Consider a dimorphic system*

$$\begin{cases} \dot{\xi}_t(x_i) = \Big(b(x_i) - d(x_i) - \alpha(x_i, x_i)\xi_t(x_i) - \alpha(x_i, x_{i+1})\xi_t(x_{i+1})\Big)\xi_t(x_i) \\ \dot{\xi}_t(x_{i+1}) = \Big(b(x_{i+1}) - d(x_{i+1}) - \alpha(x_{i+1}, x_i)\xi_t(x_i) - \alpha(x_{i+1}, x_{i+1})\xi_t(x_{i+1})\Big)\xi_t(x_{i+1}) \end{cases} \tag{2.5.1}$$

with some positive initial value. If $f_{i,i+1} < 0$, $f_{i+1,i} > 0$, *then* $\big(0, \bar{\xi}(x_{i+1})\big)$ *is the only stable equilibrium.*

The following two propositions will be used as basic ingredients to prove Theorem 2.3.1.

Proposition 2.5.2. *Admit the same conditions as in Theorem 2.3.1. Consider the case when* $L = 2$*, i.e.* $\mathcal{X} = \{x_0, x_1, x_2\}$*. Then, the limit process* $(U_t)_{t\geq 0}$ *has the form*

$$U_t \equiv \begin{cases} \bar{\xi}(x_0)\delta_{x_0} & \textit{for } 0 \leq t \leq I_1, \\ \bar{\xi}(x_1)\delta_{x_1} & \textit{for } I_1 < t \leq I_2, \\ \bar{\xi}(x_2)\delta_{x_2} & \textit{for } I_2 < t \leq I_2 + S_0, \\ \bar{\xi}(x_0)\delta_{x_0} + \bar{\xi}(x_2)\delta_{x_2} & \textit{for } t > I_2 + S_0. \end{cases} \tag{2.5.2}$$

Proof. **(a)** Firstly, suppose that the population comprises of only two types, i.e. $\mathcal{X} = \{x_0, x_1\}$. We proceed our proof by dividing the entire invasion period into four steps as in Figure 2.3.

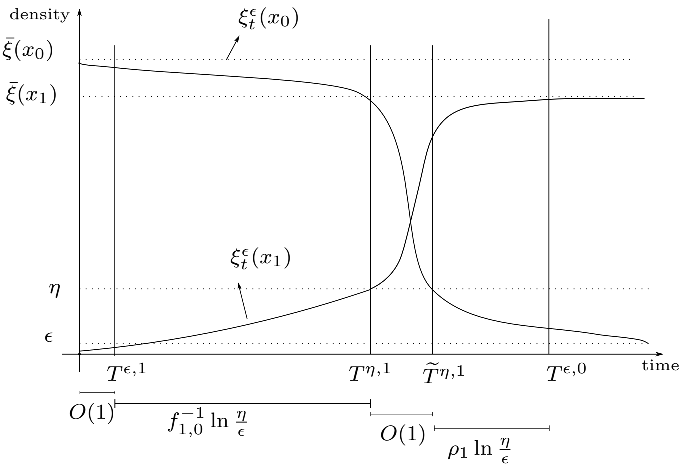

Figure 2.3: Four-step invasion analysis for a dimorphic system

.

Let $\xi_t^\epsilon(x_0) := \langle X_t^\epsilon, 1_{\{x_0\}}\rangle$ and $\xi_t^\epsilon(x_1) := \langle X_t^\epsilon, 1_{\{x_1\}}\rangle$, from (2.3.4), one obtains

$$\begin{aligned} \dot{\xi}_t^\epsilon(x_0) =& \big(b(x_0) - d(x_0) - \alpha(x_0, x_0)\xi_t^\epsilon(x_0) - \alpha(x_0, x_1)\xi_t^\epsilon(x_1)\big)\xi_t^\epsilon(x_0) \\ & - \epsilon\xi_t^\epsilon(x_0)b(x_0)m(x_0, x_1) + \epsilon\xi_t^\epsilon(x_1)b(x_1)m(x_1, x_0), \end{aligned} \tag{2.5.3}$$

and

$$\begin{aligned} \dot{\xi}_t^\epsilon(x_1) =& \big(b(x_1) - d(x_1) - \alpha(x_1, x_0)\xi_t^\epsilon(x_0) - \alpha(x_1, x_1)\xi_t^\epsilon(x_1)\big)\xi_t^\epsilon(x_1) \\ & - \epsilon\xi_t^\epsilon(x_1)b(x_1)m(x_1, x_0) + \epsilon\xi_t^\epsilon(x_0)b(x_0)m(x_0, x_1), \end{aligned} \tag{2.5.4}$$

where $\xi_0^\epsilon(x_0) = \bar{\xi}(x_0)$ and $\xi_0^\epsilon(x_1) = 0$.

Step 1. For any fixed $\eta > 0$, $\forall\, 0 < \epsilon < \eta$, let $T^{\epsilon,1}$ be the time when dynamics $\big(\xi_t^\epsilon(x_0), \xi_t^\epsilon(x_1)\big)$ leaves the $\epsilon-$neighborhood of $(\bar{\xi}(x_0), 0)$, i.e. $T^{\epsilon,1} = \inf\big\{t \geq 0 : \xi_t^\epsilon(x_1) >$

ϵ, $\xi^\epsilon_t(x_0) < \bar{\xi}(x_0) - \epsilon\big\}$. From (2.5.4), for $t < T^{\epsilon,1}$, the dynamics $\xi^\epsilon_t(x_1)$ satisfies the following differential inequality

$$\begin{aligned}\dot{\xi}^\epsilon_t(x_1) \geq &\big(b(x_1) - d(x_1) - \alpha(x_1,x_0)\bar{\xi}(x_0) - \epsilon\alpha(x_1,x_1) - \epsilon b(x_1)m(x_1,x_0)\big)\xi^\epsilon_t(x_1)\\ &+ \epsilon\big(\bar{\xi}(x_0) - \epsilon\big)b(x_0)m(x_0,x_1)\\ =&\big(f_{1,0} - \epsilon(\alpha(x_1,x_1) + b(x_1)m(x_1,x_0))\big)\xi^\epsilon_t(x_1) + \epsilon\big(\bar{\xi}(x_0) - \epsilon\big)b(x_0)m(x_0,x_1).\end{aligned} \tag{2.5.5}$$

Since $f_{1,0} = b(x_1) - d(x_1) - \alpha(x_1,x_0)\bar{\xi}(x_0) > 0$, we can choose ϵ sufficiently small such that the first term on the right hand side of above inequality is positive. By omitting the positive term, one obtains $\xi^\epsilon_t(x_1)$ can be controlled at least by $\check{\xi}_t(x_1)$, i.e. $\check{\xi}_t(x_1) \leq \xi^\epsilon_t(x_1)$, where $\check{\xi}_0(x_1) = 0$, and

$$\dot{\check{\xi}}_t(x_1) = \epsilon\big(\bar{\xi}(x_0) - \epsilon\big)b(x_0)m(x_0,x_1). \tag{2.5.6}$$

Thus, $T^{\epsilon,1}$ can be bounded from above by $\check{T}^{\epsilon,1} = \big((\bar{\xi}(x_0) - \epsilon)b(x_0)m(x_0,x_1)\big)^{-1}$, which is the time length for $\check{\xi}_t(x_1)$ reaching $\epsilon-$level. So, $T^{\epsilon,1}$ is of order $O(1)$.

Step 2. Since time $T^{\epsilon,1}$, we consider the evolution of the population $\big(\xi^\epsilon_t(x_0), \xi^\epsilon_t(x_1)\big)$ until the time (mark by $T^{\eta,1}$) when it leaves $\eta-$neighborhood of $(\bar{\xi}(x_0), 0)$. From (2.5.4), by omitting the term $\epsilon\xi^\epsilon_t(x_0)b(x_0)m(x_0,x_1)$, we get

$$\begin{aligned}&\dot{\xi}^\epsilon_t(x_1)\\ &\geq \big(b(x_1) - d(x_1) - \alpha(x_1,x_0)\xi^\epsilon_t(x_0) - \alpha(x_1,x_1)\xi^\epsilon_t(x_1)\big)\xi^\epsilon_t(x_1) - \epsilon\xi^\epsilon_t(x_1)b(x_1)m(x_1,x_0)\\ &\geq \big(b(x_1) - d(x_1) - \alpha(x_1,x_0)\bar{\xi}(x_0) - \eta\alpha(x_1,x_1)\big)\xi^\epsilon_t(x_1) - \eta\xi^\epsilon_t(x_1)b(x_1)m(x_1,x_0)\\ &= \big(f_{1,0} - \eta\check{C}\big)\xi^\epsilon_t(x_1),\end{aligned} \tag{2.5.7}$$

where $\check{C} = \alpha(x_1,x_1) + b(x_1)m(x_1,x_0)$. On the other hand, by omitting some negative terms in (2.5.4), we get

$$\begin{aligned}\dot{\xi}^\epsilon_t(x_1) &\leq \big(b(x_1) - d(x_1) - \alpha(x_1,x_0)\xi^\epsilon_t(x_0)\big)\xi^\epsilon_t(x_1) + \epsilon\xi^\epsilon_t(x_0)b(x_0)m(x_0,x_1)\\ &\leq \big(b(x_1) - d(x_1) - \alpha(x_1,x_0)(\bar{\xi}(x_0) - \eta)\big)\xi^\epsilon_t(x_1) + \epsilon\bar{\xi}(x_0)b(x_0)m(x_0,x_1)\\ &\leq (f_{1,0} + \eta\hat{C})\xi^\epsilon_t(x_1),\end{aligned} \tag{2.5.8}$$

where $\hat{C} = \alpha(x_1,x_0) + \bar{\xi}(x_0)b(x_0)m(x_0,x_1)$.

By applying Gronwall's inequality to (2.5.7) and (2.5.8), the flow $\xi^\epsilon_t(x_1)$, starting with $\xi^\epsilon_{T^{\epsilon,1}}(x_1) = \epsilon$, can be bounded from below by $\check{\xi}_t(x_1)$ and from above by $\hat{\xi}_t(x_1)$, which satisfy the following equations

$$\dot{\check{\xi}}_t(x_1) = (f_{1,0} - \check{C}\eta)\check{\xi}_t(x_1), \tag{2.5.9}$$

and

$$\dot{\hat{\xi}}_t(x_1) = (f_{1,0} + \hat{C}\eta)\hat{\xi}_t(x_1), \tag{2.5.10}$$

with $\check{\xi}_{T^{\epsilon,1}}(x_1) = \hat{\xi}_{T^{\epsilon,1}}(x_1) = \epsilon$.

The time length needed for $\check{\xi}_t(x_1)$ and $\hat{\xi}_t(x_1)$ to reach $\eta-$level can be solved explicitly and has the form $\check{T}^{\eta,1} - T^{\epsilon,1} = (f_{1,0} - \check{C}\eta)^{-1} \ln \frac{\eta}{\epsilon}$ and $\hat{T}^{\eta,1} - T^{\epsilon,1} = (f_{1,0} + \hat{C}\eta)^{-1} \ln \frac{\eta}{\epsilon}$ respectively. Because $\hat{T}^{\eta,1} < T^{\eta,1} < \check{T}^{\eta,1}$ for any $\eta > 0$, $T^{\eta,1} - T^{\epsilon,1}$ is of order $f_{1,0}^{-1} \ln \frac{1}{\epsilon}$.

Step 3. After time $T^{\eta,1}$, dynamic flow $\big(\xi_t^\epsilon(x_0), \xi_t^\epsilon(x_1)\big)$ comprising of populations x_0 and x_1 can be approximated by the following equations as ϵ tends to 0:

$$\begin{cases} \dot{\xi}_t(x_0) = \big(b(x_0) - d(x_0) - \alpha(x_1, x_0)\xi_t(x_0) - \alpha(x_1, x_1)\xi_t(x_1)\big)\xi_t(x_0) \\ \dot{\xi}_t(x_1) = \big(b(x_1) - d(x_1) - \alpha(x_1, x_0)\xi_t(x_0) - \alpha(x_1, x_1)\xi_t(x_1)\big)\xi_t(x_1), \end{cases} \tag{2.5.11}$$

which has a stable equilibrium $(0, \bar{\xi}(x_1))$ under assumption $f_{1,0} > 0,\ f_{0,1} < 0$ (see Lemma 2.5.1). Let $\widetilde{T}^{\eta,1}$ be the time until when $(\xi_t^\epsilon(x_0), \xi_t^\epsilon(x_1))$ reaches η-neighborhood of the equilibrium $(0, \bar{\xi}(x_1))$, i.e. $\xi_{\widetilde{T}^{\eta,1}}(x_0) = \eta$. Thus, $\widetilde{T}^{\eta,1} - T^{\eta,1}$ is of order $O(1)$.

Step 4. Since time $\widetilde{T}^{\eta,1}$, we consider the time length needed for x_1 getting fixed (i.e. x_0 gets absorbed by 0). From (2.5.3), one obtains the lower bound differential inequality

$$\begin{aligned} &\dot{\xi}_t^\epsilon(x_0) \\ &\geq \big(b(x_0) - d(x_0) - \alpha(x_0, x_0)\xi_t^\epsilon(x_0) - \alpha(x_0, x_1)\xi_t^\epsilon(x_1)\big)\xi_t^\epsilon(x_0) - \epsilon\xi_t^\epsilon(x_0)b(x_0)m(x_0, x_1) \\ &\geq \big(b(x_0) - d(x_0) - \eta\alpha(x_0, x_0) - \alpha(x_0, x_1)\bar{\xi}(x_1)\big)\xi_t^\epsilon(x_0) - \eta b(x_0)m(x_0, x_1)\xi_t^\epsilon(x_0) \\ &= \big(f_{0,1} - \eta\check{C}\big)\xi_t^\epsilon(x_0), \end{aligned} \tag{2.5.12}$$

where $\check{C} = \alpha(x_0, x_0) + b(x_0)m(x_0, x_1)$. As for the upper bound, we observe that

$$\begin{aligned} \dot{\xi}_t^\epsilon(x_0) &\leq \big(b(x_0) - d(x_0) - \alpha(x_0, x_1)\xi_t^\epsilon(x_1)\big)\xi_t^\epsilon(x_0) + \epsilon\xi_t^\epsilon(x_1)b(x_1)m(x_1, x_0) \\ &\leq \big(b(x_0) - d(x_0) - \alpha(x_0, x_1)(\bar{\xi}(x_1) - \eta)\big)\xi_t^\epsilon(x_0) + \epsilon\bar{\xi}(x_1)b(x_1)m(x_1, x_0) \\ &\leq (f_{1,0} + \eta\hat{C})\xi_t^\epsilon(x_1), \end{aligned} \tag{2.5.13}$$

where $\hat{C} = \alpha(x_0, x_1) + \bar{\xi}(x_1)b(x_1)m(x_1, x_0)$.

By applying Gronwall's inequality to (2.5.12) and (2.5.13), we get that $\xi_t^\epsilon(x_0)$, starting with $\xi^\epsilon_{\widetilde{T}^{\eta,1}}(x_0) = \eta$, can be bounded from below by $\check{\xi}_t(x_0)$ and from above by $\hat{\xi}_t(x_0)$, which satisfy the following equations

$$\dot{\check{\xi}}_t(x_0) = (f_{0,1} - \check{C}\eta)\check{\xi}_t(x_0), \tag{2.5.14}$$

and

$$\dot{\hat{\xi}}_t(x_0) = (f_{0,1} + \hat{C}\eta)\hat{\xi}_t(x_0), \tag{2.5.15}$$

with $\check{\xi}_{\widetilde{T}^{\eta,1}}(x_0) = \hat{\xi}_{\widetilde{T}^{\eta,1}}(x_0) = \eta$.

Since $f_{0,1} = b(x_0) - d(x_0) - \alpha(x_0, x_1)\bar{\xi}(x_1) < 0$, we can choose η small enough such that $f_{0,1} + \hat{C}\eta < 0$. Therefore, both $\check{\xi}_t(x_0)$ and $\hat{\xi}_t(x_0)$ decay exponentially. For any $\rho_1 > 0$, the process $\hat{\xi}_t(x_0)$, in time length of order $\rho_1 \ln \frac{\eta}{\epsilon}$, gets into the $\epsilon^{-\rho_1(f_{0,1} + \eta\hat{C})}-$neighborhood of 0 while $\check{\xi}_t(x_0)$ gets into the $\epsilon^{-\rho_1(f_{0,1} - \eta\check{C})}-$neighborhood of 0. Let $T^{\epsilon,0} := \widetilde{T}^{\eta,1} + \rho_1 \ln \frac{\eta}{\epsilon}$. Hence, $\lim\limits_{\epsilon \to 0} \xi^\epsilon_{T^{\epsilon,0}}(x_0) = \lim\limits_{\epsilon \to 0} \epsilon^{-\rho_1 f_{0,1}} \cdot O(\eta) = 0$, then type x_1 eventually gets fixed.

In all, by combining the above four-step analysis, one concludes that the right time scale for the more fit population x_1 getting fixed is

$$(f_{1,0}^{-1} + \rho_1) \ln \frac{1}{\epsilon}. \tag{2.5.16}$$

(b) (See Figure 2.4) Furthermore, we consider the case when there are three phenotypes $\mathcal{X} = \{x_0, x_1, x_2\}$ supporting the system (2.3.4). At the same time as the mass on trait site x_0 migrates towards new trait site x_1 as showed in (1), the mass on trait site x_1 can also migrate to another site x_2. Let $\xi_t^\epsilon(x_2) := \langle X_t^\epsilon, 1_{\{x_2\}} \rangle$. In the following, we reanalyze the evolution process by adding one more trait x_2 into the original trait space $\{x_0, x_1\}$. There will be two kinds of resource to contribute the growth of population on x_2. One of them is indirect migration from trait site x_0. More precisely, $\xi_0^\epsilon(x_2) = \epsilon \xi_0^\epsilon(x_1) = \epsilon^2 \bar{\xi}(x_0)$. Thus, starting with mass of order ϵ^2, the time length needed for $\xi_t^\epsilon(x_2)$ to reach some η-level is of order $\frac{2}{b(x_2)-d(x_2)} \ln \frac{1}{\epsilon}$.

As observed in Figure 2.4, because of assumption (C3): $\frac{2}{b(x_2)-d(x_2)} > \frac{1}{f_{1,0}}$, population x_2 is still negligible before time $T^{\eta,1}$ when dynamics x_1 reaches the $\eta-$level. Since $\widetilde{T}^{\eta,1} - T^{\eta,1} = O(1)$, population x_2, starting with $\xi^\epsilon_{\widetilde{T}^{\eta,1}}(x_2) = \epsilon \cdot O(1)$, evolves under the competition from its resident population x_1 as follows

$$\begin{aligned} \dot{\xi}_t^\epsilon(x_2) =& \big(b(x_2) - d(x_2) - \alpha(x_2, x_1)\xi_t^\epsilon(x_1) - \alpha(x_2, x_2)\xi_t^\epsilon(x_2)\big)\xi_t^\epsilon(x_2) \\ & - \epsilon \xi_t^\epsilon(x_2) b(x_2) m(x_2, x_1) + \epsilon \xi_t^\epsilon(x_1) b(x_1) m(x_1, x_2), \end{aligned} \tag{2.5.17}$$

where $\xi^\epsilon_{\widetilde{T}^{\eta,1}}(x_1) \in (\bar{\xi}(x_1) - \eta, \bar{\xi}(x_1) + \eta)$. On the other hand, populations x_0 and x_1 still behaves as in Step 1-Step 4 before time $\widetilde{T}^{\eta,1}$. Thus, we embed Figure 2.3 into Figure 2.4 and continue the proof based on the four-step analysis in (a).

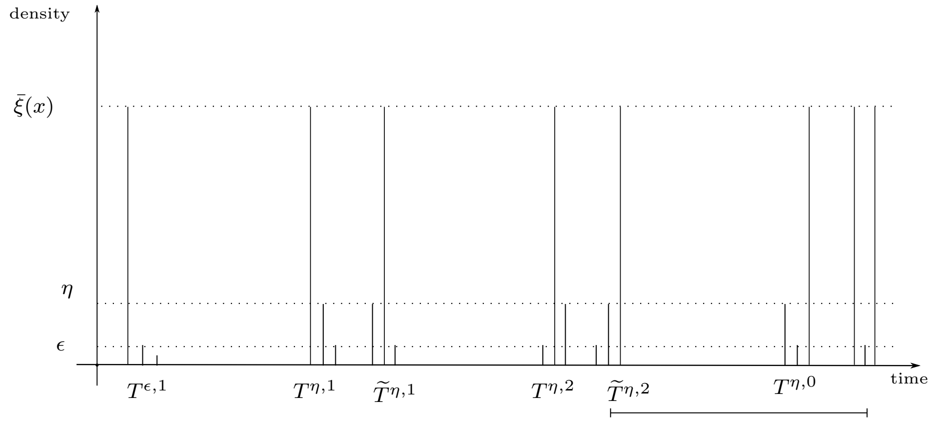

Figure 2.4: Three-type phase evolution

.

Let $T^{\eta,2}$ be the entrance time of the dynamics $\xi_t^\epsilon(x_2)$ into the $\eta-$level above 0. By similar arguments as used in Step 2, one can control $\xi_t^\epsilon(x_2)$ by way of $\check{\xi}_t(x_2) \leq \xi_t^\epsilon(x_2) \leq$

$\hat{\xi}_t(x_2)$ described as follows, for $\widetilde{T}^{\eta,1} < t < T^{\eta,2}$,

$$\dot{\check{\xi}}_t(x_2) = (f_{2,1} - \check{C}\eta)\check{\xi}_t(x_2), \tag{2.5.18}$$

and

$$\dot{\hat{\xi}}_t(x_2) = (f_{2,1} + \hat{C}\eta)\hat{\xi}_t(x_2), \tag{2.5.19}$$

where constants $\check{C}, \hat{C}$ change from line to line and $\check{\xi}_{\widetilde{T}^{\eta,1}}(x_2) = \hat{\xi}_{\widetilde{T}^{\eta,1}}(x_2) = \epsilon \cdot O(1)$. Then, it follows

$$\frac{1}{f_{2,1} + \hat{C}\eta} \ln \frac{1}{\epsilon} < T^{\eta,2} - \widetilde{T}^{\eta,1} < \frac{1}{f_{2,1} - \check{C}\eta} \ln \frac{1}{\epsilon}. \tag{2.5.20}$$

In all, from the comparison assumption (C3): $\frac{2}{b(x_2)-d(x_2)} > \frac{1}{f_{1,0}} + \frac{1}{f_{2,1}}$, we conclude that

$$\left(\frac{1}{f_{1,0}} + \frac{1}{f_{2,1}} - \delta\right) \ln \frac{1}{\epsilon} < T^{\eta,2} < \left(\frac{1}{f_{1,0}} + \frac{1}{f_{2,1}} + \delta\right) \ln \frac{1}{\epsilon}. \tag{2.5.21}$$

During time interval $[\widetilde{T}^{\eta,1}, T^{\eta,2}]$, population $\xi^\epsilon_t(x_0)$ still decays exponentially as in Step 4 with estimates $\check{\xi}_t(x_0) < \xi^\epsilon_t(x_0) < \hat{\xi}_t(x_0)$ described by inequality (2.5.12) and (2.5.13)

$$\check{\xi}_t(x_0) = \check{\xi}_{\widetilde{T}^{\eta,1}}(x_0) e^{(f_{0,1} - \check{C}\eta)(t - \widetilde{T}^{\eta,1})}, \tag{2.5.22}$$

and

$$\hat{\xi}_t(x_0) = \hat{\xi}_{\widetilde{T}^{\eta,1}}(x_0) e^{(f_{0,1} + \hat{C}\eta)(t - \widetilde{T}^{\eta,1})} \tag{2.5.23}$$

with $\check{\xi}_{\widetilde{T}^{\eta,1}}(x_0) = \hat{\xi}_{\widetilde{T}^{\eta,1}}(x_0) = \eta$.

Combining with (2.5.20), one obtains

$$\eta\epsilon^{-\frac{f_{0,1} - \check{C}\eta}{f_{2,1} - \check{C}\eta}} < \check{\xi}_{T^{\eta,2}}(x_0) < \xi^\epsilon_{T^{\eta,2}}(x_0) < \hat{\xi}_{T^{\eta,2}}(x_0) < \eta\epsilon^{-\frac{f_{0,1} + \hat{C}\eta}{f_{2,1} + \hat{C}\eta}}. \tag{2.5.24}$$

Taking the migration from neighbor site x_1 into account, the mass on x_0 should be of order $\epsilon^{\frac{|f_{0,1}|}{f_{2,1}}} \vee \epsilon$. Due to assumption $(C4): \frac{|f_{0,1}|}{f_{2,1}} < 1$, one obtains that $\epsilon^{\frac{|f_{0,1}|}{f_{2,1}}} \vee \epsilon = \epsilon^{\frac{|f_{0,1}|}{f_{2,1}}}$.

Short after $T^{\eta,2}$, as ϵ tends to 0, $\xi^\epsilon_t(x_1)$ and $\xi^\epsilon_t(x_2)$ can be approximated by Lotka-Volterra system as in Step 3. Mark by $\widetilde{T}^{\eta,2}$ the time when $(\xi^\epsilon_t(x_1), \xi^\epsilon_t(x_2))$ enters the $\eta-$neighborhood of the equilibrium $(0, \bar{\xi}(x_2))$, i.e. $\xi^\epsilon_{\widetilde{T}^{\eta,2}}(x_1) = \eta$. Also, $\widetilde{T}^{\eta,2} - T^{\eta,2}$ is of order $O(1)$. Thus, one obtains

$$\epsilon^{-\frac{f_{0,1} - \check{C}\eta}{f_{2,1} - \check{C}\eta}} \cdot O(\eta) < \xi^\epsilon_{\widetilde{T}^{\eta,2}}(x_0) = \xi^\epsilon_{T^{\eta,2}}(x_0) \cdot O(1) < \epsilon^{-\frac{f_{0,1} + \hat{C}\eta}{f_{2,1} + \hat{C}\eta}} \cdot O(\eta). \tag{2.5.25}$$

Let $T^{\eta,0}$ denote the time when $\xi^\epsilon_t(x_0)$ reaches $\eta-$level after time $\widetilde{T}^{\eta,2}$. For $\widetilde{T}^{\eta,2} < t < T^{\eta,0}$, $\xi^\epsilon_t(x_0)$ is governed approximately by a logistic equation

$$\dot{\xi}_t(x_0) = (b(x_0) - d(x_0) - \alpha(x_0, x_0)\xi_t(x_0))\xi_t(x_0). \tag{2.5.26}$$

Then, we have the differential inequality

$$(b(x_0) - d(x_0) - \alpha(x_0, x_0)\eta)\xi_t(x_0) < \dot{\xi}_t(x_0) < (b(x_0) - d(x_0))\xi_t(x_0), \tag{2.5.27}$$

with $\xi_{\widetilde{T}^{\eta,2}}(x_0)$ satisfying (2.5.25). Then, by Gronwall's inequality, one obtains

$$\begin{aligned}
-\frac{f_{0,1}+\hat{C}\eta}{(f_{2,1}+\hat{C}\eta)(b(x_0)-d(x_0))}\ln\frac{1}{\epsilon} &< T^{\eta,0}-\widetilde{T}^{\eta,2} \\
&< -\frac{f_{0,1}-\check{C}\eta}{(f_{2,1}-\check{C}\eta)(b(x_0)-d(x_0)-\alpha(x_0,x_0)\eta)}\ln\frac{1}{\epsilon}.
\end{aligned} \tag{2.5.28}$$

After $T^{\eta,0}$, $\xi_t^\epsilon(x_0)$ approaches $\bar{\xi}(x_0)$ in time of order 1.

By combining analysis in Step 2, and estimates in (2.5.20) and (2.5.28) together, because η is arbitrary, it follows that

$$\begin{aligned}
&\lim_{\eta\to 0}\frac{T^{\eta,1}}{\ln\frac{1}{\epsilon}}=\frac{1}{f_{1,0}}, \\
&\lim_{\eta\to 0}\frac{T^{\eta,2}-\widetilde{T}^{\eta,1}}{\ln\frac{1}{\epsilon}}=\frac{1}{f_{2,1}}, \\
&\lim_{\eta\to 0}\frac{T^{\eta,0}-\widetilde{T}^{\eta,2}}{\ln\frac{1}{\epsilon}}=\frac{-f_{0,1}}{f_{2,1}(b(x_0)-d(x_0))}.
\end{aligned} \tag{2.5.29}$$

Therefore, the dynamics X_t^ϵ, rescaled on time scale of order $\ln\frac{1}{\epsilon}$, converges to the TST U_t $(L=2)$ with the form (2.5.2). ☐

Proposition 2.5.3. *Admit the same conditions as in Theorem 2.3.1. Consider the case when $L=3$, i.e. $\mathcal{X}=\{x_0,x_1,x_2,x_3\}$. Then, the limit process $(U_t)_{t\geq 0}$ has the form*

$$U_t \equiv \begin{cases} \bar{\xi}(x_0)\delta_{x_0} & \textit{for } 0\leq t\leq I_1, \\ \bar{\xi}(x_1)\delta_{x_1} & \textit{for } I_1< t\leq I_2, \\ \bar{\xi}(x_2)\delta_{x_2} & \textit{for } I_2< t\leq I_3, \\ \bar{\xi}(x_3)\delta_{x_3} & \textit{for } I_3< t\leq I_3+S_1, \\ \bar{\xi}(x_1)\delta_{x_1}+\bar{\xi}(x_3)\delta_{x_3} & \textit{for } t> I_3+S_1. \end{cases} \tag{2.5.30}$$

Proof. (See Figure 2.5) From assumption (C3): $\frac{3}{b(x_3)-d(x_3)}>\frac{1}{f_{1,0}}+\frac{1}{f_{2,1}}$, it implies that population on trait site x_3 is negligible before $T^{\eta,2}$ and therefore it can not influence the evolution picture until $T^{\eta,2}$. Notice that we inherit the analysis and notations such as $T^{\eta,1},\widetilde{T}^{\eta,1}$, $T^{\eta,2},\widetilde{T}^{\eta,2}$ from the proof of Proposition 2.5.2.

Let $\xi_t^\epsilon(x_3):=\langle X_t^\epsilon,1_{\{x_3\}}\rangle$. Because $\xi_t^\epsilon(x_3)$ accounts for ϵ proportion of $\xi_t^\epsilon(x_2)$, it is negligible until $\widetilde{T}^{\eta,2}$. Then, we have

$$\begin{aligned}
\dot{\xi}_t^\epsilon(x_3)=&\big(b(x_3)-d(x_3)-\alpha(x_3,x_2)\xi_t^\epsilon(x_2)-\alpha(x_3,x_3)\xi_t^\epsilon(x_3)\big)\xi_t^\epsilon(x_3) \\
&-\epsilon\xi_t^\epsilon(x_3)b(x_3)m(x_3,x_2)+\epsilon\xi_t^\epsilon(x_2)b(x_2)m(x_2,x_3),
\end{aligned} \tag{2.5.31}$$

with initial value $\xi^\epsilon_{\widetilde{T}^{\eta,2}}(x_3)=\epsilon\cdot O(1)$.

Let $T^{\eta,0}$ and $T^{\eta,3}$ be the first time (resp.) for $\xi_t^\epsilon(x_0)$ and $\xi_t^\epsilon(x_3)$ to reach $\eta-$level after $\widetilde{T}^{\eta,2}$. By similar analysis to derive (2.5.20), we get

$$T^{\eta,3}-\widetilde{T}^{\eta,2}\sim f_{3,2}^{-1}\ln\frac{1}{\epsilon} \tag{2.5.32}$$

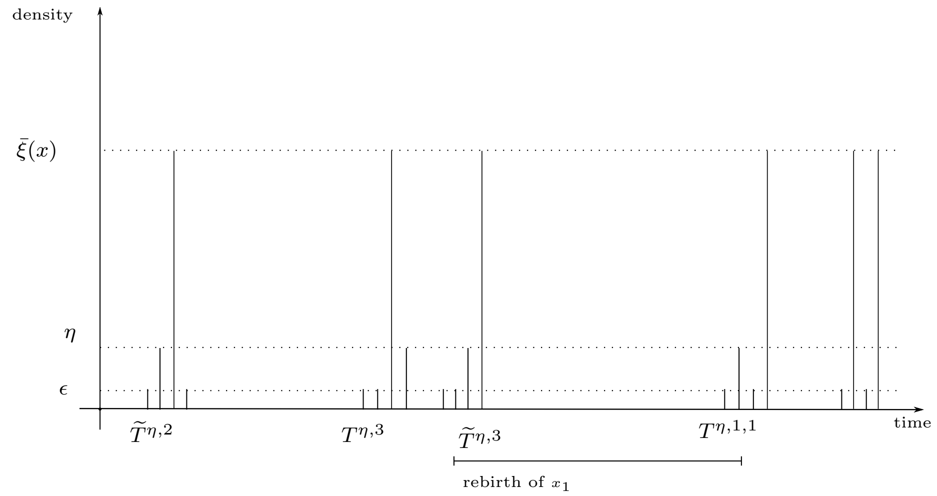

Figure 2.5: Four-type phase evolution

.

where relation $f(\epsilon) \sim g(\epsilon)$ means $\lim_{\epsilon\to 0} f(\epsilon)/g(\epsilon) = 1$.

From (2.5.28), recall that

$$T^{\eta,0} - \widetilde{T}^{\eta,2} \sim \frac{-f_{0,1}}{f_{2,1}(b(x_0)-d(x_0))} \ln\frac{1}{\epsilon}. \tag{2.5.33}$$

From assumption (C4), one obtains that

$$\frac{-f_{0,1}}{f_{2,1}(b(x_0)-d(x_0))} - \frac{1}{f_{3,2}} > \frac{-f_{1,2}}{f_{3,2}(b(x_1)-d(x_1))} > 0, \tag{2.5.34}$$

which implies $T^{\eta,3} < T^{\eta,0}$. Then, for $t \in [\widetilde{T}^{\eta,2}, T^{\eta,3}]$, $\xi^\epsilon_t(x_0)$ stays in some $\epsilon-$dependent infinitesimal neighborhood of 0. Furthermore, by some comparison results as in previous argument, $\xi^\epsilon_t(x_1)$ is mainly influenced from competition of $\bar{\xi}(x_2)$, one obtains, at time $T^{\eta,3}$,

$$\begin{aligned}\xi^\epsilon_{T^{\eta,3}}(x_1) &= \xi^\epsilon_{\widetilde{T}^{\eta,2}}(x_1)e^{f_{1,2}(T^{\eta,3}-\widetilde{T}^{\eta,2})}\\ &= \eta e^{f_{1,2}f_{3,2}^{-1}\ln\frac{1}{\epsilon}}\\ &\sim \eta\epsilon^{-\frac{f_{1,2}}{f_{3,2}}}.\end{aligned} \tag{2.5.35}$$

Similarly as in Step 3, short after $T^{\eta,3}$, $\xi^\epsilon_t(x_2)$ and $\xi^\epsilon_t(x_3)$ exchange rapidly in time of order 1 and $\xi^\epsilon_t(x_2)$ decreases into the $\eta-$neighborhood of 0 at time $\widetilde{T}^{\eta,3}$. After time $\widetilde{T}^{\eta,3}$, $\xi^\epsilon_t(x_1)$ evolves approximately as logistic growth curve since there is only negligible competition from neighbors x_0, x_2, i.e.

$$\dot{\xi}^\epsilon_t(x_1) = (b(x_1) - d(x_1) - \alpha(x_1,x_1)\xi^\epsilon_t(x_1))\xi^\epsilon_t(x_1), \tag{2.5.36}$$

with initial value $\xi^\epsilon_{\widetilde{T}^{\eta,3}}(x_1) \sim \eta\epsilon^{-\frac{f_{1,2}}{f_{3,2}}}$.

Denote by $T^{\eta,1,1}$ the second time for x_1 to reach $\eta-$level again after the first one $T^{\eta,1}$. Because of the exponential growth of logistic curve in the beginning period, in a similar method to deduce (2.5.33), one gets

$$T^{\eta,1,1}-\widetilde{T}^{\eta,3}\sim-\frac{f_{1,2}}{f_{3,2}(b(x_1)-d(x_1))}\ln\frac{1}{\epsilon}. \tag{2.5.37}$$

Again from (2.5.34), one observes that $T^{\eta,1,1}<T^{\eta,0}$, that is, $\xi_t^\epsilon(x_1)$ recovers to reach $\eta-$level earlier than $\xi_t^\epsilon(x_0)$. Consequently, $\xi_t^\epsilon(x_0)$ will drift to 0 due to competition from fitter type x_1.

Combining (2.5.32), (2.5.37) and the first two equations in (2.5.29) together, we conclude the TST limit for $L=3$ on a new time scale. □

Lemma 2.5.4. *Assumption (C4) implies the following inequalities, for any* $4\leq L\in\mathbb{N}$

$$\begin{aligned}
&\frac{-f_{0,1}}{f_{2,1}(b(x_0)-d(x_0))}>\frac{1}{f_{3,2}}+\cdots+\frac{1}{f_{L,L-1}},\\
&\frac{-f_{1,2}}{f_{3,2}(b(x_1)-d(x_1))}>\frac{1}{f_{4,3}}+\cdots+\frac{1}{f_{L,L-1}},\\
&\vdots\\
&\frac{-f_{L-3,L-2}}{f_{L-1,L-2}(b(x_{L-3})-d(x_{L-3}))}>\frac{1}{f_{L,L-1}}
\end{aligned} \tag{2.5.38}$$

and

$$\begin{aligned}
&\frac{-f_{L-4,L-3}}{f_{L-2,L-3}(b(x_{L-4})-d(x_{L-4}))}-\frac{1}{f_{L-1,L-2}}-\frac{1}{f_{L,L-1}}>\frac{-f_{L-2,L-1}}{f_{L,L-1}(b(x_{L-2})-d(x_{L-2}))}\\
&\vdots
\end{aligned} \tag{2.5.39}$$

and so on.

The proof of this Lemma follows straightforward iterations from assumption (C4). Roughly speaking, from (2.5.38), it says that, when it passes to the limit process U_t, all dynamics except $\xi_t^\epsilon(x_L)$ stay in $\epsilon-$dependent infinitesimal neighborhoods of 0 at time $T^{\eta,L}$ which denotes the dominating time for type x_L. It leads to monomorphic transportation of the mass from initial trait x_0 to the fittest trait x_L in the first half period. On the other hand, from (2.5.39), it guarantees that the fitter one recovers earlier than the unfit traits alternatively backwards to the most unfit one. .

Proof of Theorem 2.3.1. After proving the first two cases in previous propositions, we proceed our proof along two lines, one line is the case when L is an even integer while the other one is when L is an odd integer. Due to the similar fashion, we just do the recursive procedure from $L=2$ to $L=4$ (see Figure 2.2). Here, $\mathcal{X}=\{x_0,x_1,x_2,x_3,x_4\}$.

Based on the analysis in Proposition 2.5.3, after time $\widetilde{T}^{\eta,3}$, we move forward by introducing $T^{\eta,4}$, which is defined as the first hitting time of $\eta-$level for $\xi_t^\epsilon(x_4)$. As before, we can show that $T^{\eta,4}$ is far more longer than $\widetilde{T}^{\eta,3}$ from assumption (C3). One observes that

$$T^{\eta,4} - \widetilde{T}^{\eta,3} \sim \frac{1}{f_{4,3}} \ln \frac{1}{\epsilon}. \tag{2.5.40}$$

It follows mass exchange between trait site x_3 and x_4 until $\widetilde{T}^{\eta,4}$ when $\xi^{\epsilon}_{\widetilde{T}^{\eta,4}}(x_3) = \eta$. Then, $\xi^{\epsilon}_t(x_2)$ begins to recover due to the lack of competition from $\xi^{\epsilon}_t(x_3)$. The time length needed for $\xi^{\epsilon}_t(x_2)$ to reach $\eta-$level again (mark by $T^{\eta,2,2}$) can be computed explicitly

$$T^{\eta,2,2} - \widetilde{T}^{\eta,4} \sim \frac{-f_{2,3}}{f_{4,3}(b(x_2) - d(x_2))} \ln \frac{1}{\epsilon}. \tag{2.5.41}$$

Then, it will approach equilibrium $\bar{\xi}(x_2)$ according to logistic growth. Consequently, $\xi^{\epsilon}_t(x_1)$ will drift to 0 due to competition from the more fitter neighbor x_2. Recall from (2.5.33), one obtains

$$T^{\eta,0} - \widetilde{T}^{\eta,4} \sim \left[\frac{-f_{0,1}}{f_{2,1}(b(x_0) - d(x_0))} - \frac{1}{f_{3,2}} - \frac{1}{f_{4,3}}\right] \ln \frac{1}{\epsilon}. \tag{2.5.42}$$

Combining the above two estimates with assumption (2.5.39), one observes that

$$T^{\eta,2,2} - \widetilde{T}^{\eta,4} < T^{\eta,0} - \widetilde{T}^{\eta,4}. \tag{2.5.43}$$

Furthermore, we have

$$\begin{aligned} \lim_{\eta\to 0} \frac{T^{\eta,2,2}}{\ln\frac{1}{\epsilon}} &= \frac{1}{f_{1,0}} + \frac{1}{f_{2,1}} + \frac{1}{f_{3,2}} + \frac{1}{f_{4,3}} + \frac{|f_{2,3}|}{f_{4,3}(b(x_2) - d(x_2))} \\ &= I_4 + S_2, \end{aligned} \tag{2.5.44}$$

and

$$\lim_{\eta\to 0} \frac{T^{\eta,0}}{\ln\frac{1}{\epsilon}} = I_2 + S_0. \tag{2.5.45}$$

We thus get the explicit form of (2.3.5) when $L = 4$. Recursively, we get another TST process ($L = 6$) by connecting the TST process ($L = 4$) with a subtree consisting traits $\{x_4, x_5, x_6\}$ defined as in Proposition 2.5.2. So on and so forth, it follows (2.3.5) for all even integers L. ☐

2.5.2 Proof of Theorem 2.4.2

The proof of Theorem 2.4.2 should consist of two parts, firstly the convergence of exponential jump times on a proper time scale and secondly the transition rule from current configuration to the new one. To show the idea of proof, we'll just prove the first case in Definition 2.4.1. The proof of the second case is in a same fashion. We use the induction method to prove it. We list the following lemmas which are key to the proof of Theorem 2.4.2.

For nonnegative integer l, denote by $\Gamma^{(2l)}$ the atomic measure with finite support, i.e., $\Gamma^{(2l)}=\sum_{i=0}^{l}\bar{\xi}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$. Similarly, set $\Gamma^{(2l+1)}=\sum_{i=1}^{l+1}\bar{\xi}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$, whose form is described as in Definition 2.4.1. From $\Gamma^{(0)}$ to $\Gamma^{(1)}$, it is trivial to prove the theorem as done in Proposition 2.5.2 (a). To the end, we will deduce it still holds from $\Gamma^{(2l)}$ to $\Gamma^{(2l+1)}$. Denote by $\mathbb{P}_{\Gamma^{(2l)}}$ the law of the process $X_{\cdot}^{\epsilon,\sigma}$ with initial configuration $\Gamma^{(2l)}$. Denote by τ^{ϵ} the first time after 0 when there occurs a new mutant trait.

Lemma 2.5.5. *Admit the same conditions as in Theorem 2.4.2.*

$$\lim_{\epsilon\to 0}\mathbb{P}_{\Gamma^{(2l)}}\big(\tau^{\epsilon}>\frac{t}{\sigma}\big)=\exp\Big(-t\sum_{i=0}^{l}\bar{\xi}(x_{2i}^{(2l)})\mu(x_{2i}^{(2l)})\Big). \tag{2.5.46}$$

$$\lim_{\epsilon\to 0}\mathbb{P}_{\Gamma^{(2l)}}\big(\textit{at time}\,\tau^{\epsilon},\ \textit{mutant comes from trait}\,x_{2k}^{(2l)}\big)=\frac{\bar{\xi}(x_{2k}^{(2l)})\mu(x_{2k}^{(2l)})}{\sum_{i=0}^{l}\bar{\xi}(x_{2i}^{(2l)})\mu(x_{2i}^{(2l)})}. \tag{2.5.47}$$

The proof of this lemma can be deduced from the expression of its construction as done in [8, Lemma 2 (c)]. We will not show the details here.

Lemma 2.5.6. *Assume $X_0^{\epsilon,\sigma}=\Gamma^{(2l)}+\delta_{x_{2k}^{(2l)}+h}$. Then, for any $\varepsilon>0$, there exists a constant $C>0$ such that*

$$\lim_{\epsilon\to 0}\mathbb{P}\big(\sup_{t\in(C\ln\frac{1}{\epsilon},\tau^{\epsilon})}\|X_t^{\epsilon,\sigma}-\Gamma^{(2l+1)}\|<\varepsilon\big)=1 \tag{2.5.48}$$

where $\Gamma^{(2l+1)}$ is defined as in Definition 2.4.1 (i) and $\|\cdot\|$ is the total variation distance.

Proof. From Lemma 2.5.5, one concludes that, for any $C>0$,

$$\lim_{\epsilon\to 0}\mathbb{P}\big(\tau^{\epsilon}>C\ln\frac{1}{\epsilon}\big)=1.$$

According to the fitness landscape, there will be one and only one ordered position for the new trait $x_{2k}^{(2l)}+h$ in $\Gamma^{(2l)}$. Suppose there exists $x_{2j}^{(2l)}$ such that $x_{2k}^{(2l)}+h$ fits between $x_{2j}^{(2l)}$ and $x_{2j+1}^{(2l)}$. Then, one has the local fitness order

$$x_{2j-1}^{(2l)}\prec x_{2j}^{(2l)}\prec x_{2k}^{(2l)}+h\prec x_{2j+1}^{(2l)}. \tag{2.5.49}$$

Since it is unpopulated for both traits $x_{2j-1}^{(2l)}$ and $x_{2j+1}^{(2l)}$ in $\Gamma^{(2l)}$, we consider $\big(x_{2j}^{(2l)},\,x_{2k}^{(2l)}+h\big)$ as an isolated pair without competition from others. As the same analysis as being done in Proposition 2.5.2, the two-type system will converge to $\big(0,\,\bar{\xi}(x_{2k}^{(2l)}+h)\delta_{x_{2k}^{(2l)}+h}\big)$ in time $O\left(\ln\frac{1}{\epsilon}\right)$. On the right hand side of the pair, nothing changes due to their isolation. Whereas on the left hand side of the pair, trait $x_{2j-1}^{(2l)}$ increases exponentially due to the decay of its fitter neighbor $x_{2j}^{(2l)}$. So on and so forth, the mass occupation switches on the left hand side of $x_{2j}^{(2l)}$. The entire rearrangement process can be completed in time of order $O(\ln\frac{1}{\epsilon})$ as the same arguments in the finite trait case (see Section 2.5.1).

We can prove the other case when the fitness location of $x_{2k}^{(2l)}+h$ is on the left hand side of $x_{2j}^{(2l)}$, that is

$$x_{2j-1}^{(2l)} \prec x_{2k}^{(2l)}+h \prec x_{2j}^{(2l)} \prec x_{2j+1}^{(2l)},$$

in a similar method. In all, we give the new configuration $\Gamma^{(2l+1)}$ by relabeling the traits as done in Definition 2.4.1 (i). □

From Lemma 2.5.5, it shows us the mutation occurs on the time scale $O(\frac{1}{\sigma})$. Recall from Section 2.3 in the finite trait case that the fixation time scale is $O\left(\ln\frac{1}{\epsilon}\right)$. Combining them with the time scale separation constraint $\frac{1}{\sigma} \gg \ln\frac{1}{\epsilon}$, thanks to the characterization of the rate of exponentially distributed waiting time and transition rule of configurations proved in previous lemmas, the proof of Theorem 2.4.2 will be straightforward by the general construction of a Markov jump process.

2.6 Simulation algorithm

The pathwise construction of the TST process defined in Definition 2.4.1 leads to the following numerical algorithm for simulation of the TST process.

Step 0. Specify with initial condition: $\Gamma_0 = \Gamma^{(0)} = \bar{\xi}(x_0)\delta_{x_0}$.

Step 1. Simulate τ_1 exponential distributed with parameter $\bar{\xi}(x_0)\mu(x_0)$. Sample a new trait (x_0+h) with density $p(x_0, dh)$. If $f(x_0+h, x_0) > 0$, relabel $x_0^{(1)} := x_0$, $x_1^{(1)} := x_0+h$. Otherwise, relabel $x_0^{(1)} := x_0+h$, $x_1^{(1)} := x_0$.
Set $\Gamma^{(1)} = \bar{\xi}(x_1^{(1)})\delta_{x_1^{(1)}}$, and $\Gamma_t = \Gamma^{(0)}$ for $t \in [0, \tau_1)$.

Step 2. Simulate τ_2 exponential distributed with parameter $\bar{\xi}(x_1^{(1)})\mu(x_1^{(1)})$.
Set $\Gamma_t = \Gamma^{(1)}$ for $t \in [\tau_1, \tau_1+\tau_2)$.
Sample a new trait $\big(x_1^{(1)}+h\big)$ with density $p(x_1^{(1)}, dh)$.
Choose one from the following to carry out:

- if $f(x_1^{(1)}+h, x_1^{(1)}) > 0$, relabel $x_0^{(2)} := x_0^{(1)}$, $x_1^{(2)} := x_1^{(1)}$, $x_2^{(2)} := x_1^{(1)}+h$;
- if $f(x_1^{(1)}+h, x_1^{(1)}) < 0$, $f(x_1^{(1)}+h, x_0^{(1)}) > 0$, relabel $x_0^{(2)} := x_0^{(1)}$, $x_1^{(2)} := x_1^{(1)}+h$, $x_2^{(2)} := x_1^{(1)}$;
- if $f(x_1^{(1)}+h, x_0^{(1)}) < 0$, relabel $x_0^{(2)} := x_1^{(1)}+h$, $x_1^{(2)} := x_0^{(1)}$, $x_2^{(2)} := x_1^{(1)}$.

Set $\Gamma^{(2)} = \bar{\xi}(x_0^{(2)})\delta_{x_0^{(2)}} + \bar{\xi}(x_2^{(2)})\delta_{x_2^{(2)}}$.

Step 2l+1. Generate $\Gamma^{(2l+1)}$ from $\Gamma^{(2l)} = \sum\limits_{i=0}^{l} \bar{\xi}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$ for $l = 1, 2, \cdots$.

Simulate τ_{2l+1} exponential distributed with parameter $\sum\limits_{i=0}^{l} \bar{\xi}(x_{2i}^{(2l)})\mu(x_{2i}^{(2l)})$.

Set $\Gamma_t = \Gamma^{(2l)}$ for $t \in \big[\sum\limits_{i=1}^{2l}\tau_i, \sum\limits_{i=1}^{2l+1}\tau_i\big)$. Select one trait $x_{2k}^{(2l)}$, for any $0 \le k \le l$, to mutate with probability $\frac{\bar{\xi}(x_{2k}^{(2l)})\mu(x_{2k}^{(2l)})}{\sum_{i=0}^{l}\bar{\xi}(x_{2i}^{(2l)})\mu(x_{2i}^{(2l)})}$. Sample a new trait $\big(x_{2k}^{(2l)}+h\big)$ with probability density $p(x_{2k}^{(2l)}, dh)$. Choose one from the following three cases to carry out:

- if $f(x_{2k}^{(2l)}+h, x_{2l}^{(2l)}) > 0$, relabel $x_{2l+1}^{(2l+1)} := x_{2k}^{(2l)}+h$, $x_i^{(2l+1)} := x_i^{(2l)}$ for $0 \le i \le 2l$;
- if $f(x_{2k}^{(2l)}+h, x_0^{(2l)}) < 0$, relabel $x_i^{(2l+1)} := x_{i-1}^{(2l)}$ for $1 \le i \le 2l+1$, and $x_0^{(2l+1)} := x_{2k}^{(2l)}+h$;
- otherwise, there exists $0 \le j < l$ s.t. $f(x_{2k}^{(2l)}+h, x_{2i}^{(2l)}) < 0$ for $j < i \le l$, and $f(x_{2k}^{(2l)}+h, x_{2j}^{(2l)}) > 0$. Furthermore,
    - ∗ if $f(x_{2k}^{(2l)}+h, x_{2j+1}^{(2l)}) < 0$, relabel $x_i^{(2l+1)} := x_i^{(2l)}$ for $0 \le i \le 2j$, $x_i^{(2l+1)} := x_{i-1}^{(2l)}$ for $2j+2 \le i \le 2l+1$, and $x_{2j+1}^{(2l+1)} := x_{2k}^{(2l)}+h$;
    - ∗ if $f(x_{2k}^{(2l)}+h, x_{2j+1}^{(2l)}) > 0$, relabel $x_i^{(2l+1)} := x_i^{(2l)}$ for $0 \le i \le 2j+1$, $x_i^{(2l+1)} := x_{i-1}^{(2l)}$ for $2j+3 \le i \le 2l+1$, and $x_{2j+2}^{(2l+1)} := x_{2k}^{(2l)}+h$.

Set $\Gamma^{(2l+1)} = \sum\limits_{i=1}^{l+1} \bar{\xi}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$.

Step 2l+2. To generate $\Gamma^{(2l+2)}$ from $\Gamma^{(2l+1)} = \sum\limits_{i=1}^{l+1} \bar{\xi}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$ for $l = 1, 2, \cdots$.

This can be done as similar as the induction from $\Gamma^{(2l)}$ to $\Gamma^{(2l+1)}$. So forth.

3 Microscopic interpretation of the trait substitution tree model

We consider a structured population model with competition and migration between nearest neighbors. We are particularly interested in the asymptotic behavior of the total population partition on supporting trait sites, under combination of large population and rare migration. For the population without mutation on a finite-many trait space, we give the equilibrium configuration and characterize the right time scale for fixation. For the model with mutation on an infinite-many trait space, a Markov jump process-trait substitution tree model is established on the rarer mutation time scale against the rare migration constraint in terms of a large population limit.

3.1 Introduction

In recent years a spatially structured population with migration (dispersion) and local regulation, proposed by Bolker and Pacala [4], Dieckmann and Law [35] (in short BPDL process), has attracted particular interest both from biologists and mathematicians. It has several advantages over general branching processes, which make it more natural as population models: the quadratic competition term is used to prevent the population size from escaping to infinity and migration term is used to transport the population mass from one colony to unoccupied colonies for survival, and to further get colonized. There are mainly two highlights of related papers. Etheridge [19], Fournier and Méléard [22], Hutzenthaler and Wakolbinger [27], and Hutzenthaler [26] have studied the extinction and survival problems. Champagnat [8], Champagnat and Lambert [9], Méléard and Tran [38], Dawson and Greven [16] focus more on its long time behavior by multiscale analysis methods.

The main ingredient behind this model is logistic branching random walks, that is, a combination of logistic branching populations with spatial random walks (or migration) on trait sites. In [8], a so-called trait substitution sequence model (in short TSS) is derived under a combination of a large population and rare migration (namely mutation there) limit. More precisely, rare migration rate is constrained by the large population parameter so that the fixation period is not visible on the migration time scale. It is guaranteed that a single migrant arises in the resident population and no further migrants occur until the fate of the first migrant population is known. In other words, the migration time scale and branching time scale can be separated clearly.

In this paper we are interested in the case when the migrant event is still rare with respect to branching events but not that rare as in [8] (see Figure 3.1). In contrast, we assume that there are infinite migrants from a resident population on the natural time scale. Let a parameter ϵ be the migration rate and K be proportional to the initial population size. We will impose the rare migration constraint $1 \ll K\epsilon \ll K$ on the population (see parameter region II in Figure 3.1). As far as a finite-trait dynamic system is concerned, to find out the exact fixation time scale expressed in terms of the migration rate and population size is of particular interest for us. Since the original model is not easy for us to study due to the complicated interactions, we present here a slightly modified model of the one in [8] but retaining the essential machinery founded in the original model. This paper is restricted with nearest-neighbor competitions and migrations along the monotone fitness landscape. What is more, in order to study the long time behavior, we introduce mutations to drive the population to move towards more fitter configuration on a rare mutation time scale, which is longer than the fixation time scale. Note that the limit theorem arising in [8] can be applied consistently in the model developed in this paper.

The purpose of this paper and the accompanying one [6] is to justify a trait substitution tree process (in short TST) to illustrate the coexistence phenomenon with spatial structure in evolution theory, which is a purely atomic finite measure-valued process. The present one is derived from the microscopic point of view while the other one in [6] is from the macroscopic point of view. Combined these two papers together with [8], the entire framework on (rare) migration against (large) population limit can be fully characterized and it generates different rescaling limits-TSS and TST respectively on different time scales. In summary, the entire framework is as follows:

- Take large population and rare migration simultaneously by $K\epsilon \ll \frac{1}{\ln K}$, it leads to a TSS limit in [8].
- Firstly let $K \to \infty$, then add rare mutation by $\ln \frac{1}{\epsilon} \ll \frac{1}{\sigma}$ as $\epsilon \to 0$, it leads to a TST limit in [6].
- Take large population, rare migration and even rarer mutation all simultaneously constrained by $1 \ll K\epsilon \ll K$, $\ln \frac{1}{\epsilon} \ll \frac{1}{K\sigma}$. That is our goal in this paper.

The remainder of the paper is structured as follows. In Section 3.2, we present a description of the individual-based model. In Section 3.3, we consider the case without mutation but on a finite-trait space, and characterize the rare migration limit against the large population limit. In Section 3.4, concerning a modified population supported on an infinite-trait space by introducing mutations, we justify a so-called trait substitution tree process in the rare mutation limit, which already appeared in [6]. In the last section, related proofs for results in previous sections are provided.

3.2 Microscopic model

We begin with an individual-based model. Assume that the population at time t is composed of a finite number I_t individuals characterized by their phenotypic traits $x_1(t), \ldots, x_{I_t}(t)$ belonging a compact subset $\mathcal{X}$ of $\mathbb{R}^d$. We denote by $\mathcal{M}_F(\mathcal{X})$ the set

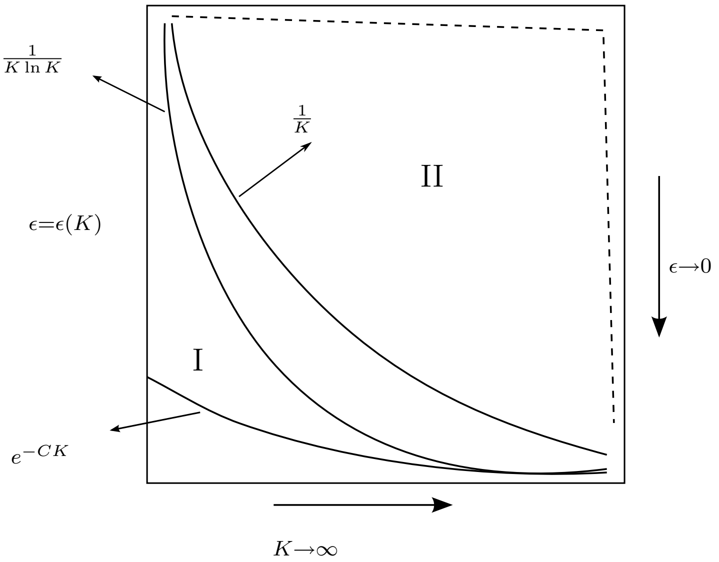

Figure 3.1: Parameter region separation: migration rate ϵ against population size K

.

of non-negative finite measures on $\mathcal{X}$. Let $\mathcal{M}_a(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$ be the set of counting measures on $\mathcal{X}$:

$$\mathcal{M}_a(\mathcal{X}) = \left\{ \sum_{i=1}^{n} \delta_{x_i} : x_1, \dots, x_n \in \mathcal{X}, n \in \mathbb{N} \right\}.$$

Then, the population process at time t can be represented as:

$$\nu_t = \sum_{i=1}^{I_t} \delta_{X_i(t)}.$$

Let $B(\mathcal{X})$ denote the totality of functions on $\mathcal{X}$ which are bounded and measurable. For any $f \in B(\mathcal{X})$, $\nu \in \mathcal{M}_F(\mathcal{X})$, we use notation $\langle \nu, f \rangle = \int f d\nu$.

Let's specify the population process $(\nu_t)_{t>0}$ by introducing a sequence of biological parameters:

- $b(x)$ is the birth rate from an individual with trait x.
- $d(x)$ is the death rate of an individual with trait x because of "aging".
- $\alpha(x, y)$ is the competition kernel felt by some individual with trait x from another individual with trait y.
- $m(x, dy)$ is the migration law of an individual from trait site x to site y.
- $\mu(x)$ is the mutation rate of an individual with trait x.
- $p(x, dh)$ is the law of mutant variation $h = y - x$ between a mutant y and its resident trait x. Since the mutant trait $y = x + h$ should belong to $\mathcal{X}$, this law has its support in $\mathcal{X} - x := \{y - x : y \in \mathcal{X}\} \subset \mathbb{R}^d$.

To specify the model without mutation mechanism, the infinitesimal generator of the $\mathcal{M}_a(\mathcal{X})$-valued process is given as follows, for any $F \in B(\mathcal{M}_a(\mathcal{X}))$:

$$\begin{aligned} LF(\nu) &= \sum_{i=1}^{I} \left[F(\nu + \delta_{x_i}) - F(\nu)\right] b(x_i) \\ &+ \sum_{i=1}^{I} \left[F(\nu - \delta_{x_i}) - F(\nu)\right] \left(d(x_i) + \sum_{j \neq i}^{I} \alpha(x_i, x_j) \right) \\ &+ \sum_{i=1}^{I} \sum_{x_j \neq x_i} \left[F(\nu + \delta_{x_j} - \delta_{x_i}) - F(\nu)\right] m(x_i, x_j). \end{aligned} \tag{3.2.1}$$

The first term above describes the clonal reproduction at the mother's site. The second term describes death of an individual x_i either due to aging or competition from another individual x_j. And the last term describes the migration of an individual from trait site x_i to site x_j.

By introducing a parameter $K \in \mathbb{N}$, we rescale the population size and competition kernel by K. We will show later, as K tends to infinity, one can get different large population limits by well-chosen rescaling procedure. Furthermore, the population process can be parameterized by another parameter ϵ governing the rate of migration law $m(x_i, x_j)$ in terms of population size scaling parameter K.

For any $K \in \mathbb{N}$, instead of studying the above process $(\nu_t^K)_{t \geq 0}$, it is more convenient to consider a sequence of rescaled measure-valued processes:

$$X_t^K := \frac{1}{K} \nu_t^K = \frac{1}{K} \sum_{i=1}^{I_t^K} \delta_{x_i} \tag{3.2.2}$$

where $X_\cdot^K$ is a $\mathcal{M}_F(\mathcal{X})-$valued process with the following infinitesimal generator:

$$\begin{aligned} L^K F(\nu) &= \int_{\mathcal{X}} \left[F(\nu + \frac{\delta_x}{K}) - F(\nu)\right] b(x) K \nu(dx) \\ &+ \int_{\mathcal{X}} \left[F(\nu - \frac{\delta_x}{K}) - F(\nu)\right] \left(d(x) + \int_{\mathcal{X}} \alpha(x, y) \nu(dy) \right) K \nu(dx) \\ &+ \epsilon \int_{\mathcal{X}} \int_{\mathcal{X}} \left[F(\nu + \frac{\delta_y}{K} - \frac{\delta_x}{K}) - F(\nu)\right] m(x, dy) K \nu(dx). \end{aligned} \tag{3.2.3}$$

Notice that we actually rescale the competition kernel α by K so that the system mathematically makes sense when we take a large population limit. From the biological point of view, K can be interpreted as scaling the resource or area available.

Let us denote by (**A**) the following assumptions.

(A1) $\exists\ \bar{b},\ \bar{d},\ \underline{\alpha}\ , \bar{\alpha}$, such that $0 < b(x) \leq \bar{b},\ 0 < d(x) \leq \bar{d},\ 0 < \underline{\alpha} \leq \alpha(x,y) \leq \bar{\alpha}$, and $b(x) - d(x) > 0,\ \forall x \in \mathcal{X}$.

(A2) $\forall\ x, y \in \mathcal{X}$, $\bar{f}(x,y) \cdot \bar{f}(y,x) < 0$, where the fitness functions

$\bar{f}(y,x) = b(y) - d(y) - \alpha(y,x)\bar{n}(x)$ and $\bar{n}(x) = \frac{b(x)-d(x)}{\alpha(x,x)}$,

$\bar{f}(x,y) = b(x) - d(x) - \alpha(x,y)\bar{n}(y)$ and $\bar{n}(y) = \frac{b(y)-d(y)}{\alpha(y,y)}$.

Notice that assumption (A1) guarantees that the process with infinitesimal generator (3.2.3) is well defined (refer to [22]). Assumptions (A2) gives the non-coexistence condition for any pair of competing populations.

3.3 Early time window on finite trait space as $K \to \infty$

We firstly review some already known results. Champagnat [8, Theorem 1] proved the following result by the time scales separation technique, which can be extended to a more general case in accelerated population dynamics [45].

Theorem 3.3.1. *Admit assumptions (**A**1) and (**A**2). Suppose that* $X_0^K = \frac{N_0^K}{K}\delta_x$ *such that* $\frac{N_0^K}{K} \overset{law}{\to} n_0 > 0$ *as* $K \to +\infty$, *and* $\forall\ C > 0$,

$$\exp\{-CK\} \ll K\epsilon \ll \frac{1}{\ln K}. \tag{3.3.1}$$

Then, $(X_{t/K\epsilon}^K, t \geq 0)$ *converges in the sense of f.d.d to*

$$Y_t = \begin{cases} n_0\delta_x, & t = 0 \\ \bar{n}(\eta_t)\delta_{\eta_t}, & t > 0 \end{cases}$$

where the Markov jump process $(\eta_t, t \geq 0)$ *satisfies* $\eta_0 = x$ *with an infinitesimal generator:*

$$A\varphi(x) = \int_{\mathcal{X}} (\varphi(y) - \varphi(x))\bar{n}(x)\frac{[\bar{f}(y,x)]_+}{b(y)}m(x,dy). \tag{3.3.2}$$

Remark 3.3.2.

- *The migration time scale is of order* $\frac{1}{K\epsilon}$ *whereas the fixation time scale starting from one migrant is of order* $\ln K$. *The population is kept monomorphic on the rare migration time scale. The rare migration parameter region imposed by* (3.3.1) *is denoted by the region **I** in Figure 3.1.*

- *As showed in Figure 3.2, it simulates a TSS model with trait space comprising of three types in the left one while in the right one it simulates a four-type case. We mark the population density of trait* x_0, x_1, x_2, x_3 *by red, blue, green and black colored curve respectively. Take* $b(x_0) = 3$, $b(x_1) = 6$, $b(x_2) = 8$, $b(x_3) = 10$ *and death rates* $d(x_i) \equiv 0$, $i = 0, 1, 2, 3$. *Take competition kernel* $\alpha \equiv 1$, *migration kernel* $m \equiv 0.5$, *and migration parameter* $\epsilon = K^{-2}$, *where initial population size* $K = 100$.

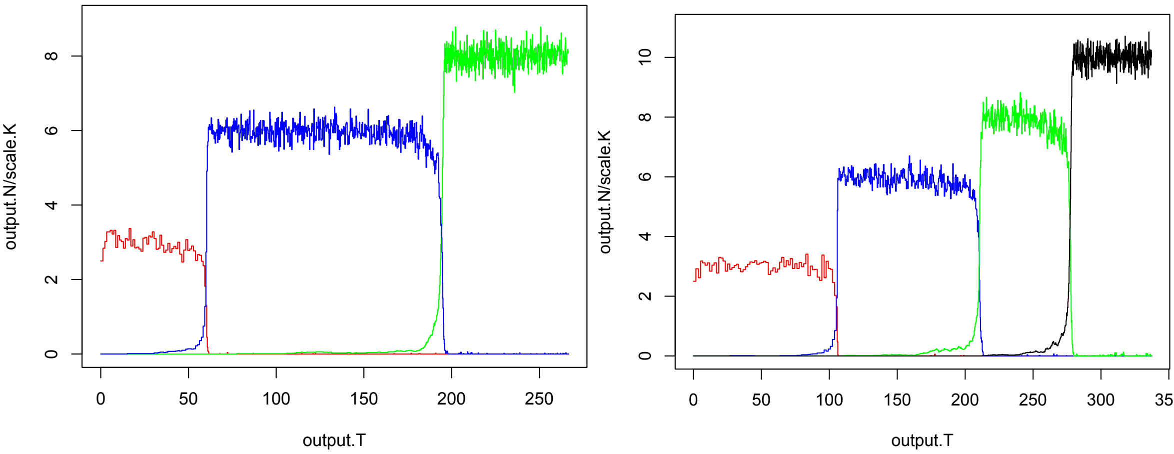

Figure 3.2: Simulations of the trait substitution sequence model arising in Theorem 3.3.1.

In [6], we firstly let K tend to infinity in (3.2.3) and get a deterministic limit. Then, we consider the rescaling limit of the deterministic system supported on a finite trait space under a slow migration limit. This is actually an extreme case where it attains the limiting process (trait substitution tree) by taking sequence limits along the marginal path (see dashed line in Figure 3.1). In terms of the individual-based population, it is of particular interest for us to give a microscopic interpretation of the TST process under some suitable conditions. Prior to the following theorem, we give assumption (**B**) to assist the following results.

(B1) For any finite number of types $L \in \mathbb{N}$, it has a monotonously increasing fitness landscape: $x_0 \prec x_1 \prec \ldots \prec x_L$, where $x_0 \prec x_1$ denotes $\bar{f}(x_0, x_1) < 0$, $\bar{f}(x_1, x_0) > 0$.

(B2) Nearest neighbor migration and competition, i.e. $m(x_i, x_j) = \alpha(x_i, x_j) = 0$ for any $\mid i - j \mid > 1$.

(B3) For any $i \geq 2$,

$$\frac{i}{b(x_i) - d(x_i)} \geq \frac{1}{\bar{f}(x_i, x_{i-1})} + \frac{1}{\bar{f}(x_{i-1}, x_{i-2})} + \cdots + \frac{1}{\bar{f}(x_1, x_0)}. \tag{3.3.3}$$

Note that assumption (B3) is not necessary for us to obtain the following theorem. There actually exist a variety of different possible paths to converge to the equilibrium configuration determined up to the ordered sequence of traits as in assumption (B1). However, thanks to assumption (B3), it brings us a lot convenience to prove the theorem without losing intrinsic content.

We inherit some notations from [6], denote configurations by $\Gamma^{(L)} := \sum_{i=0}^{l} \bar{n}(x_{2i})\delta_{x_{2i}}$ if $L = 2l$ and $\sum_{i=1}^{l+1} \bar{n}(x_{2i-1})\delta_{x_{2i-1}}$ if $L = 2l + 1$ for any $l \in \mathbb{N} \cup 0$.

Theorem 3.3.3. *Admit assumptions* $(\boldsymbol{A}1)$ *and* $\boldsymbol{B}$*. Consider the processes* $(X_t^K)_{t\geq 0}$ *on the trait space* $\mathcal{X} = \{x_0, x_1, \dots, x_L\}$*. Suppose that* $X_0^K = \frac{N_0^K}{K}\delta_{x_0}$ *such that* $\frac{N_0^K}{K} \xrightarrow{law} n_0 > 0$ *as* $K \to +\infty$*, and*

$$1 \ll K\epsilon \ll K. \tag{3.3.4}$$

Then there exists a constant $\bar{t}_L > 0$*, such that for any* $t > \bar{t}_L$*,* $\lim\limits_{K\to\infty} X^K_{t\ln\frac{1}{\epsilon}} \overset{(d)}{=} \Gamma^{(L)}$ *under the total variation norm.*

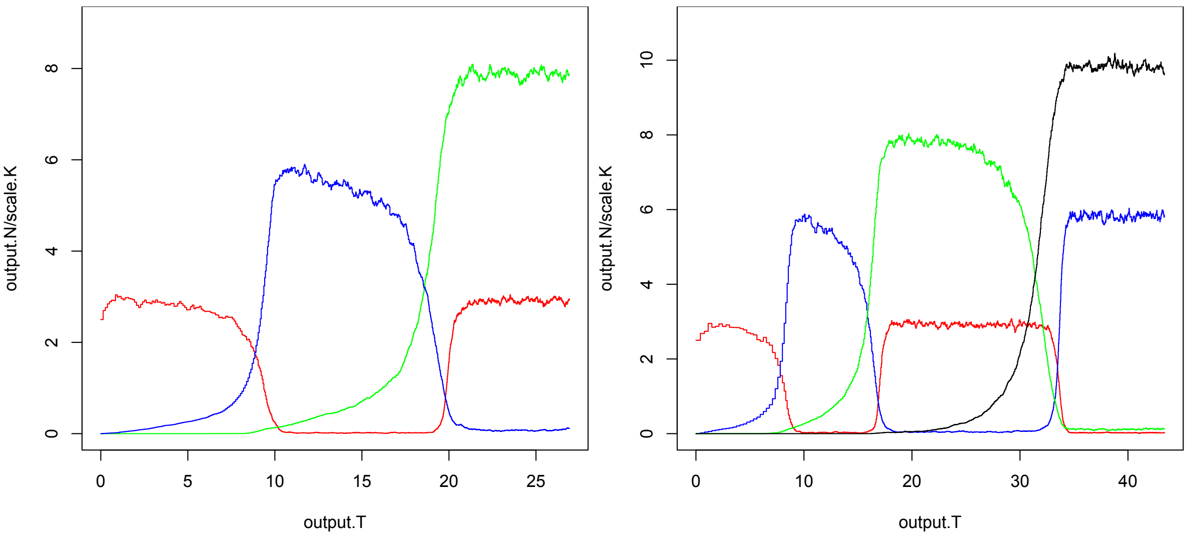

Figure 3.3: Simulations of a trait substitution tree model arising in Theorem 3.3.3 on a three- and four-type trait space.

Remark 3.3.4.

- *We illustrate the theorem by simulations (see Figure 3.3). We take all the same parameters as in Figure 3.2 except replacing* $\epsilon = K^{-\frac{4}{5}}$ *in the three-type case and* $\epsilon = K^{-\frac{3}{4}}$ *in the four-type case, initial population size* $K = 1000$*. Obviously, they both satisfy conditions* (3.3.4)*.*

- *The rare migration parameter region constrained by* (3.3.4) *is denoted by the upper right area* ***II*** *in Figure 3.1. As analyzed in Theorem 3.3.3, the fixation time scale is of order* $\ln\frac{1}{\epsilon}$*. The stable configuration for the three-type case is* $\Gamma^{(2)} = 3\delta_{x_0} + 8\delta_{x_2}$ *and it is* $\Gamma^{(3)} = 6\delta_{x_1} + 10\delta_{x_3}$ *for the four-type case. We will show in Theorem 3.4.2 that the TST process jumps from* $\Gamma^{(2)}$ *to* $\Gamma^{(3)}$ *on an even rarer mutation time scale of order* $\frac{1}{K\sigma}$ *(see Figure 3.4).*

3.4 Late time window with mutation as $K \to \infty$

Following the framework we build up in [6], in order to study the asymptotic behavior on an even longer time scale, we introduce another mutation mechanism into the population generated by (3.2.3). We now study the model with mutations formulated by the following generator supported on a compact set $\mathcal{X}$:

$$\begin{aligned} L^{K,\epsilon,\sigma}F(\nu) = &\int_{\mathcal{X}} \left[F(\nu+\frac{\delta_x}{K}) - F(\nu)\right] b(x)K\nu(dx) \\ &+ \int_{\mathcal{X}} \left[F(\nu-\frac{\delta_x}{K}) - F(\nu)\right] \left(d(x) + \int_{\mathcal{X}} \alpha(x,y)\nu(dy)\right) K\nu(dx) \\ &+ \epsilon \int_{\mathcal{X}}\int_{\mathcal{X}} \left[F(\nu+\frac{\delta_y}{K}-\frac{\delta_x}{K}) - F(\nu)\right] m(x,dy)1_{\{y\in\mathrm{supp}\{\nu\}\}}K\nu(dx) \\ &+ \sigma \int_{\mathcal{X}}\int_{\mathbb{R}^d} \left[F(\nu+\frac{\delta_{x+h}}{K}) - F(\nu)\right] \mu(x)p(x,dh)K\nu(dx). \end{aligned} \tag{3.4.1}$$

Here we denote the process by $X^{K,\epsilon,\sigma}_{\cdot}$ with one more superscript σ, distinguishing from the one without mutation in previous section.

Notice that the mutation kernel $p(x, dh)$ is used to introduce a new trait site to previous finite trait space and enlarge the supporting trait space by one each time there enters a mutant, whereas the migration kernel only acts on current support sites of the population. Later we will see, under some rare mutation constraint (with respect to migration rate), the dominating power for fixation is mainly from exponential growth of migration particles. Before proceeding towards the main theorem, we briefly give some assumptions and the definition of the trait substitution tree, which already appeared in [6].

Assumption (**C**).

(C1) For given distinct traits $\{x_0, x_1, \cdots, x_n\} \subset \mathcal{X}, n \in \mathbb{N}$, there exists a total order relation

$$x_{i_0} \prec x_{i_1} \prec \cdots \prec x_{i_{n-1}} \prec x_{i_n}, \tag{3.4.2}$$

where $x \prec y$ means that the fitness functions satisfy $\bar{f}(x, y) = b(x) - d(x) - \alpha(x, y)\bar{n}(y) < 0$, and $\bar{f}(y, x) = b(y) - d(y) - \alpha(y, x)\bar{n}(x) > 0$.

For simplicity of notation, we always assume $x^{(n)}_0 \prec x^{(n)}_1 \prec \cdots \prec x^{(n)}_n$ with $x^{(n)}_i = x_i$ for $0 \leq i \leq n$. By adding a new trait x whose fitness is between $x^{(n)}_j$ and $x^{(n)}_{j+1}$ for some $0 \leq j \leq n$, we relabel new traits as following

$$x^{(n+1)}_0 \prec x^{(n+1)}_1 \prec \cdots \prec x^{(n+1)}_n \prec x^{(n+1)}_{n+1}, \tag{3.4.3}$$

where $x^{(n+1)}_i = x^{(n)}_i$ for $0 \leq i \leq j$, $x^{(n+1)}_{j+1} = x$ and $x^{(n+1)}_i = x^{(n)}_{i-1}$ for $j+2 \leq i \leq n+1$.

(C2) Competition and migration only occurs among nearest neighbors, i.e. for totally ordered traits in (C1), we have $m(x^{(n)}_i, x^{(n)}_j) = \alpha(x^{(n)}_i, x^{(n)}_j) \equiv 0$ for $\mid i - j \mid > 1$.

Under above assumptions we can rewrite the generator (3.4.1) as following

$$\begin{aligned} L^{K,\epsilon,\sigma}F(\nu) = & \int_{\mathcal{X}} \left[F(\nu + \frac{\delta_x}{K}) - F(\nu) \right] b(x) K\nu(dx) \\ & + \int_{\mathcal{X}} \left[F(\nu - \frac{\delta_x}{K}) - F(\nu) \right] \left(d(x) + \int_{\mathcal{X}} \alpha(x,y) 1_{\{x^-,\,x,\,x^+\}} \nu(dy) \right) K\nu(dx) \\ & + \epsilon \int_{\mathcal{X}} \int_{\mathcal{X}} \left[F(\nu + \frac{\delta_y}{K} - \frac{\delta_x}{K}) - F(\nu) \right] 1_{\{x^-,\,x^+\}} m(x,dy) K\nu(dx) \\ & + \sigma \int_{\mathcal{X}} \int_{\mathbb{R}^d} \left[F(\nu + \frac{\delta_{x+h}}{K}) - F(\nu) \right] \mu(x) p(x,dh) K\nu(dx) \end{aligned} \tag{3.4.4}$$

where x^- and x^+, specified by the total order relation in assumption (C1), are elements in $\mathrm{supp}\{\nu\} \subset \mathcal{X}$ satisfying

$$x^- = \sup\{y \in \mathrm{supp}\{\nu\} : \bar{f}(y,x) < 0\}$$

and

$$x^+ = \inf\{y \in \mathrm{supp}\{\nu\} : \bar{f}(y,x) > 0\}.$$

On the migration time scale, there are a variety of different paths to approach the equilibrium configuration by giving different coefficient. However, the equilibrium configuration of a finite trait system is always the same up to the ordered sequence determined as in assumption (C1) and the time scale for convergence is always of order $O(\ln \frac{1}{\epsilon})$ as showed in Theorem 3.3.3.

Definition 3.4.1. *A Markov jump process $\{\Gamma_t : t \geq 0\}$ characterized as following is called a trait substitution tree (in short TST) with the ancestor $\Gamma_0 = \bar{n}(x_0)\delta_{x_0}$.*

(i) For any nonnegative integer l, it jumps from $\Gamma^{(2l)} := \sum_{i=0}^{l} \bar{n}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$ to $\Gamma^{(2l+1)}$ with transition rate $\bar{n}(x_{2k}^{(2l)})\mu(x_{2k}^{(2l)})p(x_{2k}^{(2l)}, dh)$ for any $0 \leq k \leq l$, where

- $\Gamma^{(2l+1)} = \sum_{i=1}^{j} \bar{n}(x_{2i-1}^{(2l)})\delta_{x_{2i-1}^{(2l)}} + \bar{n}(x_{2k}^{(2l)} + h)\delta_{x_{2k}^{(2l)}+h} + \sum_{i=j+1}^{l} \bar{n}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$
  if there exists $0 \leq j \leq l$ s.t. $x_{2j}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j+1}^{(2l)}$,

- $\Gamma^{(2l+1)} = \sum_{i=1}^{j} \bar{n}(x_{2i-1}^{(2l)})\delta_{x_{2i-1}^{(2l)}} + \sum_{i=j}^{l} \bar{n}(x_{2i}^{(2l)})\delta_{x_{2i}^{(2l)}}$
  if there exists $0 \leq j \leq l$ s.t. $x_{2j-1}^{(2l)} \prec x_{2k}^{(2l)} + h \prec x_{2j}^{(2l)}$.

Then, we relabel the traits according to the total order relation as in (C1):

$$x_0^{(2l+1)} \prec x_1^{(2l+1)} \prec \cdots \prec x_{2l}^{(2l+1)} \prec x_{2l+1}^{(2l+1)}, \tag{3.4.5}$$

where in associate with the first case

$$\begin{aligned} & x_i^{(2l+1)} := x_i^{(2l)} \text{ for } 0 \leq i \leq 2j, \qquad x_{2j+1}^{(2l+1)} := x_{2k}^{(2l)} + h, \\ & x_i^{(2l+1)} := x_{i-1}^{(2l)} \text{ for } 2j+2 \leq i \leq 2l+1, \end{aligned}$$

and in associate with the second case

$$x_i^{(2l+1)} := x_i^{(2l)} \text{ for } 0 \leq i \leq 2j-1, \quad x_{2j}^{(2l+1)} := x_{2k}^{(2l)} + h,$$
$$x_i^{(2l+1)} := x_{i-1}^{(2l)} \text{ for } 2j+1 \leq i \leq 2l+1.$$

(ii) *For nonnegative integer l, it jumps from $\Gamma^{(2l+1)} := \sum_{i=1}^{l+1} \bar{n}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$ to $\Gamma^{(2l+2)}$ with transition rate $\bar{n}(x_{2k-1}^{(2l+1)})\mu(x_{2k-1}^{(2l+1)})p(x_{2k-1}^{(2l+1)}, dh)$ for any $1 \leq k \leq l+1$, where*

- $\Gamma^{(2l+2)} = \sum_{i=1}^{j} \bar{n}(x_{2(i-1)}^{(2l+1)})\delta_{x_{2(i-1)}^{(2l+1)}} + \bar{n}(x_{2k-1}^{(2l+1)}+h)\delta_{x_{2k-1}^{(2l+1)}+h} + \sum_{i=j+1}^{l+1} \bar{n}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$
  if there exists $1 \leq j \leq l+1$ s.t. $x_{2j-1}^{(2l+1)} \prec x_{2k-1}^{(2l+1)} + h \prec x_{2j}^{(2l+1)}$,

- $\Gamma^{(2l+1)} = \sum_{i=1}^{j} \bar{n}(x_{2(i-1)}^{(2l+1)})\delta_{x_{2(i-1)}^{(2l+1)}} + \sum_{i=j}^{l+1} \bar{n}(x_{2i-1}^{(2l+1)})\delta_{x_{2i-1}^{(2l+1)}}$
  if there exists $1 \leq j \leq l+1$ s.t. $x_{2j-2}^{(2l+1)} \prec x_{2k-1}^{(2l+1)} + h \prec x_{2j-1}^{(2l+1)}$.

Then, we relabel the traits according to the total order relation as in (C1):

$$x_0^{(2l+2)} \prec x_1^{(2l+2)} \prec \cdots \prec x_{2l+1}^{(2l+2)} \prec x_{2l+2}^{(2l+2)}, \tag{3.4.6}$$

where in associate with the first case

$$x_i^{(2l+2)} := x_i^{(2l+1)} \text{ for } 0 \leq i \leq 2j-1, \quad x_{2j}^{(2l+2)} := x_{2k-1}^{(2l+1)} + h,$$
$$x_i^{(2l+2)} := x_{i-1}^{(2l+1)} \text{ for } 2j+1 \leq i \leq 2l+2,$$

and in associate with the second case

$$x_i^{(2l+2)} := x_i^{(2l+1)} \text{ for } 0 \leq i \leq 2j-2, \quad x_{2j-1}^{(2l+2)} := x_{2k-1}^{(2l+1)} + h,$$
$$x_i^{(2l+2)} := x_{i-1}^{(2l+1)} \text{ for } 2j \leq i \leq 2l+2.$$

Theorem 3.4.2. *Admit assumption **(A1)** and **(C)**. Consider the process $\{X_t^{K,\epsilon,\sigma}, t \geq 0\}$ described by the generator* (3.4.4). *Suppose that $X_0^{K,\epsilon,\sigma} = \frac{N_0^K}{K}\delta_{x_0}$ and $\frac{N_0^K}{K} \to \bar{n}(x_0)$ in law as $K \to \infty$. In addition to the condition* (3.3.4), *suppose it also holds that*

$$\ln\frac{1}{\epsilon} \ll \frac{1}{K\sigma} \ll e^{KC} \qquad \text{for any } C > 0. \tag{3.4.7}$$

Then $(X_{t/K\sigma}^{K,\epsilon,\sigma})_{t\geq 0}$ converges as $K \to \infty$ to the trait substitution tree $(\Gamma_t)_{t\geq 0}$ defined in Definition 3.4.1 in the sense of f.d.d. on $\mathcal{M}_F(\mathcal{X})$ equipped with the topology induced by mappings $\nu \mapsto \langle \nu, f\rangle$ with f a bounded measurable function on $\mathcal{X}$.

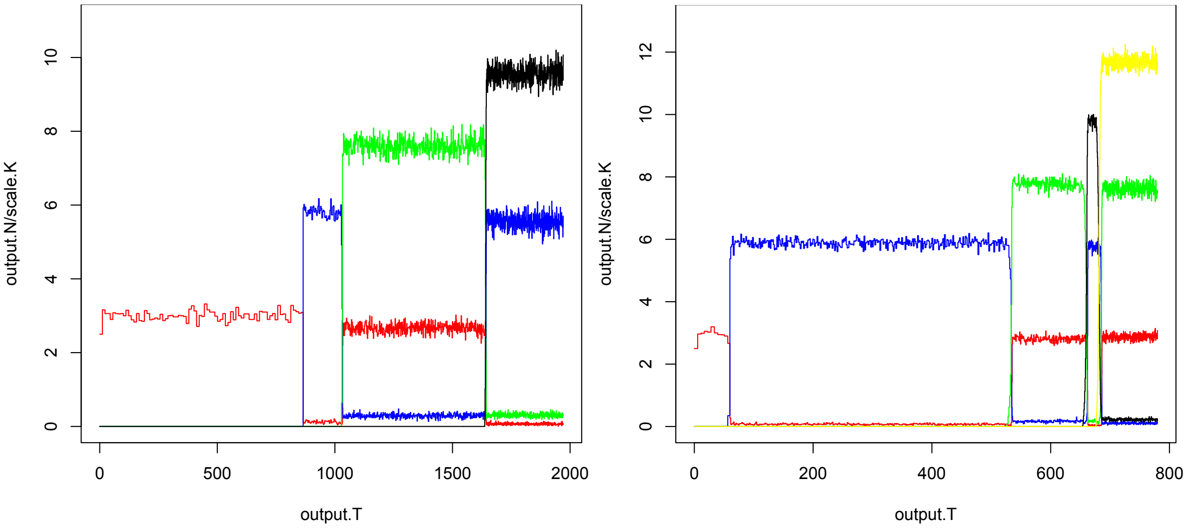

Figure 3.4: Simulations of a trait substitution tree on the mutation time scale arising in Theorem 3.4.2 on four- and five-type trait space.

Remark 3.4.3.

- *There are two time scales for the individual-based population, which can be observed from Theorem 3.3.3 and the generator* (3.4.4)*. One is the fixation time scale of order* $\ln \frac{1}{\epsilon}$ *while the other one is the mutation time scale of order* $\frac{1}{K\sigma}$*, which are constrained on LHS of the inequality* (3.4.7)*. By adopting the time scales separation technique used in [8], we can get a nice limiting structure-TST in the large population limit. The RHS of the inequality* (3.4.7) *is used to guarantee that system can not drift away from the TST equilibrium configuration on the mutation time scale (see Freidlin and Wentzell [23]).*

- *As showed in Figure 3.4, we simulate the trait substitution tree processes by introducing a mutation mechanism. Note that the simulation shows a special case where the population always reproduces a mutant which is more fitter than any of already existing traits. The birth rate of red-colored population is* 3*, while the blue one, the green one, the black one and the yellow one have birth rates* 6, 8, 10, 12 *resp.. Their death rates are constant* 0*. We take* $\epsilon = K^{-0.8}$ *and* $\sigma = K^{-1.5}$*, where initial scaling parameter* $K = 400$*. On a longer mutation time scale, the fixation process due to migration is not visible any more. However, if we zoom into the infinitesimal fixation period, we get the picture as in Figure 3.3.*

3.5 Outline of proofs

In order to illustrate the basis idea of proof, we start with a three-type toy model. But notice that our analysis is not reduced only to the three-trait case. All the machinery is still available for any finite-trait space, which will be shown later. However, the explicit proof are more difficult to write down without some restrictive conditions. That is why we impose assumption (**B**3) in Theorem 3.3.3.

Proposition 3.5.1. *Admit the same condition as in Theorem 3.3.3. Consider a sequence of processes on a trait space* $\mathcal{X} = \{x_0, x_1, x_2\}$*. Then, there exists a constant* $\bar{t}_2 > 0$*, such that for any* $t > \bar{t}_2$

$$\lim_{K\to\infty} X^K_{t\ln\frac{1}{\epsilon}} \overset{d}{=} \Gamma^{(2)} \tag{3.5.1}$$

under the total variation norm.

Proof. (see Figure 3.5).

Let $\xi^K_t(x_0) := \frac{N^K_t}{K}$ and $\xi^K_t(x_i) := \frac{N^{K,i}_t}{K} = \langle X^K_t, 1_{\{x_i\}}\rangle$ for $i = 1, 2$.

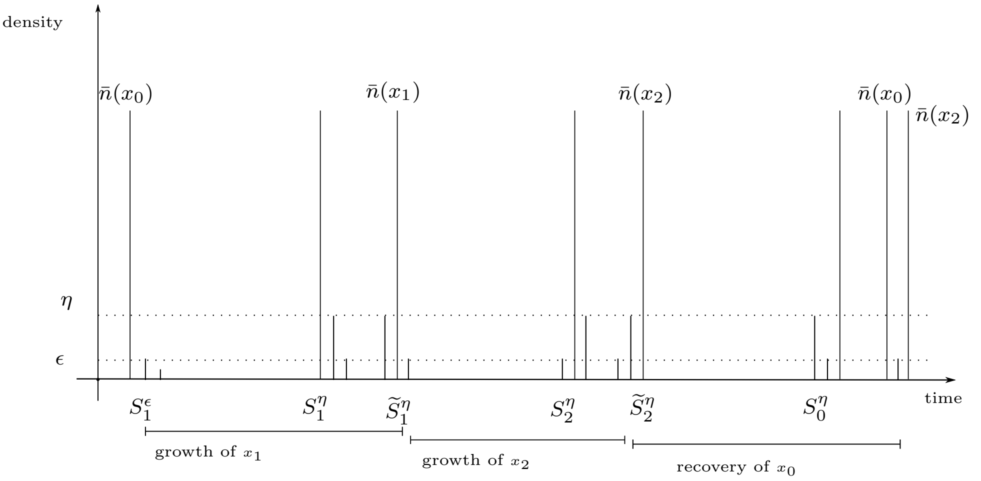

Figure 3.5: Phase evolution of mass bars in early time window on the three-trait site space

.

Step 1. Firstly, consider the emergence and growth of population at trait site x_1. Set $S^\epsilon_1 = \inf\{t > 0 : \xi^K_t(x_1) \geq \epsilon\}$. Thanks to $\frac{N^K_0}{K} \to n_0 > 0$ in law as $K \to \infty$ and by applying the law of large numbers of random processes (see Chap.11, Ethier and Kurtz 1986), one obtains from the last term in generator (3.2.3) that, for any $\delta > 0$, $T > 0$,

$$\lim_{K\to\infty} \mathbb{P}\left(\sup_{0\le t\le T}\left|\frac{\xi^K_t(x_1)}{\epsilon} - n_t(x_1)\right| < \delta\right) = 1$$

where $n_t(x_1)$ is governed by equation $\dot{n}(x_1) = m(x_0, x_1)n_0$ with initial $n_0(x_1) = 0$. Therefore,

$$\lim_{K\to\infty} \mathbb{P}\left(\frac{1}{m(x_0,x_1)n_0} - \delta < S^\epsilon_1 < \frac{1}{m(x_0,x_1)n_0} + \delta\right) = 1, \tag{3.5.2}$$

that is, S^ϵ_1 is of order 1.

For any $\eta > 0$, set $S^\eta_1 = \inf\{t : t > S^\epsilon_1,\ \xi^K_t(x_1) \geq \eta\}$. Consider a sequence of rescaled processes $\left(\frac{N^{K,1}_t}{K\epsilon}\right)_{t\ge S^\epsilon_1}$ with $\frac{N^{K,1}_{S^\epsilon_1}}{K\epsilon} = \frac{\xi^K_{S^\epsilon_1}(x_1)}{\epsilon} \to 1$ as $K \to \infty$. As before, by law of large numbers of random processes (see Chap.11 Ethier and Kurtz 1986), one obtains, for any $\delta > 0$, $T > 0$,

$$\lim_{K\to\infty} \mathbb{P}\left(\sup_{0\le t\le T}\left|\frac{N_t^{K,1}}{K\epsilon} - m_t\right| < \delta\right) = 1, \tag{3.5.3}$$

where m_t is governed by equation $\dot{m} = \bar{f}(x_1, x_0)m = (b(x_1) - d(x_1) - \alpha(x_1, x_0)\bar{n}(x_0))\, m$ with $m_0 = 1$.

Set $T_1^{\eta/\epsilon} = \inf\{t - S_1^\epsilon : t > S_1^\epsilon, \frac{N_t^{K,1}}{K\epsilon} \ge \eta/\epsilon\}$, and $t_1^{\eta/\epsilon} = \inf\{t > 0 : m_t \ge \eta/\epsilon\}$. Then, for any $\delta > 0$, there exists $\delta' > 0$ such that

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}\left(\left(\frac{1}{\bar{f}(x_1,x_0)} - \delta\right)\ln\frac{1}{\epsilon} < S_1^\eta - S_1^\epsilon < \left(\frac{1}{\bar{f}(x_1,x_0)} + \delta\right)\ln\frac{1}{\epsilon}\right)\\
=&\lim_{K\to\infty} \mathbb{P}\left(\left(\frac{1}{\bar{f}(x_1,x_0)} - \delta\right)\ln\frac{1}{\epsilon} < T_1^{\eta/\epsilon} < \left(\frac{1}{\bar{f}(x_1,x_0)} + \delta\right)\ln\frac{1}{\epsilon}\right)\\
=&\lim_{K\to\infty} \mathbb{P}\Bigg(\left(\frac{1}{\bar{f}(x_1,x_0)} - \frac{\delta}{2}\right)\ln\frac{1}{\epsilon} < t_1^{\eta/\epsilon} < \left(\frac{1}{\bar{f}(x_1,x_0)} + \frac{\delta}{2}\right)\ln\frac{1}{\epsilon},\\
&\qquad \sup_{0\le t\le t_1^{\eta/\epsilon}} \left|\frac{N_t^{K,1}}{K\epsilon} - m_t\right| < \delta'\Bigg)\\
&=1
\end{aligned} \tag{3.5.4}$$

where the last equal sign is due to (3.5.3).

After population of trait x_1 reaches some η threshold, the dynamics $\left(\xi_t^K(x_0), \xi_t^K(x_1)\right)$ can be approximated by the solution of a two-dimensional Lotka-Volterra equations. Then, it takes time of order 1 (mark this time coordinator by $\widetilde{S}_1^\eta$) for the two subpopulations switching their mass distribution and gets attracted into $\eta-$neighborhood of the stable equilibrium $(0, \bar{n}(x_1))$.

Step 2. Now consider the emerging and growth of population $\xi_t^K(x_2) := \langle X_t^K, 1_{\{x_2\}}\rangle$ at trait site x_2. Set $S_2^\epsilon = \inf\{t : t > \widetilde{S}_1^\eta, \xi_t^K(x_2) \ge \epsilon\}$. Similarly as is done for S_1^ϵ in (3.5.2), one can get that $\lim_{K\to\infty} \mathbb{P}(S_2^\epsilon - \widetilde{S}_1^\eta = O(1)) = 1$. On a longer time scale, we will not distinguish S_2^ϵ from $\widetilde{S}_1^\eta$.

Set $S_2^\eta = \inf\{t : t > S_2^\epsilon, \xi_t^K(x_2) \ge \eta\}$. One follows the same procedure to derive (3.5.4) and asserts that for any $\delta > 0$,

$$\lim_{K\to\infty} \mathbb{P}\left(\left(\frac{1}{\bar{f}(x_2,x_1)} - \delta\right)\ln\frac{1}{\epsilon} < S_2^\eta - \widetilde{S}_1^\eta < (\frac{1}{\bar{f}(x_2,x_1)} + \delta)\ln\frac{1}{\epsilon}\right) = 1. \tag{3.5.5}$$

Note that assumption (**B**3) $\frac{2}{b(x_2)-d(x_2)} \ge \frac{1}{\bar{f}(x_1,x_0)} + \frac{1}{\bar{f}(x_2,x_1)}$ guarantees that $\xi_t^K(x_2)$ can not grow so fast in exponential rate $b(x_2) - d(x_2)$ such that it reaches some η-level before S_2^η.

During time period $(\widetilde{S}_1^\eta, S_2^\eta)$, population at site x_0, on one hand, decreases due to the competition from more fitter trait x_1. On the other hand, it can not go below ϵ level due to the successive migration in a portion of ϵ from site x_1. More precisely, by neglecting migrant contribution, $\xi_t^K(x_0)$ converges $n_t(x_0)$ in probability as K tends to ∞, where

$$\dot{n}_t(x_0) = (b(x_0) - d(x_0) - \alpha(x_0, x_1)\bar{n}(x_1))\, n_t(x_0) = \bar{f}(x_0, x_1) n_t(x_0) \tag{3.5.6}$$

with $n_0(x_0) = \eta$. Let $\Delta S_2^\eta = S_2^\eta - \widetilde{S}_1^\eta$. Then, for any $\delta > 0$,

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}\left(\xi^K_{S_2^\eta}(x_0) \in (n_{\Delta S_2^\eta}(x_0) - \delta, n_{\Delta S_2^\eta}(x_0) + \delta)\right) \\
&= \lim_{K\to\infty} \mathbb{P}\left(\eta e^{\bar{f}(x_0,x_1)\Delta S_2^\eta} - \delta < \xi^K_{S_2^\eta}(x_0) < \eta e^{\bar{f}(x_0,x_1)\Delta S_2^\eta} + \delta\right) \\
&= \lim_{K\to\infty} \mathbb{P}\left(\eta \epsilon^{|\bar{f}(x_0,x_1)|/\bar{f}(x_2,x_1)} - \delta < \xi^K_{S_2^\eta}(x_0) < \eta \epsilon^{|\bar{f}(x_0,x_1)|/\bar{f}(x_2,x_1)} + \delta\right) \\
&= 1
\end{aligned} \tag{3.5.7}$$

where the second equality is due to (3.5.5). Taking the migration from site x_1 into account, we thus have

$$\lim_{K\to\infty} \mathbb{P}\left(\xi^K_{S_2^\eta}(x_0) = O(\epsilon^{|\bar{f}(x_0,x_1)|/\bar{f}(x_2,x_1)} \vee \epsilon)\right) = 1. \tag{3.5.8}$$

We proceed as before for $\widetilde{S}_1^\eta$ in step 1. After time S_2^η, the mass bars on dimorphic system $(\xi_t^K(x_1), \xi_t^K(x_2))$ can be approximated by ODEs and will be switched again in time of order 1 (marked by $\widetilde{S}_2^\eta$ as in Figure 3.5), and they are attracted into $\eta-$ neighborhood of $(0, \bar{n}(x_2))$. As for the population density on site x_0, one obtains from (3.5.8)

$$\lim_{K\to\infty} \left(\xi^K_{\widetilde{S}_2^\eta}(x_0) = O(\epsilon^{c_1})\right) = 1 \tag{3.5.9}$$

where $c_1 = \frac{|\bar{f}(x_0,x_1)|}{\bar{f}(x_2,x_1)} \wedge 1 \leq 1$.

Step 3. We now consider the recovery of subpopulation at trait site x_0. Recovery arises because of the lack of effective competitions from its neighbor site x_1, or under negligible competitions since the local population density on x_1 is very low under the control of its fitter neighbor x_2. Without lose of generality, we suppose $c_1 := \frac{|\bar{f}(x_0,x_1)|}{\bar{f}(x_2,x_1)} < 1$ in (3.5.8).

Set $S_0^\eta = \inf\{t : t > \widetilde{S}_2^\eta,\, \xi_t^K(x_0) \geq \eta\}$. We proceed as before in step 1. From (3.5.8), $\frac{\xi^K_{\widetilde{S}_2^\eta}(x_0)}{\epsilon^{c_1}}$ converges to some positive constant (say m_0) in probability as $K \to \infty$. Thus, by applying law of large numbers to the sequence of processes $\frac{N_t^K}{K\epsilon^{c_1}}$, for any $\delta > 0$, $T > 0$,

$$\lim_{K\to\infty} \mathbb{P}\left(\sup_{0\leq t\leq T} \left|\frac{\xi_t^K(x_0)}{\epsilon^{c_1}} - m_t\right| < \delta\right) = 1 \tag{3.5.10}$$

where m_t is governed by logistic equation $\dot{m} = (b(x_0) - d(x_0))\, m$ starting with a positive initial m_0.

Following the same way to obtain (3.5.4), time length $S_0^\eta - \widetilde{S}_2^\eta$ can be approximated by time needed for dynamics m to approach η/ϵ^{c_1} level, which is of order $\frac{c_1}{(b(x_0)-d(x_0))}\ln\frac{1}{\epsilon}$, i.e. for any $\delta > 0$,

$$\lim_{K\to\infty}\mathbb{P}\left(\left(\frac{c_1}{b(x_0)-d(x_0)}-\delta\right)\ln\frac{1}{\epsilon} < S_0^\eta - \widetilde{S}_2^\eta < \left(\frac{c_1}{b(x_0)-d(x_0)}+\delta\right)\ln\frac{1}{\epsilon}\right) = 1. \tag{3.5.11}$$

At the same time, $\xi_t^K(x_1)$ converges in probability to ψ_t which satisfies equation $\dot{\psi} = \bar{f}(x_1,x_2)\psi$ with $\psi_{\widetilde{S}_2^\eta} = \eta$. Then, we can justify the following estimate for population density at site x_1,

$$\lim_{K\to\infty}\mathbb{P}\left(\xi_{S_0^\eta}^K(x_1) = O(\epsilon^{c_2}\vee\epsilon)\right) = 1 \tag{3.5.12}$$

where $c_2 = \frac{c_1|\bar{f}(x_1,x_2)|}{b(x_0)-d(x_0)}$.

We now combine all these estimates (3.5.4),(3.5.5), (3.5.11) together, and conclude that

$$\lim_{K\to\infty}\mathbb{P}\left(\|X_{t\ln\frac{1}{\epsilon}}^K - \Gamma^{(2)}\| < \delta\right) = 1 \tag{3.5.13}$$

for $t > \bar{t}_2 := \frac{1}{\bar{f}(x_1,x_0)} + \frac{1}{\bar{f}(x_2,x_1)} + \frac{c_1}{b(x_0)-d(x_0)}$ under the total variation norm $\|\cdot\|$ on $\mathcal{M}_F(\mathcal{X})$.
□

Proof of Theorem 3.3.3. We proceed the proof by the induction method over the superscript $L \in \mathbb{N}$ of trait space $\mathcal{X}^{(L)} = \{x_0, x_1, \ldots, x_L\}$.

(1). When $L = 2$, it is already proved in Proposition 3.5.1 that there exists a constant $\bar{t}_2 > 0$ such that for any $t > \bar{t}_2$

$$\lim_{K\to\infty} X_{t\ln\frac{1}{\epsilon}}^K \overset{\text{(d)}}{=} \Gamma^{(2)} \tag{3.5.14}$$

under the total variation norm.

(2). Without loss of generality, suppose it holds that for any $L = 2l$ there exists a constant $\bar{t}_{2l}$ such that for any $t > \bar{t}_{2l}$

$$\lim_{K\to\infty} X_{t\ln\frac{1}{\epsilon}}^K \overset{\text{d}}{=} \Gamma^{(L)}. \tag{3.5.15}$$

We need to prove the same relation also holds for the case $L = 2l+1$.

We firstly consider the invasion time scale of population at site x_{2l+1}.

Denote by $\xi_t^K(x_{2l+1}) := \langle X_t^K, 1_{\{x_{2l+1}\}}\rangle$. If $K\epsilon^{2l+1} \ll 1$, it follows a similar proof as in Proposition 3.5.1. So, now we only need to consider the case when $K\epsilon^{2l+1} \gg 1$, that is, the mass at site x_{2l+1} is large in the very beginning. In fact, since $\frac{N_0^K}{K} \to n_0$ in law as $K \to \infty$ and the nearest-neighbor mass migrates from site x_0 to site x_{2l+1} by passing through $x_1, \ldots, x_{2l}$, one applies the law of large numbers for random processes and obtains that

$$\lim_{K\to\infty}\mathbb{P}\left(\sup_{0\le t\le T}\left|\xi_t^K(x_{2l+1}) - n_t(x_{2l+1})\right| < \delta\right) = 1 \tag{3.5.16}$$

where $n_t(x_{2l+1})$ satisfies the equation $\dot{n}_t(x_{2l+1}) = \epsilon^{2l+1} \prod_{j=1}^{2l+1} m(x_j, x_{j-1}) n_0$. So, it takes time of order 1 for $\xi_t^K(x_{2l+1})$ to reach ϵ^{2l+1} level (mark the time coordinator by S^{ϵ}_{2l+1}).

Set $S^{\eta}_{2l+1} = \inf\{t : t > S^{\epsilon}_{2l+1}, \xi^K_t(x_{2l+1}) \geq \eta\}$. For $t \in \left(S^{\epsilon}_{2l+1}, S^{\eta}_{2l+1}\right)$, again by law of large numbers, $\frac{\xi_t^K(x_{2l+1})}{\epsilon^{2l+1}}$ converges to ϕ_t which satisfies $\phi_0 = 1$ and

$$\dot{\phi} = (b(x_{2l+1}) - d(x_{2l+1}))\phi. \tag{3.5.17}$$

Thus, $\Delta S^{\eta}_{2l+1} := S^{\eta}_{2l+1} - S^{\epsilon}_{2l+1}$ can be approximated by the time length (say Δt_{2l+1}) needed for dynamics ϕ to reach η/ϵ^{2l+1} level, i.e.

$$\begin{aligned} &\lim_{K\to\infty} \mathbb{P}\left(\left(\frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} - \delta\right) \ln\frac{1}{\epsilon} < \Delta S^{\eta}_{2l+1} < \left(\frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} + \delta\right) \ln\frac{1}{\epsilon}\right) \\ &\lim_{K\to\infty} \mathbb{P}\left(\left(\frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} - \delta\right) \ln\frac{1}{\epsilon} < \Delta t_{2l+1} < \left(\frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} + \delta\right) \ln\frac{1}{\epsilon}\right) \\ &= 1. \end{aligned} \tag{3.5.18}$$

We inherit the notation S^{η}_{2l} as the hitting time of η-level for the population at site x_{2l}. Due to the hypothesis for $L = 2l$ case, we know that S^{η}_{2l} is of order

$$\left(\left[\bar{f}(x_{2l}, x_{2l-1})\right]^{-1} + \ldots + \left[\bar{f}(x_1, x_0)\right]^{-1}\right) \ln\frac{1}{\epsilon}. \tag{3.5.19}$$

Thanks to assumption (**B**3), i.e.

$$\frac{2l+1}{b(x_{2l+1}) - d(x_{2l+1})} > \frac{1}{\bar{f}(x_{2l}, x_{2l-1})} + \ldots + \frac{1}{\bar{f}(x_1, x_0)}, \tag{3.5.20}$$

it implies that before time S^{η}_{2l}, population at site x_{2l+1} is still under negligible level (of order ϵ^c for some positive constant c) and can not influence the invasion process up to x_{2l}.

Following a similar procedure as deriving (3.5.19) (see Figure 3.5) to analyze the colonization of population at site x_{2l+1} due to migration from site x_{2l} with exponential rate $\bar{f}(x_{2l+1}, x_{2l})$, one obtains that S^{η}_{2l+1} should be of order

$$\left(\left[\bar{f}(x_{2l+1}, x_{2l})\right]^{-1} + \ldots + \left[\bar{f}(x_1, x_0)\right]^{-1}\right) \ln\frac{1}{\epsilon}. \tag{3.5.21}$$

Comparing two time scale estimates (3.5.18) and (3.5.21) for S^{η}_{2l+1} under assumption (**B**3), one gets (3.5.21) is the right one for the fixation of population at site x_{2l+1}.

Now we consider the total recovery time by summing up recovery time of all subpopulation on every second site backwards from x_{2l+1} to x_0, one can do calculations repeatedly as in Step 3 of the proof for Proposition 3.5.1. More precisely, for $1 \leq i \leq l$, the initial population $\xi^K(x_{2i-1})$ on site x_{2i-1} which is prepared for recovering is no less than ϵ-level due to the consistent migration from its fitter neighbor site x_{2i}. On the other hand, it grows exponentially at least with a rate $\bar{f}(x_{2i-1}, x_{2i-2}) = b(x_{2i-1}) - d(x_{2i-1}) - \alpha(x_{2i-1}, x_{2i-2})\bar{n}(x_{2i-2})$ due to the possibly strongest competition from its unfit neighbor

x_{2i-2}. In all, the recovery time (mark by $S^{\eta,2}_{2i-1}$) for the population $\xi^K(x_{2i-1})$ to reach η-level can be bounded from above

$$\lim_{K\to\infty} \mathbb{P}\left(S^{\eta,2}_{2i-1} < \left(\frac{1}{\bar{f}(x_{2i-1}, x_{2i-2})} + \delta \right) \ln \frac{1}{\epsilon} \right) = 1. \tag{3.5.22}$$

We now combine both time estimates (3.5.21) and (3.5.22). Let

$$\bar{t}_{2l+1} := 2\left(\left[\bar{f}(x_{2l+1}, x_{2l})\right]^{-1} + \ldots + \left[\bar{f}(x_1, x_0)\right]^{-1} \right). \tag{3.5.23}$$

Then, one can conclude that for any $t > \bar{t}_{2l+1}$, for any $\delta > 0$ and $0 \leq i \leq l$,

$$\begin{aligned} &\lim_{K\to\infty} \mathbb{P}\left(\left| \langle X^K_{t\ln\frac{1}{\epsilon}}, 1_{\{x_{2i+1}\}} \rangle - \bar{n}(x_{2i+1}) \right| < \delta \right) = 1, \\ &\lim_{K\to\infty} \mathbb{P}\left(\langle X^K_{t\ln\frac{1}{\epsilon}}, 1_{\{x_{2i}\}} \rangle < \delta \right) = 1. \end{aligned} \tag{3.5.24}$$

It follows the conclusion for any $t > \bar{t}_{2l+1}$,

$$\lim_{K\to\infty} X^K_{t\ln\frac{1}{\epsilon}} \overset{\mathrm{d}}{=} \Gamma^{(2l+1)}. \tag{3.5.25}$$

□

Proof of Theorem 3.4.2. The proof of this result is similar to the proof of [8, Theorem 1]. We will not repeat all the details and only focus more on supporting lemmas which are cornerstones of the proof.

For any $\varepsilon > 0$, $t > 0$, $L \in \mathbb{N}$, $B \subset \mathcal{X}$ measurable, take the integer part $L_1 := \left\lfloor \frac{L+2}{2} \right\rfloor$ and denote by

$$\begin{aligned} A^{K,\epsilon,\sigma}(\varepsilon, t, L, B) := \Big\{ &\mathrm{Supp}(X^{K,\epsilon,\sigma}_{\frac{t}{K\sigma}}) \text{ has } L+1 \text{ elements, and } L_1 \text{ out of them, say } \{x_1, \ldots, \\ &x_{L_1}\} \subset B, \text{ satisfy } \left| \langle X^{K,\epsilon,\sigma}_{\frac{t}{K\sigma}}, 1_{\{x_i\}} \rangle - \bar{n}(x_i) \right| < \varepsilon,\ 1 \leq i \leq L_1, \\ &\text{and the other } L+1-L_1 \text{ traits, say } y_1, \ldots, y_{L-L_1}, \text{ satisfy} \\ &\langle X^{K,\epsilon,\sigma}_{\frac{t}{K\sigma}}, 1_{\{y_j\}} \rangle < \varepsilon, 1 \leq j \leq L+1-L_1 \Big\}. \end{aligned} \tag{3.5.26}$$

To the end, it is enough to establish that

$$\lim_{K\to\infty} \mathbb{P}\left(A^{K,\epsilon,\sigma}(\varepsilon, t, L, B) \right) = \mathbb{P}\left(\mathrm{Supp}(\Gamma_t) \subset B \text{ and has } L_1 \text{ elements } \right) \tag{3.5.27}$$

where $(\Gamma_t)_{t\geq 0}$ is defined in Definition 3.4.1.

The first key ingredient of the proof is the characterization of exponentially distributed waiting time of every mutation event. It can be proved from the expression of the generator (3.4.4).

Lemma 3.5.2. *Assume that* $X_0^{K,\epsilon,\sigma} = \Gamma^{(L)}$, *w.o.l., take* $L = 2l$. *Let* τ *be the first mutation time after 0. Then,*

$$\lim_{K\to\infty} \mathbb{P}\left(\tau > \frac{t}{K\sigma}\right) = \exp\left(-t\sum_{i=0}^{l} \bar{n}(x_{2i}^{(2l)})\mu(x_{2i}^{(2l)})\right). \tag{3.5.28}$$

$$\lim_{K\to\infty} \mathbb{P}\left(\textit{at time}\,\tau,\ \textit{mutant comes from trait}\ x_{2k}^{(2l)}\right) = \frac{\bar{n}(x_{2k}^{(2l)})\mu(x_{2k}^{(2l)})}{\sum_{i=0}^{l} \bar{n}(x_{2i}^{(2l)})\mu(x_{2i}^{(2l)})}. \tag{3.5.29}$$

The second ingredient can been seen as a corollary of Theorem 3.3.3. It demonstrates that fixation of new configuration takes time of order $\ln\frac{1}{\epsilon}$, which is invisible on the mutation time scale.

Lemma 3.5.3. *Assume that* $X_0^{K,\epsilon,\sigma} = \Gamma^{(2l)} + \frac{1}{K}\delta_{x_{2k}^{(2l)}+h}$ *for some* $0 \le k \le l$. *Then there exists a constant* $C > 0$, *for any* $\delta > 0$, *such that*

$$\lim_{K\to\infty} \mathbb{P}\left(\tau > C\ln\frac{1}{\epsilon},\ \sup_{t\in(C\ln\frac{1}{\epsilon},\tau)} \|X_t^{K,\epsilon,\sigma} - \Gamma^{(2l+1)}\| < \delta\right) = 1 \tag{3.5.30}$$

where $\Gamma^{(2l+1)}$ *is defined as in Definition 3.4.2 (i) and* $\|\cdot\|$ *is the total variation distance.*

The proof of this lemma is similar to the one of Lemma 2.5.6.

Thus we conclude the proof of the Theorem. □

4 Trait substitution sequence with nearly neutral mutations

We consider a spatial population model with mutation and competition undergoing an acceleration of branching rates by K^{λ}, $\lambda \in (0,1]$, where K is proportional to the initial population size. By time scales separation constraints, we obtain pure jump processes in the large population and rare mutation limit for both $0 < \lambda < 1$ and $\lambda = 1$ cases. The novelty here is that the fixation probability of the advantageous mutant is nearly neutral and proportional to $\frac{1}{K^{\lambda}}$. Therefore, we rescale the population process on an accumulation of the mutation time scale. For $0 < \lambda < 1$ it generalizes the trait substitution sequence (TSS) studied in [8] ($\lambda = 0$), where the population density jumps from one stable equilibrium of some ODE to another. For $\lambda = 1$ we need some results on branching processes in random environment to analyze the fixation period, and employ quasi-stationary theory to predict the equilibrium density of the fixed type.

4.1 Introduction

In recent years a locally regulated population model proposed by Bolker and Pacala [4] has attracted particular interest both from biologists and mathematicians. To our knowledge, Fournier and Méléard [22], using the idea of interacting branching random walks, formulated a pathwise construction of the model in terms of Poisson random measures. In parallel, Etheridge [19] studied the extinction and survival problem of this model and considered the coexistence problem of derivative models in the following papers (see [2]). In particular, for the one dimensional version (without mutation), Lambert formulated it as logistic branching processes and studied the probabilistic and analytic properties in both discrete and continuous setting (see [7, 33]).

In this paper we are interested in the long time behavior of the locally regulated (density-dependent) populations. In particular, to figure out how the mutation strength influences the evolutionary fixation makes more sense. Mutations that occur in the resident population usually can be classified into three categories (w.r.t. resident population): advantageous, neutral, and deleterious. Respectively, they correspond to the supercritical, critical and subcritical cases in the branching process setting. The advantageous mutation makes the subpopulation with its trait fixed in a positive probability while the subpopulation from deleterious mutation dies out with probability 1; the fate of the neutral mutants is still unclear for us. It is of particular interest for us to consider the fate of neutral mutants. Now we explain by a few more words how different mutations influence the fixation of populations in a specific way.

Heuristically, it is usually guaranteed that a single mutant arises in the resident population and no further mutations occur until the fate of that mutant population is known. So it is enough to just consider a two-type system. On the one hand, competition for limited resource can eliminate one type under some non-coexistence assumption. On the other hand, the logistic growth term may impede the fixed type to be stabilized near its equilibrium on a long time scale. The biologically motivated assumption of rare mutations guarantees that, on the mutation time scale, the width of time intervals during which the population is polymorphic vanishes, so that the population appears monomorphic at each time. This is the so called "trait substitution sequence model" proposed by Metz et al. [39] and mathematically justified by Champagnat [8], Champagnat and Lambert [9]. In their arguments, the fixation probability of a selective mutation is a strictly positive constant (independent of the population size K). More precisely, the fixed type is strictly advantageous in [8] while the deleterious one can also be fixed but with a random density in [9].

The purpose of this paper is to consider the case when the fixation probability of the new mutant is nearly neutral (slightly advantageous). This idea can be realized by accelerating the branching rate (birth rate and death rate) in a suitable way, i.e., take individual birth rate $b_K = K^\lambda r + b$ and death rate $d_K = K^\lambda r + d$ for $0 < \lambda \leq 1$, where the parameter K is proportional to the initial population size. This procedure is highly reminiscent of the scaling method from branching particle systems to superprocesses (see [15, 37]). To prevent the population from rapid extinction, one also has to rescale the population size, so that making assumption of large population. We will show that, on a single mutation time scale, the fixation probability of the advantageous type is of order $\frac{1}{K^\lambda}$, which converges to 0 under the large population limit. Obviously, it is not the proper time scale to rescale the population process.

For $0 < \lambda < 1$ the large population limit of the population process (branching particle system) is a deterministic measure-valued process. In particular, it satisfies an ODE which has a stable equilibrium (carrying capacity) for one-type population without mutation. To find the suitable time scale to separate successively arising mutations, our argument proceeds by a way of intermediate approximations based on the trait substitute sequence model obtained in Champagnat [8]. Eventually, we rejustify the TSS model on an accumulation of the mutation time scale. In other words, the invasion is not due to the absolute advantage of a mutant type, but due to the relatively often coming neutral mutants.

For $\lambda = 1$ the large population limit is a superprocess with a quadratic competition term. The randomness is generated due to the accelerating birth and death events and usually called "demographic stochasticity". In particular, as for the one dimensional case, the rescaled logistic branching process converges to the so called logistic Feller diffusion. As shown in [33], the diffusion will be absorbed at 0 with probability 1. Thus, to capture the long time behavior in some sense, we study the process conditioned on never attaining 0. This conditioned process , named Q-process, is studied in [7, 36], which can be realized by compensating proper immigration onto the system (see [34]). As for the fixation period, we use a classical result about the extinction probability for branching processes in random environments (see Kaplan [29]) to build up the non-coexistence condition, and further to give an implicit expression for the fixation probability. Nevertheless, we can formulate the rescaling limit processes for both $0 < \lambda < 1$ and $\lambda = 1$ in an unified regime in the context of random environments (see Remark 4.2.3).

The remainder of the paper is structured as follows. In Section 4.2, we present the description of the models and state the main results. In Section 4.3, we get the large population limit of the interacting branching particle system for both $0 < \lambda < 1$ and $\lambda = 1$, and study some properties for one and two dimensional cases. In Section 4.4, we derive the trait substitution sequence based an intermediate-approximation result for the case $0 < \lambda < 1$. In Section 4.5, for $\lambda = 1$ the conditioned trait substitution sequence is proved by employing the Q-process theory and branching processes in random environments, which are used to characterize the long time behavior and provide the non-coexistence criterion.

4.2 Model and main results

4.2.1 Locally regulated spatial population model

We assume the population at time t is composed of a finite number I_t of individuals characterized by their phenotypic traits $x_1(t), \ldots, x_{I_t}(t)$ belonging a compact subset $\mathcal{X}$ of $\mathbb{R}^d$. We denote by $\mathcal{M}_F(\mathcal{X})$ the set of non-negative finite measures on $\mathcal{X}$. Let $\mathcal{M}(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$ be the set of counting measures on $\mathcal{X}$:

$$\mathcal{M}(\mathcal{X}) = \left\{ \sum_{i=1}^{n} \delta_{x_i} : x_1, \ldots, x_n \in \mathcal{X}, n \in \mathbb{N} \right\}.$$

Then, the population process at time t can be represented as:

$$\nu_t = \sum_{i=1}^{I_t} \delta_{X_i(t)}.$$

Let $B(\mathcal{X})$ denote the totality of functions on $\mathcal{X}$ that are bounded measurable. For any $f \in B(\mathcal{X})$, $\nu \in \mathcal{M}_F(\mathcal{X})$, we use notation $\langle \nu, f \rangle = \int f d\nu$.

Let's specify the population process $(\nu_t)_{t>0}$ by introducing a sequence of biological parameters:

- $b(x)$ is the birth rate from an individual with trait x.
- $d(x)$ is the death rate of an individual with trait x because of "aging".
- $\alpha(x, y)$ is the competition kernel felt by some individual with trait x from another individual with trait y.
- $\mu(x)$ is the mutation probability of an individual with trait x.

- $m(x, dh)$ is the law of trait difference $h = y - x$ between a mutant individual with trait y born from an individual with trait x. Since the mutant trait $y = x + h$ must belong to $\mathcal{X}$, this law has its support in $\mathcal{X} - x := \{y - x : y \in \mathcal{X}\} \subset \mathbb{R}^d$. We assume that $m(x, dh)$ has a density on $\mathbb{R}^d$ which is uniformly bounded in $x \in \mathcal{X}$ by some integrable function $\bar{m}(h)$.

To specify the model, the infinitesimal generator of the $\mathcal{M}(\mathcal{X})$-valued process is given as follows, for any $\phi \in B(\mathcal{M}(\mathcal{X}))$:

$$\begin{aligned} L\phi(\nu) = &\sum_{i=1}^{I} \left[\phi(\nu + \delta_{x_i}) - \phi(\nu)\right] (1 - \mu(x_i))b(x_i) \\ &+ \sum_{i=1}^{I} \int_{\mathcal{X} - x_i} \left[\phi(\nu + \delta_{x_i + h}) - \phi(\nu)\right] \mu(x_i)b(x_i)m(x_i, dh) \\ &+ \sum_{i=1}^{I} \left[\phi(\nu - \delta_{x_i}) - \phi(\nu)\right] (d(x_i) + \sum_{j \neq i}^{I} \alpha(x_i, x_j)). \end{aligned}$$

The first term above describes the clonal reproduction without mutation. The second term describes the mutant offspring with trait $x_i + h$ from mother with trait x_i. And the last term describes death of an individual x_i either due to aging or competition from another individual x_j.

By introducing a parameter $K \in \mathbb{N}$, we rescale the population size by K. We will show later, as K tends to infinity, we can get different large population limits by well-chosen renormalization. Furthermore, the population process can be parametrized by another parameter u_K governing the strength of mutation probability $\mu(\cdot)$ w.r.t. population size scaling parameter K.

Given a sequence of processes denoted by $(\nu_t^K)_{t \geq 0}$ with an accelerating exponent $\lambda \in (0, 1]$, assume the corresponding parameters have the following form, for $K \in \mathbb{N}$:

- $b_K(x) = K^\lambda r(x) + b(x)$.
- $d_K(x) = K^\lambda r(x) + d(x)$.
- $\alpha_K(x, y) = \alpha(x, y)/K$.

Notice that, we rescale the competition kernel by K as above, so that the system mathematically makes sense when we take large population limit. Biologically, K can be interpreted as scaling the resource or area available.

Instead of studying the above processes $(\nu_t^K)_{t \geq 0}$, it is more convenient to consider a sequence of rescaled measure-valued processes:

$$X_t^K := \frac{1}{K} \sum_{i=1}^{I_t^K} \delta_{x_i} \tag{4.2.1}$$

where X_t^K is in the space $\mathcal{M}_F(\mathbb{R}^d)$ comprising of finite measures.

The infinitesimal generator of the Markov process $(X_t^K)_{t\geq 0}$ can be expressed as follows:

$$\begin{aligned}L^K\phi(\nu) &= \int_{\mathcal{X}} \left[\phi(\nu+\frac{\delta_x}{K})-\phi(\nu)\right](1-u_K\mu(x))b_K(x)K\nu(dx)\\ &+\int_{\mathcal{X}}\int_{\mathbb{R}^d\setminus\{0\}}\left[\phi(\nu+\frac{\delta_{x+h}}{K})-\phi(\nu)\right]u_K\mu(x)b_K(x)m(x,dh)K\nu(dx) \qquad (4.2.2)\\ &+\int_{\mathcal{X}}\left[\phi(\nu-\frac{\delta_x}{K})-\phi(\nu)\right]\left(d_K(x)+\int_{\mathcal{X}}\alpha(x,y)\nu(dy)\right)K\nu(dx).\end{aligned}$$

Let us denote by (**A**) the following assumptions.

(A1) $\exists\ \bar{b},\ \bar{d},\ \bar{r},\ \underline{\alpha}\ ,\bar{\alpha}$, such that $0<b(x)\leq\bar{b},\ 0<d(x)\leq\bar{d},\ 0<r(x)\leq\bar{r},\ 0<\underline{\alpha}\leq\alpha(x,y)\leq\bar{\alpha}$, and $b(x)-d(x)>0,\ \forall x\in\mathcal{X}$.

(A2) $\forall\ x,y\in\mathcal{X},\ \bar{f}(x,y)\cdot\bar{f}(y,x)<0$, where

$\bar{f}(y,x)=b(y)-d(y)-\alpha(y,x)\bar{n}(x)$ and $\bar{n}(x)=\frac{b(x)-d(x)}{\alpha(x,x)}$,

$\bar{f}(x,y)=b(x)-d(x)-\alpha(x,y)\bar{n}(y)$ and $\bar{n}(y)=\frac{b(y)-d(y)}{\alpha(y,y)}$.

(A3) $\forall\ x,y\in\mathcal{X},\ Ef(x,y)\cdot Ef(y,x)<0$, where e.g. $f(y,x)=b(y)-d(y)-\alpha(y,x)n(x)$, and random variable $n(x)$ has the distribution $\pi^{(x)}$ as in Remark 4.3.11. Fitness function $f(x,y)$ takes the symmetric form.

For fixed $K\in\mathbb{N}$, assumption (A1) guarantees that the process with the infinitesimal generator (4.2.2) is well defined (refer to [22]). Assumptions (A2) and (A3) give the non-coexistence condition for $0<\lambda<1$ and $\lambda=1$, resp..

4.2.2 Statement of main results

Theorem 4.2.1 ($0<\lambda<1$)**.** *Admit assumptions (**A**1) and (**A**2). Suppose that* $X_0^K=\frac{N_0^K}{K}\delta_x$ *and are such that* $\frac{N_0^K}{K}\overset{law}{\to}n_0>0$*, as* $K\to+\infty$*, and* $\exp\{-CK^{1-\lambda}\}\ll u_K\ll\dfrac{1}{K^{1+\lambda}\ln K}$*,* $\forall\ C>0$*. Then,* $(X_{t/Ku_K}^K,t\geq 0)$ *converges in the sense of f.d.d to*

$$Y_t=\begin{cases}n_0\delta_x, & t=0\\ \bar{n}(\eta_t)\delta_{\eta_t}, & t>0\end{cases}$$

where the Markov jump process $(\eta_t,t\geq 0)$ *satisfies* $\eta_0=x$ *with an infinitesimal generator:*

$$A\varphi(x)=\int_{\mathbb{R}^d}(\varphi(x+h)-\varphi(x))\,\mu(x)\bar{n}(x)\frac{r(x)}{r(x+h)}[\bar{f}(x+h,x)]_+m(x,dh). \qquad (4.2.3)$$

Definition 4.2.2 ($\lambda=1$)**.** *For any* $t\geq 0$*, define a measure-valued Markov jump process* $V_t=n(\eta_t)\delta_{\eta_t}$ *with singular support as following* *(i)* $(\eta_t, t \geq 0)$ *is a Markov jump process on* $\mathcal{X}$ *with initial value* $\eta_0 = x$ *and the transition density from* x *to* $x + h$ *given by*

$$q(x, dh) = \gamma(x)\mathbb{E}\left[\frac{g_x(\xi(x), h)}{g_x(\xi(x))}\right] m(x, dh), \tag{4.2.4}$$

where

$$\gamma(x) = \mu(x)r(x)\mathbb{E}\left[n(x)g_x(n(x))\right] \tag{4.2.5}$$

and

$$\mathbb{P}\left(\xi(x) \in dz\right) = \frac{zg_x(z)\pi^{(x)}(dz)}{\mathbb{E}\left[n(x)g_x(n(x))\right]} \tag{4.2.6}$$

and

$$g_x(z) = \int_{\mathbb{R}^d} g_x(z, h)m(x, dh) \tag{4.2.7}$$

with

$$g_x(z, h) = 1_{\{Ef(x+h,x)>0\}} \int_{C([0,+\infty),\mathbb{R}_+^*)} Q_z^{(x)}(d\omega)$$

$$\left(r(x+h)\int_0^\infty \exp\left\{-\left(tb(x+h) - td(x+h) - \int_0^t \alpha(x+h, x)\omega_\tau(x)d\tau\right)\right\}dt\right)^{-1}.$$

i.e. the infinitesimal generator of $(\eta_t, t \geq 0)$ *has the form:*

$$Af(x) = \int_{\mathbb{R}^d\setminus\{0\}} \left(f(x+h) - f(x)\right) q(x, dh),$$

where $\mathbb{Q}_z^{(x)}$ *is a law on path space* $C([0,\infty), \mathbb{R}_+^*)$ *such that* $\omega_0(x) = z$ *and* $\pi^{(x)}$ *is the law of random variable* $n(x)$ *on* $\mathbb{R}_+^*$*. They are both defined in Remark 4.3.11.*

(ii) Conditioned on $(\eta_{t_1}, \ldots, \eta_{t_n}) = (x_1, \ldots, x_n)$ *for any* $0 < t_1 < \ldots < t_n,$*, all* $n(\eta_{t_i})$ *are independent and distributed as* $\pi^{(x_i)}$*, respectively.*

Observe that $\gamma(x)$ can be seen as the mean successful mutant production rate of a stationary (in some sense) x-type population and that $\mathbb{E}\left[\frac{g_x(\xi(x),h)m(x,dh)}{g_x(\xi(x))}\right]$ is the probability density of fixation of a successful mutant $x+h$-type population entering a pure (resident) x-type population with a biased stationary size $\xi(x)$.

Remark 4.2.3. *(1) When* $0 < \lambda < 1$*, the TSS limit is a measure-valued Markov jump process with singular support and the corresponding density is determined (non-random) by its trait. When* $\lambda = 1$*, which will be shown, the TSS limit is still a Markov jump process as in Definition 4.2.2 but with some random density.*

(2) In fact, the above two TSS limits are consistent. The first result can be embedded into the second one by taking $\pi^{(x)} = \delta_{\bar{n}(x)}$ and $\mathbb{Q}_z^{(x)}(d\omega) = \delta_{\bar{n}(x)}$ in the formula of transition rate $q(x, dh)$. In this degenerate case, $\omega_t(x) \equiv \bar{n}(x)$, $t \geq 0$ and $n(x) \equiv \bar{n}(x)$. Thus, we can compute the transition rate explicitly,

$$\begin{aligned} q(x, dh) &= \frac{\mu(x) r(x) 1_{\{Ef(x+h,x)>0\}}}{r(x+h)} \bar{n}(x) \\ &\quad \cdot \left(\int_0^\infty \exp\left\{ -(tb(x+h) - td(x+h) - \int_0^t \alpha(x+h, x)\bar{n}(x) d\tau) \right\} dt \right)^{-1} \\ &= \frac{\mu(x) r(x) \bar{n}(x) 1_{\{\bar{f}(x+h,x)>0\}}}{r(x+h)} \left(\int_0^\infty \exp\left\{ -t\bar{f}(x+h, x) \right\} dt \right)^{-1} m(x, dh) \\ &= \frac{\mu(x) r(x) \bar{n}(x)}{r(x+h)} [\bar{f}(x+h, x)]_+ m(x, dh). \end{aligned} \tag{4.2.8}$$

4.3 Auxiliary results on convergence

4.3.1 Convergence to a deterministic flow for $0 < \lambda < 1$

Firstly, we give the following martingale properties for $(X_t^K)_{t \geq 0}$, which will take a key role in the proof of weak convergence results. As a convention, we inherit all the notations and parameters in Section 4.2.

Proposition 4.3.1. *Admit assumption ($\boldsymbol{A}1$) and for some $p \geq 2$, $\sup_K \mathbb{E}[\langle X_0^K, 1 \rangle]^p < \infty$. Then, for any bounded measurable function f, the process*

$$\begin{aligned} & M_t^{K,f} \\ &= \langle X_t^K, f \rangle - \langle X_0^K, f \rangle - \int_0^t \int_{\mathbb{R}^d} \big(\, b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x, y) X_s^K(dy) \, \big) \, f(x) X_s^K(dx) ds \\ &- \int_0^t \int_{\mathbb{R}^d} u_K \mu(x) \, \big(\, K^\lambda r(x) + b(x) \, \big) \big(\int_{\mathbb{R}^d \setminus \{0\}} f(x+h) m(x, dh) - f(x) \, \big) \, X_s^K(dx) ds \end{aligned} \tag{4.3.1}$$

is a square integrable martingale with quadratic variation

$$\begin{aligned} & \langle M_\cdot^{K,f} \rangle_t \\ &= \frac{1}{K} \Bigg\{ \int_0^t \int_{\mathbb{R}^d} \left(2K^\lambda r(x) + b(x) + d(x) + \int_{\mathbb{R}^d} \alpha(x, y) X_s^K(dy) \right) f^2(x) X_s^K(dx) ds \\ &+ \int_0^t \int_{\mathbb{R}^d} u_K \mu(x) \left(K^\lambda r(x) + b(x) \right) \left(\int_{\mathbb{R}^d \setminus \{0\}} f^2(x+h) m(x, dh) - f^2(x) \right) X_s^K(dx) ds \Bigg\}. \end{aligned} \tag{4.3.2}$$

Remark 4.3.2. *By choosing different range of the exponent λ and studying the explicit representation of semi-martingale $\langle X^K, f\rangle$, one can have either a deterministic or a stochastic large population limit depending on whether the asymptotic limit of quadratic variation of the martingale part vanishes or not.*

Proof. By martingale problem theory, for any bounded measurable functional ϕ on $\mathcal{M}_F(\mathcal{X})$, one knows that the process

$$\phi(X_t^K) - \phi(X_0^K) - \int_0^t L^K\phi(X_s^K)ds \tag{4.3.3}$$

is a càdlàg martingale. Let $\phi(\nu) = \langle \nu, f\rangle$, $\forall f \in B(\mathcal{X})$, we obtain that

$$\begin{aligned}
&M_t^{K,f}\\
&:= \langle X_t^K, f\rangle - \langle X_0^K, f\rangle - \int_0^t \int_{\mathbb{R}^d} \big(\, b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x,y)X_s^K(dy)\,\big)\, f(x)X_s^K(dx)ds\\
&- \int_0^t \int_{\mathbb{R}^d} u_K\mu(x)(K^\lambda r(x) + b(x))\,\big(\int_{\mathbb{R}^d\setminus\{0\}} f(x+h)m(x,dh) - f(x)\,\big)\, X_s^K(dx)ds
\end{aligned}$$

is a square integrable martingale.

By applying Itô formula to $\langle X_t^K, f\rangle^2$ w.r.t. semimartingale $\langle X_t^K, f\rangle$, one obtains

$$\begin{aligned}
&\langle X_t^K, f\rangle^2 - \langle X_0^K, f\rangle^2 - 2\int_0^t \int_{\mathbb{R}^d} \Big\{\big(\, b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x,y)X_s^K(dy)\,\big)\, f(x)\\
&+ u_K\mu(x)(K^\lambda r(x) + b(x))\,\big(\int_{\mathbb{R}^d\setminus\{0\}} f(x+h)m(x,dh) - f(x)\,\big)\Big\}\, \langle X_s^K, f\rangle X_s^K(dx)ds\\
&- \langle M_\cdot^{K,f}\rangle_t
\end{aligned}$$

is a martingale.

On the other hand, let $\phi(\nu) = \langle \nu, f\rangle^2$ in (4.3.3), we get the following martingale

$$\begin{aligned}
&\langle X_t^K, f\rangle^2 - \langle X_0^K, f\rangle^2 - \int_0^t L^K\langle X_s^K, f\rangle^2 ds\\
=&\langle X_t^K, f\rangle^2 - \langle X_0^K, f\rangle^2\\
&- \frac{1}{K}\Big\{\int_0^t \int_{\mathbb{R}^d} \big(\, 2K^\lambda r(x) + b(x) + d(x) + \int_{\mathbb{R}^d} \alpha(x,y)X_s^K(dy)\,\big)\, f^2(x)X_s^K(dx)ds\\
&+ \int_0^t \int_{\mathbb{R}^d} u_K\mu(x)\,\big(\, K^\lambda r(x) + b(x)\,\big)\big(\int_{\mathbb{R}^d\setminus 0} f^2(x+h)m(x,dh) - f^2(x)\,\big)\, X_s^K(dx)ds\Big\}\\
&- \int_0^t \int_{\mathbb{R}^d} \Big\{\big(\, b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x,y)X_s^K(dy)\,\big)\, f(x)\\
&+ u_K\mu(x)(K^\lambda r(x) + b(x))\,\big(\int_{\mathbb{R}^d\setminus\{0\}} f(x+h)m(x,dh) - f(x)\,\big)\Big\}\, 2\langle X_s^K, f\rangle X_s^K(dx)ds.
\end{aligned}$$

By uniqueness of a semimartingale decomposition, the quadratic variation would be

$$\langle M^{K,f}_{\cdot}\rangle_t$$
$$= \frac{1}{K}\left\{ \int_0^t \int_{\mathbb{R}^d} \left(2K^{\lambda} r(x) + b(x) + d(x) + \int_{\mathbb{R}^d} \alpha(x,y) X^K_s(dy)\right) f^2(x) X^K_s(dx) ds \right.$$
$$\left. + \int_0^t \int_{\mathbb{R}^d} u_K \mu(x) \left(K^{\lambda} r(x) + b(x)\right) \left(\int_{\mathbb{R}^d\setminus\{0\}} f^2(x+h) m(x,dh) - f^2(x)\right) X^K_s(dx) ds \right\},$$

which concludes the proof. □

More precisely, we obtain the following convergence results for both $0 < \lambda < 1$ and $\lambda = 1$ in the subsequent subsection.

Proposition 4.3.3. *Assume $\lambda \in (0,1)$, $u_K = 0$, and X^K_0 converges in law to X_0 under the weak topology on $\mathcal{M}_F(\mathcal{X})$ as $K \to +\infty$, and* $\sup_K \mathbb{E}(\langle X^K_0, 1\rangle^3) < \infty$*.*

Then, $\forall T > 0$, a sequence of processes $(X^K_t, t \in [0,T]) \in D([0,T], \mathcal{M}_F(\mathcal{X}))$ converges in law to an unique process $(X_t, t \in [0,T]) \in C([0,T], \mathcal{M}_F(\mathcal{X}))$, satisfying the following:

$$\sup_{t\in[0,T]} \mathbb{E}(\langle X_t, 1\rangle^3) < \infty, \tag{4.3.4}$$

and for any bounded measurable function f,

$$\langle X_t, f\rangle = \langle X_0, f\rangle + \int_0^t \int_{\mathbb{R}^d} (b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x,y) X_s(dy)) f(x) X_s(dx) ds. \tag{4.3.5}$$

Remark 4.3.4. *(a) For the one-dimensional case, if $X^K_0 = \frac{N^K_0}{K}\delta_x$ and $\frac{N^K_0}{K} \to n_0$ in law, then $(\frac{N^K_t}{K}, t \in [0,T])$ converges to $(n_t, t \in [0,T])$ which satisfies logistic equation*

$$\dot{n}_t = (b - d - \alpha n_t) n_t. \tag{4.3.6}$$

(b)Similarly, for the dimorphic case, we have the following limit

$$\left\{ \begin{array}{l} \dot{n}_t = (b(x) - d(x) - \alpha(x,x) n_t - \alpha(x,y) m_t) n_t \\ \dot{m}_t = (b(y) - d(y) - \alpha(y,x) n_t - \alpha(y,y) m_t) m_t \end{array} \right. . \tag{4.3.7}$$

This Proposition and the next one can be deduced from Proposition 4.3.1 by the classical uniqueness-compactness argument. Please refer to [22] for similar proofs.

4.3.2 Convergence to a superprocess with competition for $\lambda = 1$

Classical Lotka-Volterra model is described by a set of ordinary differential equations as in last subsection. By stability analysis, we can determine the domain of attraction for coexistence or competitive exclusion according to the relative strengths of competitions within and between the species. In this section, we justify a stochastic version of Lotka-Volterra models in the settings of one-type and two-type system.

Firstly, we give a general form of stochastic Lotka-Volterra systems by a superprocess characterization. We always assume $\lambda = 1$ and $u_K = 0$ in this section.

Proposition 4.3.5. *Assume* X_0^K *converges in law to* X_0 *under the weak topology on* $\mathcal{M}_F(\mathcal{X})$ *as* $K \to +\infty$*, and that* $\sup_K \mathbb{E}(\langle X_0^K, 1\rangle^3) < \infty$*.*

Then, $\forall T > 0$*, the sequence of processes* $(X_t^K, t \in [0,t]) \in D([0,T], \mathcal{M}_F(\mathcal{X}))$ *converges in law to an unique process* $(X_t, t \in [0,T]) \in C([0,T], \mathcal{M}_F(\mathcal{X}))$*, satisfying the following:*

$$\sup_{t\in[0,T]} \mathbb{E}(\langle X_t, 1\rangle^3) < \infty, \tag{4.3.8}$$

and for any bounded measurable function f*,*

$$M_t^f := \langle X_t, f\rangle - \langle X_0, f\rangle - \int_0^t \int_{\mathbb{R}^d} \Big(b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x,y) X_s(dy) \Big) f(x) X_s(dx) ds \tag{4.3.9}$$

is a continuous martingale with quadratic variation

$$\langle M^f\rangle_t = 2\int_0^t \int_{\mathbb{R}^d} r(x) f^2(x) X_s(dx) ds. \tag{4.3.10}$$

Remark 4.3.6. *The process* X *above can be rewritten in a weak form*

$$\partial_t X_t(x) = \Big(b(x) - d(x) - \int_{\mathbb{R}^d} \alpha(x,y) X_t(dy) \Big) X_t(x) + \dot{M}_t. \tag{4.3.11}$$

Proof. We need to prove uniqueness of the solution of the martingale problem by Dawson's Girsanov transform. Tightness of process sequence $(X_t^K, t \in [0,T])$ can be proved by Aldous-Rebolledo criterion. Convergence limit of the martingale sequence can be deduced by Proposition 4.3.1. ☐

Logistic type Feller diffusion and its Q-process

When there is only one type without mutation, it is called logistic branching process (jump case) or logistic Feller diffusion (diffusion case) studied by Lambert [33] and Etheridge [19], respectively.

Definition 4.3.7 (Logistic branching process)**.** *For any fixed* $K \in \mathbb{N}$*, an integer-valued process* N_t^K *is called a (binary) logistic branching process if the transition rates have the following form:*

$$q_{i,j}^K = \begin{cases} b_K i & \text{if } j = i+1, \\ (d_K + \alpha\frac{i-1}{K})i, & \text{if } j = i-1, \\ -(b_K + d_K + \alpha\frac{i-1}{K})i, & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$

Corollary 4.3.8 (Logistic Feller diffusion)**.** *Consider the process* $(Z_t^K)_{t\geq 0}$ *defined by* $Z_t^K := N_t^K/K$*. Suppose* $Z_0^K \overset{law}{\to} Z_0$ *as* $K \to \infty$*. For all* $T > 0$*,* $(Z_t^K, t \in [0,T])$ *converges in law to a logistic Feller diffusion* $(Z_t, t \in [0,T])$ *which satisfies the solution of the following stochastic differential equation:*

$$dZ_t = (b(x) - d(x) - \alpha(x,x)Z_t)Z_t dt + \sqrt{2r(x)Z_t} dB_t, \qquad t > 0. \tag{4.3.12}$$

Proof. According to Proposition 4.3.5, we know $(Z_t^K, 0 \leq t \leq T)$ converges to $(Z_t, 0 \leq t \leq T)$ which is a solution of the following equation

$$\partial_t Z_t = (b(x) - d(x) - \alpha(x,x)Z_t)Z_t + \dot{M}_t, \tag{4.3.13}$$

where M_t is a continuous martingale satisfying

$$\langle M \rangle_t = 2r(x) \int_0^t Z_s ds.$$

Since the continuous martingale can be represented by Itô integral, we have

$$M_t = \int_0^t \sqrt{2r(x)Z_s} dB_s,$$

where B_t is one dimensional standard Brownian motion. Then (4.3.13) can be rewritten as

$$dZ_t = (b(x) - d(x) - \alpha(x,x)Z_t)Z_t dt + \sqrt{2r(x)Z_t} dB_t, \qquad t > 0,$$

which concludes the proof. □

Further more, as proved in [33], the diffusion limit defined by (4.3.12) will get extinct in finite time a.s. provided there is no immigration from outside. The point 0 is thus an absorbing state for the process. Nevertheless, the time for extinction can be large compared to human time scale and it may fluctuate for a long time before extinction actually occurs. On the other hand, the long time behavior of an absorbed Markov process can be well described by the distribution of the trajectories which never attains 0. When it exists, we define a new conditioned process and study its stationary distribution.

Definition 4.3.9 (Q-process)**.** *The distribution* $\mathbb{Q}_z$ *is the law of a process issued from* $z > 0$ *and conditioned to never attain* 0*. When it exists, it is defined as follows: for* $s > 0$ *and for any Borel set* $B \subset C(([0,s], \mathbb{R}_+^*)$*,*

$$\mathbb{Q}_z(Z \in B) = \lim_{t \to \infty} \mathbb{P}_z(Z \in B | t < T_0), \tag{4.3.14}$$

where $T_0 = \inf\{t > 0, Z_t = 0\}$*. This limiting procedure defines the law of a diffusion process that never reaches* 0 *called its corresponding Q-process. Denote the new conditional process of Z. by* $Z_\cdot^{\uparrow}$*.*

By spectral theory arguments, one can get the quasi-stationary distribution of a diffusion process. What is more, one can express the law of the Q-process and its stationary distribution in terms of the quasi-stationary measure. Concerning the existence and uniqueness of a quasi-stationary measure for the process defined by (4.3.12), we can refer to [7]. Since it can not be written down in an explicit form, Villemonais [43] provides an approximation method to simulate the quasi-stationary distribution based on a Fleming-Viot system. Concerning the Q-process, one can view it as a modification of the original process by adding some immigration structure on it (see [34]). For fixed $x \in \mathcal{X}$, consider the logistic Feller diffusion

$$dZ_t = (b(x) - d(x) - \alpha(x,x)Z_t)Z_t dt + \sqrt{2r(x)Z_t} dB_t, \qquad t \geq 0. \tag{4.3.15}$$

By letting $\tilde{Z}_t = \sqrt{\frac{2Z_t}{r(x)}}$, one obtains a drifted Brownian motion on $(0,\infty)$

$$d\tilde{Z}_t = dB_t - q(\tilde{Z}_t)dt. \tag{4.3.16}$$

Lemma 4.3.10 ([7]). *(i) For all $z > 0$, $s \geq 0$, the law of the Q-process of the process $(\tilde{Z}_t)_{t\geq 0}$ exists as a probability measure $\tilde{\mathbb{Q}}_z^{(x)}$ on path space $C([0,\infty), \mathbb{R}_+^*)$ with transition probability given by*

$$q(s,z,y) = e^{\lambda_1 s}\frac{\xi_1(y)}{\xi_1(z)}p(s,z,y)e^{-Q(y)}, \tag{4.3.17}$$

where

$$Q(y) := 2\int_0^y q(z)dz,$$

and ξ_1 is the corresponding eigenfunction of the first (positive) spectrum λ_1 of the operator:

$$Lg = \frac{1}{2}g'' - qg',$$

where $q(z) = \frac{1}{2z} - \frac{(b(x)-d(x))z}{2} + \frac{\alpha(x,x)r(x)z^3}{4}$.

(ii) For any Borel set $A \subset \mathbb{R}_+^*$ and any $z > 0$,

$$\lim_{s\to\infty} \tilde{\mathbb{Q}}_z^{(x)}(Z_s \in A) = \int_A \xi_1^2(y)e^{-Q(y)}dy =: \tilde{\pi}^{(x)}(A). \tag{4.3.18}$$

Remark 4.3.11. *The Q-process distribution for $\tilde{Z}$ and Z are related by an immediate change of variables, so that above results for $\tilde{Z}$ can be straightforward translated to results on Z. We denote by $\mathbb{Q}_z^{(x)}(\cdot)$ the distribution of the Q-process of Z issued from $z > 0$ indexed with a phenotype $x \in \mathcal{X}$. Furthermore, we denote by $\pi^{(x)}(\cdot)$ the invariant measure of the Q-process.*

Lotka-Volterra type Feller diffusion

Concerning another example-a binary branching system with competition, we can obtain a set of Feller diffusions with Lotka-Volterra drift in a time-space scaling limit. Firstly, we define the following binary branching process with density dependent competition as a $\mathbb{N}^2$-valued process (N_t^K, M_t^K).

Suppose that the transition rates of (N_t^K, M_t^K) are

$$\mathcal{Q}^K = \begin{cases} b_K(x)i, & (i,j) \to (i+1,j) \\ (d_K(x) + \alpha(x,x)\frac{i-1}{K} + \alpha(x,y)\frac{j}{K})i, & (i,j) \to (i-1,j) \\ b_K(y)j, & (i,j) \to (i,j+1) \\ (d_K(y) + \alpha(y,x)\frac{i}{K} + \alpha(y,y)\frac{j-1}{K})j, & (i,j) \to (i,j-1). \end{cases}$$

Consider a sequence of rescaled processes $(\frac{N_t^K}{K}, \frac{M_t^K}{K})_{t\geq 0}$.

Corollary 4.3.12 (Lotka-Volterra type Feller diffusion). *Suppose* $(\frac{N_0^K}{K}, \frac{M_0^K}{K})$ *converges in distribution to* (Z_0^1, Z_0^2) *(maybe random) as* $K \to \infty$*. Then for any* $T > 0$, $\big(Z_t^{1,K}, Z_t^{2,K}\big)$ $:= \big((\frac{N_t^K}{K}, \frac{M_t^K}{K}), t \in [0,T]\big)$ *converges to* $\big((Z_t^1, Z_t^2), t \in [0,T]\big)$ *in distribution as* $K \to \infty$*, where* $(Z_t^1, Z_t^2)_{t\geq 0}$ *satisfies the following SDEs*

$$\left\{ \begin{array}{l} dZ_t^1 = (b(x) - d(x) - \alpha(x,x)Z_t^1 - \alpha(x,y)Z_t^2)Z_t^1 dt + \sqrt{2r(x)Z_t^1} dB_t^1 \\ dZ_t^2 = (b(y) - d(y) - \alpha(y,x)Z_t^1 - \alpha(y,y)Z_t^2)Z_t^2 dt + \sqrt{2r(y)Z_t^2} dB_t^2. \end{array} \right. \tag{4.3.19}$$

Here, $(B_t^1)_{t\geq 0}$ *and* $(B_t^2)_{t\geq 0}$ *are two independent standard Brownian motion.*

Remark 4.3.13. *By using some comparison argument, we can easily get that* $(Z_t^1, Z_t^2)_{t\geq 0}$ *will be absorbed by* $(0,0)$ *in finite time with probability 1. In fact, assume that* $(Z_t^3, Z_t^4)_{t\geq 0}$ *(without interspecific competition) satisfies the following equations:*

$$\left\{ \begin{array}{l} dZ_t^3 = (b(x) - d(x) - \alpha(x,x)Z_t^3)Z_t^3 dt + \sqrt{2r(x)Z_t^3} dB_t^1 \\ dZ_t^4 = (b(y) - d(y) - \alpha(y,y)Z_t^4)Z_t^4 dt + \sqrt{2r(y)Z_t^4} dB_t^2. \end{array} \right. \tag{4.3.20}$$

Obviously, $(Z_t^1, Z_t^2)_{t\geq 0}$ *can be dominated by* $(Z_t^3, Z_t^4)_{t\geq 0}$ *a.s.. Since it is already known that* $(Z_t^3, Z_t^4)_{t\geq 0}$ *can be absorbed by* $(0,0)$ *in finite time with probability 1, so does* $(Z_t^1, Z_t^2)_{t\geq 0}$*. Hence, we will study its long time behavior by the stationary distribution of its Q-process (see Proposition 4.5.4).*

Proof. Let $X_t^K = \frac{N_t^K}{K}\delta_x + \frac{M_t^K}{K}\delta_y$ in Proposition 4.3.5 and suppose the limiting process $(X_t)_{t\geq 0}$ has the form $X_t = Z_t^1\delta_x + Z_t^2\delta_y$. If we take $f = \mathbf{1}_{\{x\}}$, then

$$M_t^f = Z_t^1 - Z_0^1 - \int_0^t (b(x) - d(x) - \alpha(x,x)Z_s^1 - \alpha(x,y)Z_s^2)Z_s^1 ds$$

is a continuous martingale with quadratic variation

$$\langle M^f \rangle_t = 2\int_0^t r(x)Z_s^1 ds.$$

So, we get process $(Z_t^1)_t \geq 0$ satisfying the equation

$$dZ_t^1 = (b(x) - d(x) - \alpha(x,x)Z_t^1 - \alpha(x,y)Z_t^2)Z_t^1 dt + \sqrt{2r(x)Z_t^1} dB_t^1$$

for some standard Brownian motion $(B_t^1)_{t\geq 0}$. By taking $f = \mathbf{1}_{\{y\}}$, we can prove $(Z_t^2)_{t\geq 0}$ satisfying the equation

$$dZ_t^2 = (b(y) - d(y) - \alpha(y,x)Z_t^1 - \alpha(y,y)Z_t^2)Z_t^2 dt + \sqrt{2r(y)Z_t^2} dB_t^2,$$

where $(B_t^2)_{t\geq 0}$ is a standard Brownian motion independent from $(B_t^1)_{t\geq 0}$. ☐

4.4 TSS limit for $0 < \lambda < 1$

In this section we will attain the trait substitute sequence model for accelerating exponent $\lambda \in (0, 1)$ by taking a rare mutation limit. Since the fixation probability of the selectively neutral mutation approaches 0 under the large population limit, the successive substitutions are not visible any more on a single mutation time scale. Therefore, we choose the accumulation of mutation time scale as a proper renormalization scale. Our result proceeds by the way of an intermediate approximation based on the result in Champagnat [8].

4.4.1 Intermediate-scaling approximation on the mutation time scale

For any $K \in \mathbb{N}$, define

$$Y_t^K := \begin{cases} n_0\delta_x, & t = 0 \\ \bar{n}(\eta_t^K)\delta_{\eta_t^K}, & t > 0 \end{cases} \tag{4.4.1}$$

where η^K is a Markov jump process starting from x and with an infinitesimal generator:

$$A^K\varphi(x) = \int_{\mathbb{R}^d} (\varphi(x+h) - \varphi(x))\mu(x)\bar{n}(x)\frac{r(x)}{K^\lambda r(x+h)}[\bar{f}(x+h, x)]_+ m(x, dh). \tag{4.4.2}$$

The next Lemma is used to give an intermediate approximation of X^K on the mutation time scale $1/K^{1+\lambda}u_K$.

Proposition 4.4.1. *Admit assumptions* ($\boldsymbol{A}1$) *and* ($\boldsymbol{A}2$). *Suppose that* $X_0^K = \frac{N_0^K}{K}\delta_x$ *and are such that* $\frac{N_0^K}{K} \xrightarrow{law} n_0 > 0$ *as* $K \to +\infty$. *Further suppose that* $\exp\{-CK^{1-\lambda}\} \ll u_K \ll \dfrac{1}{K^{1+\lambda}\ln K}$, $\forall C > 0$.

Then $(X^K_{t/K^{1+\lambda}u_K}, t \geq 0)$ *can be approximated by* $(Y_t^K)_{t\geq 0}$ *in the sense of f.d.d., i.e. for any* $n \geq 1, \varepsilon > 0, 0 < t_1 < t_2 < ... < t_n < \infty$, *and for any measurable sets* $\Gamma_1, ..., \Gamma_n \subset \mathcal{X}$,

$$\lim_{K\to\infty} \frac{\mathbb{P}(\forall 1 \leq i \leq n, \exists x_i \in \Gamma_i : Supp(X^K_{t_i/K^{1+\lambda}u_K}) = \{x_i\}, |\langle X^K_{t_i/K^{1+\lambda}u_K}, 1\rangle - \bar{n}(x_i)| < \varepsilon)}{\mathbb{P}(\forall 1 \leq i \leq n, \eta^K_{t_i} \in \Gamma_i)} = 1. \tag{4.4.3}$$

Remark 4.4.2. *(a) Since the individual growth rate* $b_K(x) - d_K(x) = b(x) - d(x)$ *is preserved independently from K, the selectively advantageous type evolves like a supercritical branching process with exponential growth. Thus, the fixation time is of order* $\ln K$. *On the other hand, the large population limit of a monomorphic population without mutation is deterministic as the solution of a logistic differential equation* (4.3.6). *Thus, it has a locally stable equilibrium density.*

(b) The fixation probability of the slightly advantageous mutant is the surviving probability of the asymptotic critical branching process:

$$\frac{b_K(x+h) - d_K(x+h) - \alpha(x+h,x)\bar{n}(x)}{b_K(x+h)} \vee 0 = \frac{[\bar{f}(x+h,x)]_+}{K^\lambda r(x+h) + b(x+h)}$$

while the transition rate is $\mu(x)\bar{n}(x)r(x)$ *on the mutation time scale* $O(1/K^{1+\lambda}u_K)$.

The proof of this proposition is similar to the proof in [8, Theorem 1] by just replacing the auxiliary lemmas with the following ones. Instead of proving the proposition itself, we list the following results which are crucial for the detailed proof.

Denote by τ_n the first mutation time after τ_{n-1}, $n \geq 1$. Denote by θ_n the first time after τ_n when the population gets monomorphic again and γ_n be the corresponding survival type. Set $\tau_0 = 0$. Inductively, we can define $\tau_n, \theta_n, \gamma_n, n \geq 2$. Obviously, we have the relation: $\theta_{n-1} < \tau_n < \theta_n, n \geq 2$.

We list the following Lemmas to furnish the proof of the Proposition. The first Lemma gives the characterization of the exponentially distributed waiting time of mutation occurrence.

Lemma 4.4.3. *Under the same condition as in Proposition 4.4.1,*

$$\lim_{K\to+\infty} \mathbb{P}_{\frac{N_0^K}{K}\delta_x}\left(\tau_1 > \frac{t}{K^{1+\lambda}u_K}\right) = \exp\{-\bar{n}(x)\mu(x)r(x)t\}. \tag{4.4.4}$$

Lemma 4.4.3 and its counterpart Lemma 4.5.9 for $\lambda = 1$ case, can be proved by a similar approach used in [8, Lemma 2]. The next Lemma characterizes the fixation probability of the selectively advantageous population in a dimorphic system.

Lemma 4.4.4. *Under the same condition as in Proposition 4.4.1, and further assume* $\frac{N_0^K}{K} \xrightarrow{law} \bar{n}(x)$. *Then,*

$$\lim_{K\to+\infty} K^\lambda \cdot \mathbb{P}_{\frac{N_0^K}{K}\delta_x + \frac{1}{K}\delta_{x+h}}(\gamma_0 = x+h) = \frac{[\bar{f}(x+h,x)]_+}{r(x+h)}.$$

Rather than proving this lemma, we will prove its counterpart Lemma 4.5.8 in next section. The arguments are nearly the same by taking λ in different ranges.

As for the time needed for the process to drift out of the stable equilibrium, we give the following estimation.

Lemma 4.4.5. *Assume the same condition as in Remark 4.3.4 (a). Obviously,* $\bar{n} = \frac{b-d}{\alpha}$ *is the stable equilibrium.* $\forall\ 0 < \delta < \bar{n}$, *define* $\sigma_K = \inf\{t > 0 : \frac{N_t^K}{K} \in [\bar{n}-\delta, \bar{n}+\delta]^c\}$. *For any subset* $U \subseteq [\bar{n}-\delta, \bar{n}+\delta]$, *there exists a constant* $\bar{V} > 0$, *s.t.*

$$\lim_{K\to+\infty} \inf_{x\in U} \mathbb{P}_x(\sigma_K \geq e^{K^{1-\lambda}\bar{V}}) = 1. \tag{4.4.5}$$

Proof. The infinitesimal generator of $(\frac{N_t^K}{K}, t \geq 0)$ can be written, for $g(\cdot) \in C_b^2(R)$,

$$\begin{aligned}
L^K g(z) &= (g(z+\frac{1}{K}) - g(z))(K^\lambda r + b)Kz \\
&\quad + g(z-\frac{1}{K}) - g(z))(K^\lambda r + d + \alpha z)Kz \\
&= g'(z)(b-d-\alpha z)z + \frac{K^\lambda r + b}{K}[g(z+\frac{1}{K}) - g(z) - \frac{1}{K}g'(z)]K^2 z \\
&\quad + \frac{K^\lambda r + d + \alpha z}{K}[g(z-\frac{1}{K}) - g(z) + \frac{1}{K}g'(z)]K^2 z. \qquad (4.4.6)
\end{aligned}$$

Concerning the problem of exit from the domain on the above equation perturbed by random small noise, we can handle it in the same approach as the following diffusion

$$dZ_t^K = (b-d-\alpha Z_t^K)Z_t^K dt + \sqrt{\frac{1}{K^{1-\lambda}}}\sqrt{2rZ_t^K}dB_t. \qquad (4.4.7)$$

From the well known Freidlin-Wentzell theory (see [23]), there exists a constant $\bar{V} > 0$, s.t. $\forall \delta > 0$,

$$\lim_{K\to+\infty} \inf_{x\in U} \mathbb{P}_x(e^{K^{1-\lambda}(\bar{V}-\delta)} < \sigma_K < e^{K^{1-\lambda}(\bar{V}+\delta)}) = 1. \qquad (4.4.8)$$

So we can conclude the results.

□

4.4.2 Rescaling on an accumulation of mutations time scale

Theorem 4.4.6 $(0 < \lambda < 1)$**.** *Admit assumptions* $(\boldsymbol{A}1)$ *and* $(\boldsymbol{A}2)$*. Suppose that* $X_0^K = \frac{N_0^K}{K}\delta_x$ *and are such that* $\frac{N_0^K}{K} \overset{law}{\to} n_0 > 0$*, as* $K \to +\infty$*, and* $\exp\{-CK^{1-\lambda}\} \ll u_K \ll \dfrac{1}{K^{1+\lambda}\ln K}$*,* $\forall\, C > 0$*. Then,* $(X_{t/Ku_K}^K, t \geq 0)$ *converges in the sense of f.d.d to*

$$Y_t = \begin{cases} n_0\delta_x, & t = 0 \\ \bar{n}(\eta_t)\delta_{\eta_t}, & t > 0 \end{cases} \qquad (4.4.9)$$

where the Markov jump process $(\eta_t, t \geq 0)$ *satisfies* $\eta_0 = x$ *with an infinitesimal generator:*

$$A\varphi(x) = \int_{\mathbb{R}^d} (\varphi(x+h) - \varphi(x))\mu(x)\bar{n}(x)\frac{r(x)}{r(x+h)}[\bar{f}(x+h,x)]_+ m(x,dh). \qquad (4.4.10)$$

Remark 4.4.7. *For large* K*, we can see the mutant type evolves as an asymptotic critical branching process with (*K*-dependent) fixation probability* $\frac{[\bar{f}(x+h,x)]_+}{K^\lambda r(x+h)+b(x+h)}$ *while the accumulation rate of mutants is of order* $K^\lambda = \frac{K^{1+\lambda}u_K}{Ku_K}$*. Therefore, the total transition rate is approximated by* $\frac{r(x)}{r(x+h)}[\bar{f}(x+h,x)]_+$*. In other words, the phase transition is not because the new mutant has absolute preferability than the resident one but because the mutations occur more intensively.*

Proof. Because of Proposition 4.4.1, in order to prove the convergence from $(X^K_{t/Ku_K})_{t\geq 0}$ to $(Y_t)_{t\geq 0}$, it is sufficient to prove $(Y^K_{t\cdot K^\lambda})_{t\geq 0}$ converges to $(Y_t)_{t\geq 0}$.

From (4.4.3), one obtains the following by replacing $t_i = s_i K^\lambda$,

$$\lim_{K\to\infty} \frac{\mathbb{P}(\forall 1 \leq i \leq n, \exists x_i \in \Gamma_i : Supp(X^K_{s_i/Ku_K}) = \{x_i\}, |\langle X^K_{s_i/Ku_K}, 1\rangle - \bar{n}(x_i)| < \varepsilon)}{\mathbb{P}(\forall 1 \leq i \leq n, \eta^K_{s_i\cdot K^\lambda} \in \Gamma_i)} = 1. \tag{4.4.11}$$

Furthermore, it is not hard to prove $(\eta^K_{s\cdot K^\lambda})_{s\geq 0}$ converges to $(\eta_s)_{s\geq 0}$ from the convergence of generators $K^\lambda A^K$ to A as $K \to \infty$. Meanwhile, the convergence from $\bar{n}(\eta^K_{s\cdot K^\lambda})$ to $\bar{n}(\eta_s)$ is implied. As $K \to \infty$, the limit of $\mathbb{P}(\forall 1 \leq i \leq n, \eta^K_{s_i\cdot K^\lambda} \in \Gamma_i)$ exists and

$$\lim_{K\to\infty} \mathbb{P}(\forall 1 \leq i \leq n, \eta^K_{s_i\cdot K^\lambda} \in \Gamma_i) = \mathbb{P}(\forall 1 \leq i \leq n, \eta_{s_i} \in \Gamma_i), \tag{4.4.12}$$

where η is defined as in (4.4.10). By combining (4.4.11) and (4.4.12), one obtains

$$\lim_{K\to\infty} \mathbb{P}(\forall 1 \leq i \leq n, \exists x_i \in \Gamma_i : Supp(X^K_{s_i/Ku_K}) = \{x_i\}, |\langle X^K_{s_i/Ku_K}, 1\rangle - \bar{n}(x_i)| < \varepsilon) = \mathbb{P}(\forall 1 \leq i \leq n, \eta_{s_i} \in \Gamma_i), \tag{4.4.13}$$

which yields the f.d.d. convergence. ☐

4.5 Conditioned TSS limit for $\lambda = 1$

4.5.1 Birth and death processes in random environments

Kaplan (see [29]) formulated a continuous time branching model in random environments and gave a sufficient and necessary condition for extinction. In our regime, it is of particular interest to consider the binary branching case, that is, with linear birth and death rates. More precisely, one is given a stationary ergodic process $(\omega_t)_{t\geq 0}$ taking values in $\mathbb{R}^*_+$ with the initial distribution π. It amounts to assuming that for process $(X_t)_{t\geq 0}$ both birth rate $b(t,\omega)$ and death rate $d(t,\omega)$ per individual at time t are specified by an environmental process ω. For a realization of environmental process ω, we can get a continuous time non-homogeneous birth and death process as in [25, Chapter 5]. Notice that there are two different sense of probability measures in our framework. One is probability measure $P^\omega(\cdot)$ called "quenched" and its corresponding expectation E^ω if the environment ω is specified in advance. Another one is P, which is defined on the environment space $\Omega = \{(\omega_t)_{t\geq 0} : \omega \in C([0,+\infty), \mathbb{R}^*_+)\}$. We denote by $\mathbb{P} = P \otimes P^\omega$ the product of measures.

Firstly, we give sufficient and necessary conditions for a.s. extinction as well as the probability for non-extinction case.

Lemma 4.5.1. *Let* $\rho(\omega) := b(0,\omega) - d(0,\omega)$. *Assume* $P(X_0 = 1) = 1$ *and* 0 *is an absorbing state. Then* *(i)* $P^\omega(\lim_{t\to\infty} X_t = 0) + P^\omega(\lim_{t\to\infty} X_t = \infty) = 1.$

(ii) $P^\omega(\lim_{t\to\infty} X_t = 0) = 1$ *a.e., iff*

$$E\rho(\omega) = \int (b(0,\omega) - d(0,\omega))d\pi \le 0. \tag{4.5.1}$$

(iii) If $E\rho(\omega) > 0$*, then the quenched extinction probability*

$$\begin{aligned} q_\omega &:= P^\omega(\lim_{t\to\infty} X_t = 0) \\ &= \frac{\int_0^\infty d(t,\omega)e^{\int_0^t (d(\tau,\omega)-b(\tau,\omega))d\tau}dt}{1+\int_0^\infty d(t,\omega)e^{\int_0^t (d(\tau,\omega)-b(\tau,\omega))d\tau}dt} \end{aligned} \tag{4.5.2}$$

Proof. We can refer to [29] for the proof of (i) and (ii).

From related results of non-homogeneous branching process (Page 104, [25]), we get the quenched extinction probability with one ancestor initially

$$\begin{aligned} q_\omega &:= P^\omega(\lim_{t\to\infty} X_t = 0) = \lim_{t\to\infty} P_{1,0}(t) \\ &= \lim_{t\to\infty} \frac{\int_0^t d(s,\omega)e^{\int_0^s (d(\tau,\omega)-b(\tau,\omega))d\tau}ds}{1+\int_0^t d(s,\omega)e^{\int_0^s (d(\tau,\omega)-b(\tau,\omega))d\tau}ds}. \end{aligned}$$

This concludes the proof of (iii). □

Define $T_K = inf\{t \ge 0 : X_t = K\}$ and T_0 is the extinction time. The following lemma will give estimation of order of time needed to exceed some given level provided non-extinction.

Lemma 4.5.2. *Assume* $E\rho(\omega) > 0, E\rho(\omega)^2 < \infty$*. Then,*

(i) as $t \to \infty$*,*

$$(X_t \exp(-t\mathbb{E}\rho))^{t^{-1/2}} \longrightarrow We^V \tag{4.5.3}$$

in distribution, where W *and* V *are independent,* $\mathbb{P}(W = 0) = 1 - \mathbb{P}(W = 1) = \mathbb{P}(\lim_{t\to\infty} X_t = 0) = E_P(q_\omega) =: q$ *and* V *has a normal distribution* $\mathcal{N}(0,\sigma^2)$*.*

(ii)

$$\lim_{K\to\infty} \mathbb{P}(T_K < T_0) = 1 - q \tag{4.5.4}$$

$$\forall\ t_K \gg \ln K,\ \lim_{K\to\infty} \mathbb{P}(T_K < t_K \mid T_K < T_0) = 1 \tag{4.5.5}$$

$$\forall\ s_K \ll \ln K,\ \lim_{K\to\infty} \mathbb{P}(T_K > s_K \mid T_K < T_0) = 1 \tag{4.5.6}$$

Remark 4.5.3. *Actually, (i) is deduced by the central limit theorem for* $\ln X_t$ *(see [30, 31]). Heuristically, it can be written as*

$$X_t \sim W e^{tE\rho + t^{1/2}V + \circ(t^{1/2})} \qquad \textit{for large } t. \tag{4.5.7}$$

We see that, provided non-extinction $(W = 1)$*, the process* $(X_t)_{t\geq 0}$ *behaves almost like an exponential growth model.*

(4.5.5) *and* (4.5.6) *together give a sharp estimation of order of the hitting time* T_K*, which is* $O(\ln K)$*.*

Proof. Proof of (ii).

$$\begin{aligned}\lim_{K\to\infty} \mathbb{P}(T_K < T_0) &= \mathbb{P}(T_0 = \infty) = \mathbb{P}(\lim_{t\to\infty} X_t = \infty)\\ &= 1 - q\end{aligned}$$

Now we prove (4.5.5). $\forall\ t_K \gg \ln K$,

$$\begin{aligned}\lim_{K\to\infty} \mathbb{P}(T_K < t_K \mid T_K < T_0) &= \lim_{K\to\infty} \mathbb{P}(X_{t_K} > K \mid T_K < T_0)\\ &= \lim_{K\to\infty} \mathbb{P}(W e^{t_K E\rho} > K \mid W = 1)\\ &= 1\end{aligned}$$

where the second equal sign is due to (4.5.3).

The proof of (4.5.6) is similar. □

4.5.2 Fixation and extinction analysis

According to Lemma 4.3.10, each equation in (4.3.20) defines its Q-process and the Q-processes have stationary distributions as random variables $n(x)$ and $n(y)$ defined in Remark 4.3.11. The following results provide us a non-coexistence condition for a dimorphic system based on the results in Section 4.5.1. In contrast to the deterministic fitness function defined in [8], the fitness function here will be random.

Proposition 4.5.4 (non-coexistence condition). *Consider Lotka-Volterra type Feller diffusion* $(Z_t^1, Z_t^2)_{t\geq 0}$ *(see* (4.3.19)*) conditioned to be never attaining (0,0)* $(Z_t^1 + Z_t^2 > 0)$*, starting with some positive initial state* (z_0^1, z_0^2)*. Assume*

$$Ef(x,y)f(y,x) < 0, \tag{4.5.8}$$

where

$$\begin{aligned} f(x,y) &= b(x) - d(x) - \alpha(x,y)n(y),\\ f(y,x) &= b(y) - d(y) - \alpha(y,x)n(x).\end{aligned} \tag{4.5.9}$$

Then, the Q-process $(Z_t^1, Z_t^2)_{t\geq 0}^{\uparrow}$ *of* $(Z_t^1, Z_t^2)_{t\geq 0}$ *has stationary distribution as* $(0, n(y))$ *or* $(n(x), 0)$*, depending on whether* $Ef(y,x) > 0$ *or* $Ef(x,y) > 0$ *resp. More precisely,*

(i)

$$\mathbb{P}\big((Z_t^1, Z_t^2)^{\uparrow} \xrightarrow{\mathcal{L}} (n(x), 0) \text{ } or \text{ } (0, n(y))\big) = 1. \tag{4.5.10}$$

(ii)

$$\begin{aligned} &\mathbb{P}\big((Z_t^1, Z_t^2)^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(y))\big) = 1_{\{Ef(x,y)<0\}} \\ &\mathbb{P}\big((Z_t^1, Z_t^2)^{\uparrow} \xrightarrow{\mathcal{L}} (n(x), 0)\big) = 1_{\{Ef(y,x)<0\}}. \end{aligned} \tag{4.5.11}$$

Proof. (i) Because of the stochastic domination in (4.3.20), for the double variables system $(Z_t^1, Z_t^2)_{t\geq 0}$ conditioned on never absorbed by $(0,0)$, the union of axes $\mathbb{R}_+^* \times 0 \cup 0 \times \mathbb{R}_+^*$ is accessible and absorbing. And its complementary set is transient.

(ii)According to the definition of Q-process, we have

$$\mathbb{P}\big((Z_t^1, Z_t^2)^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(y))\big) = \mathbb{P}\big((Z_t^1, Z_t^{2^{\uparrow}}) \xrightarrow{\mathcal{L}} (0, n(y))\big). \tag{4.5.12}$$

Therefore, it is sufficient to prove

$$\mathbb{P}\big((Z_t^1, Z_t^{2^{\uparrow}}) \xrightarrow{\mathcal{L}} (0, n(y))\big) = 1_{\{Ef(x,y)<0\}}. \tag{4.5.13}$$

Obviously, by neglecting some terms in (4.3.19), we get the following stochastic domination relation:

$$\big(Z_t^1, Z_t^2\big) \preceq \big(Z_t^5, Z_t^4\big), \tag{4.5.14}$$

where

$$\begin{cases} dZ_t^5 = (b(x) - d(x) - \alpha(x,y)Z_t^4)Z_t^5 dt + \sqrt{2r(x)Z_t^5}dB_t^1 \\ dZ_t^4 = (b(y) - d(y) - \alpha(y,y)Z_t^4)Z_t^4 dt + \sqrt{2r(y)Z_t^4}dB_t^2. \end{cases} \tag{4.5.15}$$

Then, it implies

$$\big(Z_t^1, Z_t^{2^{\uparrow}}\big) \preceq \big(Z_t^5, Z_t^{4^{\uparrow}}\big), \tag{4.5.16}$$

where Z_t^5 is the updated solution by substituting coefficient Z_t^4 with $Z_t^{4^{\uparrow}}$, i.e.

$$dZ_t^5 = (b(x) - d(x) - \alpha(x,y)Z_t^{4^{\uparrow}})Z_t^5 dt + \sqrt{2r(x)Z_t^5}dB_t^1. \tag{4.5.17}$$

By Lemma 4.3.10, we get $Z_t^{4^{\uparrow}}$ converges in distribution to a random variable $n(y)$ governed by measure $\pi^{(y)}$. So, we can view Z_t^5 as a continuous state branching process in random environments. If the drift coefficient satisfies $Ef(x,y) < 0$, it means Z_t^5 is subcritical and further implies

$$\mathbb{P}\big(Z_t^5 \to 0\big) = 1. \tag{4.5.18}$$

Then, by (4.5.16), we have $\mathbb{P}\big(Z_t^1 \to 0\big) = 1$. Finally, it follows

$$\mathbb{P}\big((Z_t^1, Z_t^{2^{\uparrow}}) \xrightarrow{\mathcal{L}} (0, n(y))\big) = 1_{\{Ef(x,y)<0\}}. \tag{4.5.19}$$

The other formula can be proved similarly. □

Let's denote by Assumption (**B**) the following: for any $x \in \mathcal{X}$,

$$b(x) \equiv b,\ d(x) \equiv d,\ \alpha(x,x) \equiv \alpha,\ r(x) \equiv r.$$

This assumption guarantees that the killing rate of logistic Feller diffusion (its minimal positive spectrum) is independent of the traits (refer to [7]). Unfortunately, this condition is very restrictive. However, the interspecific competition parameter is still free of choice.

Under the above assumptions, we can get the main result of this section.

Theorem 4.5.5 ($\lambda = 1$)**.** *Admit assumptions (**A**1), (**A**3) and (**B**). Denote by* $\beta_1^K, \ldots, \beta_l^K$ *the occurrence times of the first* l *successful mutations of* $X^K_{\cdot/Ku_K}$ *and* $\beta_1, \ldots, \beta_l$ *of the first* l *jump times of the process* V *defined in Definition 4.2.2. Let* $\mathbf{P}$ *be the distribution of the process* V.

Suppose that $X_0^K = \frac{N_0^K}{K}\delta_x$ *and are such that* $\frac{N_0^K}{K} \overset{law}{\to} n_0 > 0$ *as* $K \to +\infty$, *and* $K^2 u_K \ll \frac{1}{\ln K}$. *Then, for any* $l \in \mathbb{N}$, *and* $0 = t_0 < t_1 < \ldots < t_l = t$, *and for any measurable subsets* $A_1, \ldots, A_l$ *of* $\mathbb{R}_+^*$, $B_1, B_2, \ldots, B_l$ *of* $\mathcal{X}$,

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}\Big(\textit{there exist merely } l \textit{ successful mutations before time } t, \textit{ such that } \forall 1 \le i \le l,\\
&\quad \beta_i^K \in (t_{i-1}, t_i);\ \exists x_i \in B_i : \textit{Supp}(X^K_{t_i/Ku_K}) = x_i;\ \langle X^K_{t_i/Ku_K}, 1\rangle \in A_i \;\Big|\; \langle X^K_{t/Ku_K}, 1\rangle > 0\Big)\\
&= \mathbf{P}\Big(\textit{there exist merely } l \textit{ successful mutations before time } t, \textit{ such that } \forall 1 \le i \le l,\\
&\quad \beta_i \in (t_{i-1}, t_i);\ \eta_{t_i} \in B_i;\ n(\eta_{t_i}) \in A_i\Big).
\end{aligned} \tag{4.5.20}$$

Remark 4.5.6. *We divide the entire invasion process into three steps:*

(a) ***Quasi-stationary behavior of resident type before the first mutation.*** *Suppose there is a single type* x *initially. Conditioned on non-extinction, its long time behavior can be characterized by the stationary distribution of its Q-process, which is governed by the invariant measure* $\pi^{(x)}(\cdot)$ *as in Remark* 4.3.11. *By ergodic theorem, we will see that the first mutation happens asymptotically in exponentially distributed time with a parameter* $\gamma(x)$. *Since* $1/K^2u_K \gg \ln K$, *one obtains that* $\left\langle X^K_{\frac{\beta_1^K-}{Ku_K}}, 1 \right\rangle$, *conditioned on non-extinction, converges a random variable with a biased distribution of stationary size* $n(x)$.

(b) ***Population arising from new mutant evolves in stationary random environments constituted by the resident type.*** *Short after there comes a mutant, we can view mutant population* $(x+h)$ *evolving in stationary random environments constituted by resident type* x. *The environmental process has initial distribution* $\pi^{(x)}$. *According to Lemma* 4.5.1, *either the mutant type is subcritical,*

i.e.

$$\begin{aligned} E\rho(\omega) &= E\big(b(0,\omega) - d(0,\omega)\big) \\ &= E\big(b(x+h) - (d(x+h) + \alpha(x+h,x)n(x))\big) \\ &= Ef(x+h,x) < 0, \end{aligned} \tag{4.5.21}$$

then it dies out with probability 1, or the mutant type is supercritical ($E\rho(\omega) = Ef(x+h,x) > 0$*), then we get the quenched probability for the population* $(x+h)$ *to reach some given* $\varepsilon-$*level is* $(1-q_\omega)$*. As for the latter case, according to Lemma* 4.5.2*, the time needed for population* $(x+h)$ *to exceed the* $\varepsilon-$*level is of order* $\ln K$*.*

(c) ***Extinction of disadvantageous type.*** *Suppose that we have condition* $Ef(x+h,x) > 0, Ef(x,x+h) < 0$*. Once mutant type population reaches* $\varepsilon-$*level, according to Proposition* 4.5.4*, the two-type population process conditioned on never absorbed by* $(0,0)$ *would converge in distribution to a stationary distribution* $(0, n(x+h))$ *with probability 1. That is the so-called "invasion implies fixation" principle. In all, the entire invasion period takes time of order* $\ln K$*.*

As in Section 4.4.1, we endow $(\tau_n, \theta_n, \gamma_n)_{n\geq 0}$ with the same explanation. In parallel, we list the following Lemmas to furnish the proof of the main theorem. Lemma 4.5.8 gives the fixation probability which is of order $1/K$. Lemma 4.5.9 gives the right time scale of the mutation occurrence governed by exponential distribution.

Lemma 4.5.7. *Consider the processes* $Z^K_\cdot$ *and* $Z_\cdot$ *given in Corollary 4.3.8 and Corollary 4.3.12. Then it implies that* $Z^{K\uparrow}_\cdot$ *converges weakly* $Z^\uparrow_\cdot$ *as* $K \to \infty$*.*

Proof. Due to similarity, we just prove the case in Corollary 4.3.8. For any Borel $B \in C([0,+\infty), \mathbb{R}^*_+)$,

$$\begin{aligned} &\lim_{K\to\infty} \mathbb{P}(Z^{K\uparrow}_\cdot \in B) \\ &= \lim_{K\to\infty} \lim_{s\to\infty} \mathbb{P}(Z^K_\cdot \in B \mid T^{(K)}_0 > s) \\ &= \lim_{K\to\infty} \lim_{s\to\infty} \frac{\mathbb{P}(Z^K_\cdot \in B,\, T^{(K)}_0 > s)}{\mathbb{P}(T^{(K)}_0 > s)} \\ &= \lim_{s\to\infty} \frac{\mathbb{P}(Z_\cdot \in B,\, T_0 > s)}{\mathbb{P}(T_0 > s)} \\ &= \mathbb{P}(Z^\uparrow_\cdot \in B) \end{aligned} \tag{4.5.22}$$

where $T^{(K)}_0$ denotes the hitting time of 0 by $Z^K_\cdot$ and T_0 denotes the one by $Z_\cdot$. ☐

Lemma 4.5.8. *Under the same condition as in Theorem* 4.5.5*, consider the processes defined in Corollary 4.3.12.*

(i) *If* $Ef(x+h,x) < 0$*, then for any* $z > 0$*,*

$$\lim_{K\to\infty} \mathbb{P}_{z\delta_x + \frac{1}{K}\delta_{x+h}}\big((Z^{1,K}_t, Z^{2,K}_t)^\uparrow \xrightarrow{\mathcal{L}} (n(x), 0)\big) = 1. \tag{4.5.23}$$

(ii) If $Ef(x+h,x) > 0$, then for any $z > 0$,

$$
\begin{aligned}
&\lim_{K\to\infty} K \cdot \mathbb{P}_{z\delta_x + \frac{1}{K}\delta_{x+h}}\big((Z_t^{1,K}, Z_t^{2,K})^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(x+h))\big) \\
&= \int_{C([0,+\infty),\mathbb{R}_+^*)} Q_z^{(x)}(d\omega) \\
&\qquad \Big(r(x+h)\int_0^{\infty} \exp\big\{ -(tb(x+h) - td(x+h) - \int_0^t \alpha(x+h,x)\omega_\tau(x)d\tau)\big\}dt\Big)^{-1} \\
&=: g_x(z,h).
\end{aligned}
$$

Proof. As analyzed in Remark 4.5.6, after there comes a mutant type $(x+h)$, the population of type $(x+h)$ can be approximated by a birth and death process evolving in random environments constituted by resident population x. The proof of (i) is simply implied in Remark 4.5.6 (b). We only need to prove (ii). Obviously, one has

$$
\begin{aligned}
&\mathbb{P}_{z\delta_x + \frac{1}{K}\delta_{x+h}}\big((Z_t^{1,K}, Z_t^{2,K})^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(x+h))\big) \\
&\quad = \mathbb{E}_{Q^{(x)}}\Big[P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}\big((Z_t^{1,K}, Z_t^{2,K})^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(x+h))\big)\Big].
\end{aligned}
\tag{4.5.24}
$$

By a similar arguments as in [8], as analyzed in Remark 4.5.6, the fixation probability of population $(x+h)$ issued from a single mutant is the probability for it to exceed some given $\varepsilon-$level. Denote by $T_k^{(x)}$ the hitting time of level k by population of type x. By Markov property, for given environment ω, one obtains

$$
\begin{aligned}
&\lim_{K\to\infty} K \cdot P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}\big((Z_t^{1,K}, Z_t^{2,K})^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(x+h))\big) \\
&= \lim_{K\to\infty} K \cdot P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}\big((Z_t^{1,K}, Z_t^{2,K})^{\uparrow} \xrightarrow{\mathcal{L}} (0, n(x+h)) \mid T^{(x+h)}_{\lfloor \varepsilon K \rfloor} < T_0^{(x+h)}\big) \\
&\qquad \cdot P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}(T^{(x+h)}_{\lfloor \varepsilon K \rfloor} < T_0^{(x+h)}) \\
&= \lim_{K\to\infty} K \cdot P_{z\delta_x + \varepsilon\delta_{x+h}}\big((Z_t^1, Z_t^2)^{\uparrow} \to (0, n(x+h))\big) \cdot P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}(T^{(x+h)}_{\lfloor \varepsilon K \rfloor} < T_0^{(x+h)}) \\
&= 1_{\{Ef(x,x+h)<0\}} \cdot \lim_{K\to\infty} K \cdot P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}(T^{(x+h)}_{\lfloor \varepsilon K \rfloor} < T_0^{(x+h)}) \\
&= \lim_{K\to\infty} K \cdot P^{\omega}_{z\delta_x + \frac{1}{K}\delta_{x+h}}(T^{(x+h)}_{\lfloor \varepsilon K \rfloor} < T_0^{(x+h)})
\end{aligned}
\tag{4.5.25}
$$

where the second equality is due to Lemma 4.5.7 and the third equality is due to Proposition 4.5.4 (ii). By substituting back into (4.5.24), the proof ends up.

For Lemma 4.5.1 (iii), one gets $b(t,\omega) \equiv Kr(x+h) + b(x+h)$ and $d(t,\omega) = Kr(x+$

$h) + d(x+h) + \alpha(x+h,x)\omega_t(x)$. From Lemma 4.5.2 (ii), one obtains, for $Q^{(x)}-$ a.s.,

$$\begin{aligned}
&\lim_{K\to\infty} K \cdot P^{\omega}_{z\delta_x+\frac{1}{K}\delta_{x+h}}(T^{(x+h)}_{\lfloor \varepsilon K\rfloor} < T^{(x+h)}_0) \\
&\qquad = \lim_{K\to\infty} K(1-q_\omega) \\
&\qquad = \lim_{K\to\infty} K\Big(1 - \frac{\int_0^\infty d(t,\omega) e^{\int_0^t (d(\tau,\omega)-b(\tau,\omega))d\tau} dt}{1+\int_0^\infty d(t,\omega) e^{\int_0^t (d(\tau,\omega)-b(\tau,\omega))d\tau} dt}\Big) \\
&\qquad = \lim_{K\to\infty} K\Big(1 + \int_0^\infty (Kr(x+h) + d(x+h) + \alpha(x+h,x)\omega_t) \\
&\qquad\qquad \times \exp\big\{ - (tb(x+h) - td(x+h) - \int_0^t \alpha(x+h,x)\omega_\tau(x) d\tau)\,\big\} dt\Big)^{-1} \\
&\qquad = \Big(r(x+h)\int_0^\infty \exp\big\{ - (tb(x+h) - td(x+h) - \int_0^t \alpha(x+h,x)\omega_\tau(x) d\tau)\,\big\} dt\Big)^{-1}.
\end{aligned}$$

It eventually follows the conclusion by substituting above equation back into (4.5.25).
□

Lemma 4.5.9. *Under the same condition as in Theorem* 4.5.5, $\left(\beta_1^K, \left\langle X^K_{\frac{\beta_1^K-}{Ku_K}}, 1\right\rangle\right)$ *conditioned on event* $\left\{\left\langle X^K_{\frac{\beta_1^K}{Ku_K}}, 1\right\rangle > 0\right\}$ *converges to a couple of independent random variables* $(\beta_1,\, \xi(x))$*, where* β_1 *is exponentially distributed with parameter* $\gamma(x)$ *defined in* (4.2.5)*, and* $\xi(x)$ *is obtained as a biased distribution of* $n(x)$*:*

$$\mathbb{P}\left(\xi(x) \in dz\right) = \frac{z g_x(z)\pi^{(x)}(dz)}{\mathbb{E}\left[n(x) g_x(n(x))\right]} \tag{4.5.26}$$

.

Proof. Denote by $Z^K_{\frac{t}{Ku_K}} = \left\langle X^K_{\frac{t}{Ku_K}}, 1_{\{x\}}\right\rangle$ for $t < \beta_1^K$. For any bounded measurable function ϕ on $\mathbb{R}^*_+$,

$$\begin{aligned}
&\lim_{K\to\infty}\ \mathbb{E}\left[\phi\left(\left\langle X^K_{\frac{\beta_1^K-}{Ku_K}}, 1\right\rangle\right), \beta_1^K < t \;\middle|\; \left\langle X^K_{\frac{\beta_1^K}{Ku_K}}, 1\right\rangle > 0\right] \\
&= \lim_{K\to\infty}\lim_{\Delta t\to 0}\ \mathbb{E}\left[\phi\left(Z^K_{\frac{\beta_1^K-\Delta t}{Ku_K}}\right), \beta_1^K < t \;\middle|\; Z^K_{\frac{\beta_1^K-\Delta t}{Ku_K}+\frac{\Delta t}{Ku_K}} > 0\right] \\
&= \lim_{K\to\infty}\ \mu(x)r(x)\int_0^t ds \mathbb{E}\Big[\phi\left(Z^K_{\frac{s}{Ku_K}}{}^{\uparrow}\right) Z^K_{\frac{s}{Ku_K}}{}^{\uparrow} g_x\left(Z^K_{\frac{s}{Ku_K}}{}^{\uparrow}\right) \\
&\qquad \exp\Big(- Ku_K\mu(x)r(x)\int_0^{s/Ku_K} Z_u^{K\uparrow} g_x(Z_u^{K\uparrow}) du\Big)\Big]
\end{aligned} \tag{4.5.27}$$

$$\begin{aligned} &= \mu(x)r(x)\int_0^t ds\mathbb{E}\left[\phi(n(x))n(x)g_x(n(x))\right]\exp\left(-s\mu(x)r(x)\mathbb{E}\left[n(x)g_x(n(x))\right]\right) \\ &= \frac{\mathbb{E}\left[\phi(n(x))n(x)g_x(n(x))\right]}{\mathbb{E}\left[n(x)g_x(n(x))\right]}\int_0^t \gamma(x)e^{-\gamma(x)s}ds \end{aligned} \tag{4.5.28}$$

where the second equality above is due to definition of Q-process. The third equality is due to Lemma 4.5.7, and one obtains $Ku_K\int_0^{s/Ku_K} Z_u^{K\uparrow}g_x(Z_u^{K\uparrow}) \to \mathbb{E}[n(x)g_x(n(x))]$ a.s. as $K \to \infty$ by applying ergodic theorem on the Q-process. $\square$

Proof of Theorem 4.5.5. Let $\mathbf{P}_x$ denote the law of V_t as defined in Definition 4.2.2 with $\mathrm{Supp}(V_0) = \eta_0 = x$. For any $i \geq 1$, any fixed $t > 0$, measurable subset $A \subset \mathbb{R}_+^*$ and $B \subset \mathcal{X}$, applying the strong Markov property at β_1

$$\begin{aligned} &\mathbf{P}_x\left(\beta_i < t < \beta_{i+1}, \exists y \in B : V_t = n(y)\delta_y, n(y) \in A\right) \\ &= \int_0^t \gamma(x)e^{-s\gamma(x)}ds\int_{\mathbb{R}^d}\mathbf{P}_{x+h}\left(\beta_{i-1} < t-s < \beta_i, \exists y \in B : V_t = n(y)\delta_y, n(y) \in A\right) \\ &\qquad \times \mathbb{E}\left[\frac{g_x(\xi(x),h)}{g_x(\xi(x))}\right]m(x,dh). \end{aligned} \tag{4.5.29}$$

In particular,

$$\mathbf{P}_x\left(0 < t < \beta_1, \exists y \in B, n(y) \in A\right) = 1_{\{x\in B\}}e^{-t\gamma(x)} \times \mathbf{P}(n(x) \in A). \tag{4.5.30}$$

A finite dimensional distribution of the process V can be fully characterized by the two relations above. The idea to prove the Theorem is to show that the same relations hold for process $X^K_{\cdot/Ku_K}$, conditioned on some event, when we replace β_i by β_i^K and the support of V_t by the support of $X^K_{\cdot/Ku_K}$ in the limit $K \to \infty$. More precisely, we have the following claim, which will be proved later.

Claim 4.5.10. *Admit same conditions as in Theorem 4.5.5, and* $\beta_0^K = 0$. *For any* $i \geq 0$,

$$\begin{aligned} p_i(t,A,B,x) \overset{def}{=} \lim_{K\to\infty}\mathbb{P}_{\frac{N_0^K}{K}\delta_x}\Big(&\beta_i^K < t < \beta_{i+1}^K, \exists y \in B : X^K_{\frac{t}{Ku_K}} = \left\langle X^K_{\frac{t}{Ku_K}}, 1\right\rangle\delta_y, \\ &\left\langle X^K_{\frac{t}{Ku_K}}, 1\right\rangle \in A \,\Big|\, \left\langle X^K_{\frac{\beta_{i+1}^K}{Ku_K}}, 1\right\rangle > 0\Big) \end{aligned}$$

exists and satisfies the following relation, for $i \geq 1$,

$$p_i(t,A,B,x) = \int_0^t \gamma(x)e^{-\gamma(x)s}ds\int_{\mathbb{R}^d}p_{i-1}(t-s,A,B,x+h)\mathbb{E}\left[\frac{g_x(\xi(x),h)}{g_x(\xi(x))}\right]m(x,dh) \tag{4.5.31}$$

$$p_0(t,A,B,x) = 1_{\{x\in B\}}e^{-t\gamma(x)} \times \mathbb{P}(n(x) \in A). \tag{4.5.32}$$

By comparing chain relations (4.5.29), (4.5.30) with (4.5.31) and (4.5.32), one obtains that

$$p_i(t,A,B,x) = \mathbf{P}_x\left(\beta_i < t < \beta_{i+1}, \exists y \in B : V_t = n(y)\delta_y, n(y) \in A\right), \qquad i \geq 0. \quad (4.5.33)$$

For any measurable subsets $A_1, \dots, A_l$ of $\mathbb{R}_+^*$, $B_1, B_2, \dots, B_l$ of $\mathcal{X}$,

$$\begin{aligned}
&\lim_{K\to\infty} \mathbb{P}\Big(\text{there exist merely } l \text{ successful mutations before time } t\text{, such that } \forall 1 \leq i \leq l, \\
&\qquad \beta_i^K \in [t_{i-1}, t_i);\ \exists x_i \in B_i : \mathrm{Supp}(X^K_{t_i/Ku_K}) = x_i;\ \langle X^K_{t_i/Ku_K}, 1\rangle \in A_i \ \Big|\ \langle X^K_{t/Ku_K}, 1\rangle > 0\Big) \\
&= \lim_{K\to\infty} \mathbb{P}\Big(\forall 0 \leq i \leq l-1,\ \beta_i^K < t_i \leq \beta_{i+1}^K; \\
&\qquad \exists x_i \in B_i : \mathrm{Supp}(X^K_{t_i/Ku_K}) = x_i;\ \langle X^K_{t_i/Ku_K}, 1\rangle \in A_i \ \Big|\ \langle X^K_{t/Ku_K}, 1\rangle > 0\Big) \\
&= \lim_{K\to\infty} \prod_{i=0}^{l-1} \mathbb{P}\Big(\beta_i^K < t_i \leq \beta_{i+1}^K;\ \exists x_i \in B_i : \mathrm{Supp}(X^K_{t_i/Ku_K}) = x_i;\ \langle X^K_{t_i/Ku_K}, 1\rangle \in A_i; \\
&\qquad \langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0;\ \langle X^K_{t/Ku_K}, 1\rangle > 0\Big) \Big/ \mathbb{P}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right) \\
&= \lim_{K\to\infty} \prod_{i=0}^{l-1} \Big[\mathbb{P}\Big(\beta_i^K < t_i \leq \beta_{i+1}^K;\ \exists x_i \in B_i : \mathrm{Supp}(X^K_{t_i/Ku_K}) = x_i;\ \langle X^K_{t_i/Ku_K}, 1\rangle \in A_i \\
&\qquad \Big|\ \langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0\Big) \cdot \mathbb{P}_{X^K_{\beta^K_i/Ku_K}}\Big(\langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0\Big)\Big] \\
&\qquad \cdot \mathbb{P}_{X^K_{\beta^K_l/Ku_K}}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right) \Big/ \mathbb{P}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right) \\
&= \prod_{i=0}^{l-1} \mathbf{P}\Big(\beta_i < t_i \leq \beta_{i+1};\ \eta_{t_i} \in B_i;\ n(\eta_{t_i}) \in A_i\Big) \cdot \lim_{K\to\infty} \frac{\prod_{i=0}^{l-1} \mathbb{P}_{X^K_{\beta^K_i/Ku_K}}\Big(\langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0\Big)}{\mathbb{P}\Big(\langle X^K_{t/Ku_K}, 1\rangle > 0\Big)} \\
&\qquad \cdot \mathbb{P}_{X^K_{\beta^K_l/Ku_K}}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right) \\
&= \mathbf{P}\Big(\forall 0 \leq i \leq l-1,\ \beta_i < t_i \leq \beta_{i+1};\ \eta_{t_i} \in B_i;\ n(\eta_{t_i}) \in A_i\Big) \\
&= \mathbf{P}\Big(\text{there exist merely } l \text{ successful mutations before time } t\text{, such that } \forall 1 \leq i \leq l, \\
&\qquad \beta_i \in [t_{i-1}, t_i);\ \eta_{t_i} \in B_i;\ n(\eta_{t_i}) \in A_i\Big).
\end{aligned}$$

In above arguments, the third equal sign is due to strong Markov property at β_i^K/Ku_K for any $1 \leq i \leq l$, the fourth one can be deduced from (4.5.33), and the fifth equal sign arises from the following argument.

In fact, conditioned on $\mathrm{Supp}(X^K_{t_i/Ku_K}) = x_i$ for $0 \leq i \leq l-1$,

$$\mathbb{P}_{X^K_{\beta^K_i/Ku_K}}\Big(\langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0\Big)$$

can be expressed in terms of the killing rate $\kappa(x_i)$ which is determined by spectral analysis developed in [7], i.e.

$$\mathbb{P}_{X^K_{\beta^K_i/Ku_K}}\left(\langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0\right) \simeq e^{-\kappa(x_i)\frac{\beta^K_{i+1}-\beta^K_i}{Ku_K}} \text{ for large } K, \tag{4.5.34}$$

$$\mathbb{P}_{X^K_{\beta^K_l/Ku_K}}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right) \simeq e^{-\kappa(x_l)\frac{t-\beta^K_l}{Ku_K}} \text{ for large } K. \tag{4.5.35}$$

Due to assumption (B), one can conclude that the killing rate of any type is the same constant κ. For instance, as in previous approximation, $\kappa(x_i) \equiv \kappa$ for $1 \leq i \leq l$. Furthermore,

$$\begin{aligned}\mathbb{P}\left(\langle X^K_{\frac{t}{Ku_K}}, 1\rangle > 0\right) &\simeq \sum_{l=1}^{\infty} m_l \exp\left(-\sum_{j=0}^{l-1}\kappa(x^{(l)}_j)(\beta^K_{j+1}-\beta^K_j)/Ku_K\right)\\ &\quad\cdot \exp\left(-\kappa(x^{(l)}_l)(t-\beta^K_l)/Ku_K\right)\\ &\equiv \sum_{l=1}^{\infty} m_l \exp\left(-\kappa t/Ku_K\right)\\ &= e^{-\kappa t/Ku_K}\end{aligned} \tag{4.5.36}$$

where $\{m_l\}_{l\in\mathbb{N}}$ denotes the probability in which the number of successful mutations before time t/Ku_K is l, and $x^{(l)}_j$, $1 \leq j \leq l$, denote the l successful mutant traits. Thus, one can conclude

$$\begin{aligned}&\lim_{K\to\infty}\frac{\prod_{i=0}^{l-1}\mathbb{P}_{X^K_{\beta^K_i/Ku_K}}\left(\langle X^K_{\beta^K_{i+1}/Ku_K}, 1\rangle > 0\right)}{\mathbb{P}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right)}\cdot \mathbb{P}_{X^K_{\beta^K_l/Ku_K}}\left(\langle X^K_{t/Ku_K}, 1\rangle > 0\right)\\ &= \lim_{K\to\infty}\frac{\prod_{i=0}^{l-1} e^{-\kappa(x_i)\frac{\beta^K_{i+1}-\beta^K_i}{Ku_K}}\cdot e^{-\kappa(x_l)\frac{t-\beta^K_l}{Ku_K}}}{e^{-\kappa t/Ku_K}}\\ &= 1\end{aligned} \tag{4.5.37}$$

□

To close the proof of the Theorem, we just need to prove the Claim stated in above proof by induction methods over $i \geq 0$.

Proof of Claim 4.5.10. We first prove an easy result.

$$\begin{aligned}&p_0(t, A, B, x)\\ &\overset{\text{def}}{=} \lim_{K\to\infty}\mathbb{P}_{\frac{N^K_0}{K}\delta_x}\left(0 < t < \beta^K_1,\ \exists y \in B : X^K_{\frac{t}{Ku_K}} = \left\langle X^K_{\frac{t}{Ku_K}}, 1\right\rangle\delta_y,\ \left\langle X^K_{\frac{t}{Ku_K}}, 1\right\rangle \in A\right.\\ &\qquad\qquad\left.\Big|\ \left\langle X^K_{\frac{\beta^K_1}{Ku_K}}, 1\right\rangle > 0\right)\end{aligned}$$

$$
\begin{aligned}
&= \lim_{K\to\infty} 1_{\{x\in B\}} \mathbb{E}\left[1_{\left\{\left\langle X^K_{\frac{t}{Ku_K}},1\right\rangle\in A\right\}} \exp\left(-\mu(x)r(x)Ku_K\int_0^{t/Ku_K} \langle X^K_s,1\rangle g_x(\langle X^K_s,1\rangle)ds\right) \right.\\
&\qquad\qquad \left. \Bigm| \left\langle X^K_{\frac{\beta^K_1}{Ku_K}},1\right\rangle > 0\right]\\
&= 1_{\{x\in B\}}\mathbb{P}(n(x)\in A)\exp\left(-t\mu(x)r(x)\mathbb{E}\left[n(x)g_x(n(x))\right]\right),
\end{aligned}
$$

in the last line where it holds due to the ergodic theorem .

Let

$$
p^K_i(t,A,B,\nu) := \mathbb{P}_\nu\Big(\beta^K_i < t < \beta^K_{i+1}, \exists y\in B : X^K_{\frac{t}{Ku_K}} = \left\langle X^K_{\frac{t}{Ku_K}},1\right\rangle\delta_y,
$$

$$
\left\langle X^K_{\frac{t}{Ku_K}},1\right\rangle \in A \Bigm| \left\langle X^K_{\frac{\beta^K_{i+1}}{Ku_K}},1\right\rangle > 0\Big).
$$

Suppose that $p_{i-1}(t,A,B,x) = \lim_{K\to\infty} p^K_{i-1}(t,A,B,n\delta_x)$ exists. To the end, we need prove $\lim_{K\to\infty} p^K_i(t,A,B,n\delta_x)$ exists.

By applying strong Markov property two times,we have

$$
\begin{aligned}
p^K_i(t,A,B,n\delta_x) &= \int_{\mathbb{R}^d} \mathbb{E}_{n\delta_x}\left[1_{\{\beta^K_1<t\}}\cdot p^K_{i-1}\left(t-\beta^K_1, A, B, \left\langle X^K_{\frac{\beta^K_1-}{Ku_K}},1\right\rangle\delta_x + \frac{1}{K}\delta_{x+h}\right)\right.\\
&\qquad \left. \cdot \frac{g_x\left(\left\langle X^K_{\frac{\beta^K_1-}{Ku_K}},1\right\rangle, h\right)}{g_x\left(\left\langle X^K_{\frac{\beta^K_1-}{Ku_K}},1\right\rangle\right)} \cdot m(x,dh) \Bigm| \left\langle X^K_{\frac{\beta^K_{i+1}}{Ku_K}},1\right\rangle > 0\right]\\
&= \int_{\mathbb{R}^d} \mathbb{E}_{n\delta_x}\left[1_{\{\beta^K_1<t\}}\cdot p^K_{i-1}\left(t-\beta^K_1 - Ku_K\rho_0, A, B, \langle X^K_{\rho_0},1\rangle\delta_{x+h}\right)\right.\\
&\qquad \left. \cdot \frac{g_x\left(\left\langle X^K_{\frac{\beta^K_1-}{Ku_K}},1\right\rangle, h\right)}{g_x\left(\left\langle X^K_{\frac{\beta^K_1-}{Ku_K}},1\right\rangle\right)} \Bigm| \left\langle X^K_{\frac{\beta^K_{i+1}}{Ku_K}},1\right\rangle > 0\right] m(x,dh),
\end{aligned}
\tag{4.5.38}
$$

where ρ_0 is the fixation time of type $x+h$.

Applying Lemma 4.5.9 to terms on RHS of the above equation, we obtain

$$
\begin{aligned}
\lim_{K\to\infty} p^K_i(t,A,B,n\delta_x) &= \int_0^t ds\gamma(x)e^{-s\gamma(x)}\\
&\cdot \int_{\mathbb{R}^d} p_{i-1}(t-s,A,B,x+h)\mathbb{E}\left[\frac{g_x(\xi(x),h)}{g_x(\xi(x))}\right] m(x,dh).
\end{aligned}
\tag{4.5.39}
$$

□

5 Fluctuation limit of a locally regulated population

We consider a locally regulated spatial population model introduced by Bolker and Pacala. Based on the deterministic approximation obtained by Fournier and Méléard, a fluctuation theorem is proved under a second order moment condition. The limiting process is justified to be an infinite-dimensional Gaussian process solving a generalized Langevin equation. In particular, we further study its properties in the one dimensional setting. Finally, we consider the stationary behavior of the Bolker-Pacala processes and its fluctuation limit under some specific conditions. We prove that the fluctuation limit and the long term limit are interchangeable.

5.1 Introduction

It is well known that branching processes have been widely used to model the evolution in biological populations. If, in addition, the individuals are assumed to follow some independent motions (like Brownian motion or random walks), the system can be approximated by the so-called Dawson-Watanabe superprocess (refer to [15, 18, 37]). The most common feature of these processes is that branching and spatial motion are independent.

Since individuals can reproduce, mutate and die in varying rates according to their different quantitative characteristics (phenotypes), one reasonable improvement we can make is to add spatial components to both branching and dispersal parameters. Nevertheless, the spatial-dependent components destroy the independent relationship between branching and dispersal while bringing us abundant information from the phenotypic point of view, and even though, the model is still deficient: such as in the finite-dimensional branching process model, a population either dies out or escapes to infinity, depending on the mean matrix of its offspring distribution. The model thus can not predict a non-trivial equilibrium which actually happens quite often in the biological world. Bolker and Pacala [4] propose a self-regulated model which attains the above two improved features. By employing the idea of the ordinary logistic growth equation, they introduce a competition term in the density-dependent population, which can help us to build some equilibria under specific conditions. However, the loss of branching property can also cause some new technical difficulties when we study some properties such as weak convergence from branching particle systems to a continuum limit.

Law and Dieckmann [35] study this model in parallel with Bolker and Pacala [4]. We simply call it BPDL model. In recent years, this model has been extensively studied by

Etheridge [19], Fournier and Méléard [22], Champagnat [8], Lambert [33], Dawson and Greven [16]. Etheridge [19] studies two diffusion limits, one is a *stepping stone version* of the BPDL model (interacting diffusions indexed by $\mathbb{Z}^d$) and another is a *superprocess version* of it. In that paper, sufficient conditions are given for survival and local extinction. Fournier and Méléard [22] formulate a pathwise construction of the BPDL process in terms of Poisson point processes. Under the finiteness of third moment condition, they rigorously obtain a deterministic approximation (law of large numbers) of the BPDL processes from a macroscopic perspective. Our work originates from Bolker and Pacala [4] and is based on the formalization of Fournier and Méléard [22]. In the papers Champagnat [8], Champagnat and Méléard [10], Dawson and Greven [16], they investigate long term behaviors by the method of multiple time scale analysis in respective populations.

In this paper we aim to present and prove the fluctuation limits in an unified general framework, which covers the cases mentioned above, including applications in the derivative models studied by referred authors. As for a sequence of density-dependent population processes with only finite-many types of individuals, Kurtz [32] proves its central limit theorem, which is characterized by some finite-dimensional diffusion process. As for infinite-dimensional population models, Gorostiza and Li [24] prove the high-density fluctuations of a branching particle system with immigration, where they use the classical Laplace transform method owing to the branching property. In our setting, this approach doesn't work anymore due to the loss of branching property.

The remainder of the paper is structured as follows. In Section 5.2, we list some preliminary results on the model and give the fluctuation theorem. We recall the law of large numbers of the BPDL processes under a finite second moment rather than a third moment condition as in Fournier and Méléard [22]. In Section 5.3, in order to better understand the limiting process, we justify it to be the solution of an infinite-dimensional inhomogeneous Langevin equation, which can be viewed as living in a deterministic medium. Concerning applications, we will explore them in two directions as following. In Section 5.3.2, we consider a degenerate case, the one dimensional version of the fluctuation limit. A precise characterization of the fluctuating diffusion is given as a time-inhomogeneous Ornstein-Uhlenbeck process. We study its stationary distribution as well. In Section 5.4, we consider the stationary behavior in a specific setting. The solution of the above Langevin equation turns out to have a well-defined stationary limit. We show that this limit coincides with the one obtained by first letting time tending to infinity in the original BPDL processes and then attaining its fluctuation limit. In some sense, this interchangeability of limits also extends our fluctuation theorem from time interval $[0, \infty)$ to time interval $[0, \infty]$. In Section 5.5, we provide the proofs on the tightness and the finite-dimensional convergence based on moment estimates in subsequent sections.

5.2 Model and main results

5.2.1 Notations and description of the processes

Following [4], we assume the population at time t is composed of a finite number $I(t)$ of individuals characterized by their phenotypic traits $x_1(t), \cdots, x_{I(t)}(t)$ taking values in a compact subset $\mathcal{X}$ of $\mathbb{R}^d$.
We denote by $\mathcal{M}_F(\mathcal{X})$ the set of finite measures on $\mathcal{X}$ (including negative measures). Let $\mathcal{M}_a(\mathcal{X}) \subset \mathcal{M}_F(\mathcal{X})$ be the set of counting measures on $\mathcal{X}$:

$$\mathcal{M}_a(\mathcal{X}) = \left\{ \sum_{i=1}^{n} \delta_{x_i} : x_1, \cdots, x_n \in \mathcal{X}, n \in \mathbb{N} \right\}.$$

Then, the population process can be represented as:

$$\nu_t = \sum_{i=1}^{I(t)} \delta_{X_i(t)}.$$

Let $B(\mathcal{X}), C^\infty(\mathcal{X})$ denote the totality of functions on $\mathcal{X}$ that are bounded measurable, and infinitely differentiable, respectively. Let $\mathcal{S}(\mathcal{X})$ denote the Schwartz space of (infinitely differentiable, rapidly decreasing) testing functions on $\mathcal{X}$ whose topological dual space is $\mathcal{S}'(\mathcal{X})$, and $\langle \cdot, \cdot \rangle$ the canonical bilinear form on $\mathcal{S}'(\mathcal{X}) \times \mathcal{S}(\mathcal{X})$. When $\mu \in \mathcal{S}'(\mathcal{X})$ is a (signed) measure, then $\langle \mu, \phi \rangle = \int \phi d\mu$, $\phi \in \mathcal{S}(\mathcal{X})$.
Let's specify the population processes $(\nu_t^n)_{t>0}$ by introducing a sequence of biological parameters, for n$\in \mathbb{N}$:

- $b_n(x)$ is the rate of birth from an individual with trait x.
- $d_n(x)$ is the rate of death of an individual with trait x because of "aging".
- $\alpha_n(x, y)$ is the competition kernel felt by some individual with trait x from another individual with trait y.
- $D_n(x, dz)$ is the children's dispersion law from the mother with trait x. In particular, it can be decomposed into two parts-a majority of clonal copies and a small portion of mutants based on birth, i.e.

  $$D_n(x, dz) = (1 - \epsilon)1_{z=0} + \epsilon m_n(x, dz)1_{z\neq 0}. \tag{5.2.1}$$

  Here, $m_n(x, dz)$ is the probability density of mutation, which satisfies

  $$\int_{z\in\mathbb{R}^d, x+z\in\mathcal{X}} m_n(x, dz) = 1.$$

  We will omit the superscript ϵ in D_n in the sequel when this leads no ambiguity.

Fournier and Méléard [22] has formulated a pathwise construction of the BPDL process $\{(\nu_t^n)_{t\geq 0}; n \in \mathbb{N}\}$ in terms of Poisson random measures and justified its infinitesimal generator defined for any $\Phi \in B(\mathcal{M}_a(\mathcal{X}))$:

$$\begin{aligned} L_0^n\Phi(\nu) = &\int_{\mathcal{X}} \nu(dx) \int_{\mathbb{R}^d} \Big(\Phi(\nu + \delta_{x+z}) - \Phi(\nu)\Big) b_n(x) D_n(x, dz) \\ &+ \int_{\mathcal{X}} \nu(dx) \Big(\Phi(\nu - \delta_x) - \Phi(\mu)\Big) \left(d_n(x) + \int_{\mathcal{X}} \alpha_n(x, y)\nu(dy)\right). \end{aligned} \tag{5.2.2}$$

The first term is used to model birth events, while the second term which is nonlinear is interpreted as natural death and competing death.
Instead of studying the original BPDL processes defined by (5.2.2), our goal is to study the rescaled processes

$$X_t^n := \frac{\nu_t^n}{n}, \qquad t \geq 0 \tag{5.2.3}$$

since it provides us a macroscopic approximation when we take the large population limits (we will see later, the initial population is proportional to n in some sense). The infinitesimal generator of the rescaled BPDL process has the form, for any $\Phi \in B(\mathcal{M}_F(\mathcal{X}))$:

$$\begin{aligned} L^n\Phi(\mu) = &\int_{\mathcal{X}} n\mu(dx) \int_{\mathbb{R}^d} \Big(\Phi(\mu + \frac{\delta_{x+z}}{n}) - \Phi(\mu)\Big) b_n(x) D_n(x, dz) \\ &+ \int_{\mathcal{X}} n\mu(dx) \Big(\Phi(\mu - \frac{\delta_x}{n}) - \Phi(\mu)\Big) \left(d_n(x) + \int_{\mathcal{X}} \alpha_n(x, y) n\mu(dy)\right). \end{aligned} \tag{5.2.4}$$

5.2.2 Preliminary results

Let's denote by (A) the following assumptions:

(A1) There exist $b(x)$, $d(x)$, $m(x, dy)$, $\alpha(x, y) \in C^\infty(\mathcal{X})$ such that, for x, $y \in \mathcal{X}$, $n \in \mathbb{N}$,

$$\begin{aligned} &0 < b_n(x) \equiv b(x), \qquad 0 < d_n(x) \equiv d(x), \qquad m_n(x, dz) \equiv m(x, dz), \\ &0 < \alpha_n(x, y) = \frac{\alpha(x, y)}{n}. \end{aligned}$$

(A2) $b(x) - d(x) > 0$.

The first assumption implies that there exist constants $\bar{b}$, $\bar{d}$, $\bar{\alpha}$ such that $b(x) \leq \bar{b}$, $d(x) \leq \bar{d}$, $\alpha(x, y) \leq \bar{\alpha}$.

By neglecting the high order moment, Bolker and Pacala [4] use the "moment closure" procedure to approximate the stochastic population processes. As we can see from the generator formula (5.2.4), it should be enough to "close" the second order moment due to the quadratic nonlinear term. Actually, we indeed can improve the result of Fournier and Méléard by giving a second moment condition $\sup_{n\geq 1} \mathbb{E}\langle X_0^n, 1\rangle^2 < \infty$ rather than the finiteness of the third moment condition. Since there is no essential difficulty in the improved proof, we only list the result here without giving the detailed proof repeatedly.

Theorem 5.2.1 (Fournier and Méléard [22], convergence to an integrodifferenial equation)**.** *Under the assumption (A1), and consider the sequence of processes* $(X^n_t)_{t\geq 0}$ *defined in* (5.2.3)*. Suppose that* (X^n_0) *converges in law to some deterministic finite measure* $X_0 \in \mathcal{M}_F(\mathcal{X})$ *as* $n \to \infty$ *and satisfies* $\sup\limits_{n\geq 1} \mathbb{E}\langle X^n_0, 1\rangle^2 < \infty$*. Then, a sequence of processes* $(X^n_t)_{t\geq 0}$ *converges in law as* $n \to \infty$*, on* $D([0,\infty), \mathcal{M}_F(\mathcal{X}))$*, to a deterministic measure-valued process* $(X_t)_{t\geq 0} \in C([0,\infty), \mathcal{M}_F(\mathcal{X}))$*, where* $(X_t)_{t\geq 0}$ *is an unique solution satisfying*

$$\sup_{t\in[0,T]} \langle X_t, 1\rangle < \infty, \tag{5.2.5}$$

and for any $\phi \in B(\mathcal{X})$*,*

$$\begin{aligned}\langle X_t, \phi\rangle =& \langle X_0, \phi\rangle + \int_0^t ds \int_{\mathcal{X}} X_s(dx) b(x) \int_{\mathbb{R}^d} \phi(x+z) D(x, dz) \\ &- \int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi(x)\Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s(dy)\Big).\end{aligned} \tag{5.2.6}$$

Finally, it turns out a natural question: how does $(X^n_t)_{t\geq 0}$ fluctuate around the macroscopic limit $(X_t)_{t\geq 0}$ given above? A natural candidate to investigate is the centralized processes sequence:

$$Y^n_t := \frac{\nu^n_t - nX_t}{\sqrt{n}} = \sqrt{n}(X^n_t - X_t). \tag{5.2.7}$$

In the following proposition, we will give some martingale properties of the processes $(Y^n_t)_{t\geq 0}$, which will play a key role in the proof of the main theorem.

Proposition 5.2.2. *Admit the same assumptions as in Theorem 5.2.1. For fixed* $n \in \mathbb{N}$ *and* $\phi \in B(\mathcal{X})$*, the process*

$$\begin{aligned}M^n_t(\phi) :=& \langle Y^n_t, \phi\rangle - \langle Y^n_0, \phi\rangle - \int_0^t ds \int_{\mathcal{X}} Y^n_s(dx) b(x) \int_{\mathbb{R}^d} \phi(x+z) D(x,dz) \\ &+ \int_0^t ds \int_{\mathcal{X}} \phi(x) d(x) Y^n_s(dx) \\ &+ \sqrt{n} \int_0^t ds \int_{\mathcal{X}} X^n_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) X^n_s(dy) \\ &- \sqrt{n} \int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) X_s(dy)\end{aligned} \tag{5.2.8}$$

is a càdlàg square integrable martingale with quadratic variation

$$\begin{aligned}\langle M^n_\cdot(\phi)\rangle_t =& \int_0^t ds \int_{\mathcal{X}} X^n_s(dx) b(x) \int_{\mathbb{R}^d} \phi^2(x+z) D(x,dz) \\ &+ \int_0^t ds \int_{\mathcal{X}} X^n_s(dx)\phi^2(x)\Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X^n_s(dy)\Big).\end{aligned} \tag{5.2.9}$$

Proof. Recall the generator (5.2.4), for bounded measurable functional Φ on $\mathcal{M}_F(\mathcal{X})$, the process

$$\Phi(X_t^n) - \Phi(X_0^n) - \int_0^t L^n \Phi(X_s^n) ds$$

is a càdlàg square integrable martingale. If we take $\Phi(\mu) = \langle \mu, \phi \rangle$, $\forall \phi \in B(\mathcal{X})$, one obtains that

$$\begin{aligned} N_t^n(\phi) :=& \langle X_t^n, \phi \rangle - \langle X_0^n, \phi \rangle - \int_0^t ds \int_{\mathcal{X}} X_s^n(dx) b(x) \int_{\mathbb{R}^d} \phi(x+z) D(x,dz) \\ &+ \int_0^t ds \int_{\mathcal{X}} X_s^n(dx) \phi(x) \Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s^n(dy) \Big) \end{aligned} \tag{5.2.10}$$

is a c$\grave{a}$dl$\grave{a}$g martingale. By applying It$\hat{o}$'s formula to $\langle X_t^n, \phi \rangle^2$, we have

$$\begin{aligned} \langle X_t^n, \phi \rangle^2 - \langle X_0^n, \phi \rangle^2 - 2 \int_0^t ds \langle X_s^n, \phi \rangle \int_{\mathcal{X}} X_s^n(dx) \Big\{ b(x) \int_{\mathbb{R}^d} \phi(x+z) D(x,dz) \\ - \phi(x) \Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s^n(dy) \Big) \Big\} - \langle N_\cdot^n(\phi) \rangle_t \end{aligned}$$

is a martingale. On the other hand, if we take $\Phi(\mu) = \langle \mu, \phi \rangle^2$, it follows

$$\begin{aligned} &\langle X_t^n, \phi \rangle^2 - \langle X_0^n, \phi \rangle^2 - \int_0^t L^n \Phi(X_s^n) ds \\ &= \langle X_t^n, \phi \rangle^2 - \langle X_0^n, \phi \rangle^2 - 2 \int_0^t ds \langle X_s^n, \phi \rangle \int_{\mathcal{X}} X_s^n(dx) \Big\{ b(x) \int_{\mathbb{R}^d} \phi(x+z) D(x,dz) \\ &\quad - \phi(x) \Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s^n(dy) \Big) \Big\} \\ &\quad - \frac{1}{n} \int_0^t ds \int_{\mathcal{X}} X_s^n(dx) b(x) \int_{\mathbb{R}^d} \phi^2(x+z) D(x,dz) \\ &\quad - \frac{1}{n} \int_0^t ds \int_{\mathcal{X}} X_s^n(dx) \phi^2(x) \Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s^n(dy) \Big) \end{aligned} \tag{5.2.11}$$

is a martingale. By comparing the above two decompositions of semimartingale $\langle X_t^n, \phi \rangle^2$, one obtains

$$\begin{aligned} \langle N_\cdot^n(\phi) \rangle_t = & \frac{1}{n} \int_0^t ds \int_{\mathcal{X}} X_s^n(dx) b(x) \int_{\mathbb{R}^d} \phi^2(x+z) D(x,dz) \\ &+ \frac{1}{n} \int_0^t ds \int_{\mathcal{X}} X_s^n(dx) \phi^2(x) \Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s^n(dy) \Big). \end{aligned} \tag{5.2.12}$$

Owing to (5.2.10) and (5.2.6), do the operation $\big(\langle X_t^n, \phi \rangle - \langle X_t, \phi \rangle \big)$ and let $M_t^n(\phi) := \sqrt{n} N_t^n(\phi)$, we conclude the proof by the definition of (Y_t^n) in (5.2.7). □

5.2.3 Fluctuation theorem

In this section, our aim is to study the asymptotic behavior of $(Y_t^n)_{t\geq 0}$ as $n \to \infty$. The following theorem, the main result of the paper, shows that $(Y_t^n)_{t\geq 0}$ indeed converges to an unique solution of a martingale problem.

In the following sections, we will use the notations given in Section 5.2 without declaration. We always assume assumption (A1) holds.

Theorem 5.2.3. *Admit assumption (A1) and suppose there exists a deterministic finite nonnegative measure* $X_0 \in \mathcal{M}_F(\mathcal{X})$ *such that* $Y_0^n = \sqrt{n}(X_0^n - X_0)$, *for any* $\phi \in B(\mathcal{X})$, *satisfying*

$$\sup_{n\geq 1} \sup_{\|\phi\|_\infty \leq 1} \mathbb{E}\langle Y_0^n, \phi\rangle^2 < \infty. \tag{5.2.13}$$

Suppose that (Y_0^n) *converges in law to a finite (maybe random) measure* γ *as* $n \to \infty$.

Then, the process $(Y_t^n)_{t\geq 0}$ *converges in law as* $n \to \infty$ *on* $D([0,\infty), \mathcal{S}'(\mathcal{X}))$ *to a process* $(Y_t)_{t\geq 0} \in C([0,+\infty), \mathcal{S}'(\mathcal{X}))$ *where* $(Y_t)_{t\geq 0}$ *satisfies, for any* $\phi \in \mathcal{S}(\mathcal{X})$,

$$\begin{aligned} \langle Y_t, \phi\rangle = \langle \gamma, \phi\rangle &+ \int_0^t ds \Big\langle Y_s,\, b(\cdot)\int_{\mathbb{R}^d} \phi(\cdot + z) D(\cdot, dz)\Big\rangle \\ &- \int_0^t ds \Big\langle Y_s,\, d(\cdot)\phi(\cdot)\Big\rangle \\ &- \int_0^t ds \Big\langle Y_s,\, \int_{\mathcal{X}} \alpha(x,\cdot)\phi(x) X_s(dx)\Big\rangle \\ &- \int_0^t ds \Big\langle Y_s,\, \phi(\cdot)\int_{\mathcal{X}} \alpha(\cdot, y) X_s(dy)\Big\rangle \\ &+ M_t(\phi). \end{aligned} \tag{5.2.14}$$

Here, $(X_t)_{t\geq 0}$ *is the solution of the deterministic nonlinear equation* (5.2.6), *while* $M_t(\phi)$ *is a continuous martingale with quadratic variation*

$$\begin{aligned} \langle M_\cdot(\phi)\rangle_t = &\int_o^t ds \int_{\mathcal{X}} X_s(dx) b(x) \int_{\mathbb{R}^d} \phi^2(x+z) D(x, dz) \\ &+ \int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi^2(x)\Big(d(x) + \int_{\mathcal{X}} \alpha(x,y) X_s(dy)\Big). \end{aligned} \tag{5.2.15}$$

Remark 5.2.4. *The argument above makes essentially use of the initial moment* (5.2.13) *condition and initial convergence condition. These conditions fulfil the assumptions needed in Theorem 5.2.1. Therefore, the large number limit* $(X_t)_{t\geq 0}$ *is well defined (see Lemma 5.5.1).*

Proving the theorem is the content of the last section.

Corollary 5.2.5. *Given a deterministic nonnegative finite measure* X_0 *in* $\mathcal{M}_F(\mathcal{X})$, *for any* $n \geq 1$, *suppose the original BPDL process has initial state* $\nu_0^n \in \mathcal{M}(\mathcal{X})$ *which is a Poisson random measure with intensity measure* (nX_0). *Then, the same conclusion holds as in Theorem 5.2.3.*

It is easy to verify the conditions in Theorem 5.2.3. In fact,

$$\begin{aligned}\sup_{n\geq 1}\sup_{\|\phi\|_\infty\leq 1}\mathbb{E}\langle Y_0^n,\phi\rangle^2 &= \sup_{n\geq 1}\sup_{\|\phi\|_\infty\leq 1}\frac{1}{n}\mathbb{E}(\langle \nu_0^n,\phi\rangle - n\langle X_0,\phi\rangle)^2\\ &= \sup_{n\geq 1}\sup_{\|\phi\|_\infty\leq 1}\frac{1}{n}\times n\langle X_0,\phi\rangle\\ &= \sup_{\|\phi\|_\infty\leq 1}\langle X_0,\phi\rangle\\ &= \langle X_0,1\rangle<\infty.\end{aligned}$$

Furthermore, $Y_0^n = \frac{\nu_0^n - nX_0}{\sqrt{n}}$ converges in law as $n\to\infty$ to an isonormal Gaussian random measure with variance intensity measure X_0 on $\mathcal{X}$, i.e., $\mathcal{N}(0,X_0(dx))$.

5.3 Links with a generalized Langevin equation

5.3.1 Statement of the result

A criterion for an infinite-dimensional Gaussian process (distribution-valued process) to satisfy a generalized Langevin equation is given in [3], where both of the evolution term and the white noise term are time-inhomogeneous. In this section, we apply the criterion to our fluctuation limit obtained in previous section.

Definition 5.3.1. *An $\mathcal{S}'(\mathcal{X})$-valued process $\{W_t; t\in\mathbb{R}^+\}$ is called (centered) Gaussian if $\{\langle W_t,\phi\rangle; t\in\mathbb{R}^+,\phi\in\mathcal{S}(\mathcal{X})\}$ is a (centered) Gaussian system.*

Definition 5.3.2. *A centered Gaussian $\mathcal{S}'(\mathcal{X})$-valued process $W=\{W_t;t\in\mathbb{R}^+\}$ is called a generalized Wiener process if it has continuous path and its covariance functional $C(s,\phi;t,\psi) := \mathbb{E}[\langle W_s,\phi\rangle\langle W_t,\phi\rangle]$ has the form*

$$C(s,\phi;t,\psi) = \int_0^{s\wedge t}\langle Q_u\phi,\psi\rangle du, \qquad s,t\in\mathbb{R}^+,\phi,\psi\in\mathcal{S}(\mathcal{X}), \tag{5.3.1}$$

where the operators $Q_u:\mathcal{S}(\mathcal{X})\to\mathcal{S}'(\mathcal{X})$ have the following properties:

1. *Q_u is linear, continuous, symmetric and positive for each $u\in\mathbb{R}^+$,*
2. *the function $u\to\langle Q_u\phi,\psi\rangle$ is right continuous with left limit for each $\phi,\psi\in\mathcal{S}(\mathcal{X})$.*

We then say that W is associated to Q.

Let's remind that we inherit the same notations as in Section 5.2.
Define $Q_t\phi\in\mathcal{S}'(\mathcal{X})$ for any $\phi\in\mathcal{S}(\mathcal{X})$ and $t\in\mathbb{R}^+$ by

$$\begin{aligned}\langle Q_t\phi,\psi\rangle := \int_{\mathcal{X}}X_t(dx)\Bigg[&b(x)\int_{\mathbb{R}^d}\phi(x+z)\psi(x+z)D(x,dz)\\ &+\phi(x)\psi(x)\Big(d(x)+\int_{\mathcal{X}}\alpha(x,y)X_t(dy)\Big)\Bigg], \qquad \text{for } \psi\in\mathcal{S}(\mathcal{X}).\end{aligned} \tag{5.3.2}$$

Recall the quadratic variation form of $M_t(\phi)$ in (5.2.15). It follows a direct fact that $\langle M_\cdot(\phi)\rangle_t = \int_0^t \langle Q_u\phi, \phi\rangle du$. Then, we have

Theorem 5.3.3. *The fluctuation limit process* $(Y_t)_{t\geq 0}$ *obtained in Theorem 5.2.3 is an unique solution of a time inhomogeneous Langevin equation*

$$\begin{cases} dY_t = A_t^* Y_t dt + dW_t, & t > 0 \\ Y_0 = \gamma \end{cases}, \tag{5.3.3}$$

where A_t^* *denotes the adjoint operator of* A_t *defined by*

$$\begin{aligned} A_t\phi(x) = b(x)\int_{\mathbb{R}^d} \phi(x+z)D(x,dz) - \phi(x)\big(d(x) + \int_{\mathcal{X}} \alpha(x,y)X_t(dy)\big) \\ - \int_{\mathcal{X}} \alpha(y,x)\phi(y)X_t(dy), \end{aligned} \tag{5.3.4}$$

and $(W_t)_{t\geq 0}$ *is an* $\mathcal{S}'(\mathcal{X})$*-valued Wiener process with covariance*

$$\mathbb{E}\big[\langle W_s, \phi\rangle\langle W_t, \psi\rangle\big] = \int_0^{s\wedge t} \langle Q_u\phi, \psi\rangle du, \qquad s,t \geq 0,\ \phi,\psi \in \mathcal{S}(\mathcal{X}). \tag{5.3.5}$$

Remark 5.3.4.

1. *An* $\mathcal{S}'(\mathcal{X})$*-valued process* $(Y_t)_{t\geq 0}$ *is said to be a solution of* (5.3.3) *if for each* $\phi \in \mathcal{S}(\mathcal{X})$,

$$\langle Y_t, \phi\rangle = \langle \gamma, \phi\rangle + \int_0^t \langle Y_u, A_u\phi\rangle du + \langle W_t, \phi\rangle, \qquad \text{for } t \in \mathbb{R}^+. \tag{5.3.6}$$

2. $(W_t)_{t\geq 0}$ *has independent increments but not stationary property since covariance functional* Q *depends on time.*

Proof. **Existence.** According to Theorem 5.2.3, the covariance functional of continuous martingale M_t on testing functions is deterministic, which implies $(M_t)_{t\geq 0}$ is a $\mathcal{S}'(\mathcal{X})$-valued mean zero Gaussian process (see Walsh[44, Proposition 2.10]). Hence, $(Y_t)_{t\geq 0}$ is also an $\mathcal{S}'(\mathcal{X})$-valued Gaussian process.
Set $K(s,\phi;t,\psi) := \mathbb{E}[\langle Y_s, \phi\rangle\langle Y_t, \psi\rangle]$. To the end, one needs eventually to show that

$$\frac{\partial}{\partial t}K(t,\phi;t,\psi) - K(t,A_t\phi;t,\psi) - K(t,\phi;t,A_t\psi) = \langle Q_t\phi, \psi\rangle. \tag{5.3.7}$$

By using (5.2.14), we have

$$\begin{aligned} \frac{\partial}{\partial t}K(t,\phi;t,\psi) &= \frac{\partial}{\partial t}\mathbb{E}[\langle Y_t,\phi\rangle\langle Y_t,\psi\rangle] \\ &= \mathbb{E}\Big\{\langle Y_t,\psi\rangle\frac{\partial}{\partial t}\langle Y_t,\phi\rangle + \langle Y_t,\phi\rangle\frac{\partial}{\partial t}\langle Y_t,\psi\rangle\Big\} \\ &= \mathbb{E}\Big\{\langle Y_t,\psi\rangle\big(\langle Y_t,A_t\phi\rangle + \dot{M}_t(\phi)\big) + \langle Y_t,\phi\rangle\big(\langle Y_t,A_t\psi\rangle + \dot{M}_t(\psi)\big)\Big\} \\ &= \mathbb{E}[\langle Y_t,\psi\rangle\langle Y_t,A_t\phi\rangle] + \mathbb{E}[\langle Y_t,\phi\rangle\langle Y_t,A_t\psi\rangle] \\ &\qquad + \mathbb{E}[M_t(\psi)\dot{M}_t(\phi)] + \mathbb{E}[M_t(\phi)\dot{M}_t(\psi)] \\ &= K(t,A_t\phi;t,\psi) + K(t,\phi;t,A_t\psi) + \frac{\partial}{\partial t}\mathbb{E}[M_t(\psi)M_t(\phi)] \\ &= K(t,A_t\phi;t,\psi) + K(t,\phi;t,A_t\psi) + \langle Q_t\phi,\psi\rangle, \end{aligned} \tag{5.3.8}$$

where the last equality is due to (5.2.15) and (5.3.2).
On the other hand, it is not hard to check that $(Q_t)_{t\geq 0}$ satisfies the conditions required in Definition 5.3.2. Finally, by the results of [3, Theorem 2], there exists an $\mathcal{S}'(\mathcal{X})$-valued Wiener process $(W_t)_{t\geq 0}$ associated to the covariance functional $(Q_t)_{t\geq 0}$ such that $(Y_t)_{t\geq 0}$ satisfying the generalized Langevin equation (5.3.3) driven by the generalized Wiener process $(W_t)_{t\geq 0}$.

Uniqueness. Since all coefficients are bounded, the linear operator A_t on $\mathcal{S}(\mathcal{X})$ is uniformly bounded for $t \in [0, T]$ for any $T > 0$. Therefore, the equation (5.3.3) has an unique $\mathcal{S}'(\mathcal{X})$-valued solution given by the mild form:

$$Y_t = T^*_{0,t}\gamma + \int_0^t T^*_{r,t} dW_r, \tag{5.3.9}$$

where $\{T_{r,t} : 0 \leq r \leq t < +\infty\}$ is an unique reversed evolution system generated by $(A_t)_{t\geq 0}$ and $T^*_{r,t}$ is its adjoint operator of $T_{r,t}$. We refer the reader to [28, Theorem 2.1] for more details on evolution systems. □

5.3.2 One dimensional case: inhomogeneous OU process

In this subsection, we will study a simple case as an example of Theorem 5.3.3. Consider the case when there is no spatial dispersal and all the individuals stay at the same position, i.e., $\epsilon = 0$ in dispersal kernel (5.2.1).

Proposition 5.3.5. *Suppose that the same assumptions hold as in Theorem 5.2.3. In particular, assume $X^n_t = \xi^n_t \delta_x$ and $\epsilon = 0$ in (5.2.1). Then, $(\xi^n_t, \eta^n_t)_{t\geq 0}$ converge in law to $(\xi_t, \eta_t)_{t\geq 0}$ as $n \to \infty$ which satisfies the following equations:*

$$\begin{cases} d\xi_t = \big(b(x) - d(x) - \alpha(x,x)\xi_t\big)\xi_t dt \\ d\eta_t = \big(b(x) - d(x) - 2\alpha(x,x)\xi_t\big)\eta_t dt + \sqrt{\big(b(x) + d(x) + \alpha(x,x)\xi_t\big)\xi_t} dB_t, \end{cases} \tag{5.3.10}$$

where $\eta^n_t := \sqrt{n}(\xi^n_t - \xi_t)$.

Remark 5.3.6. *We can regard the above system as an inhomogeneous Ornstein-Uhlenbeck (OU) process living in a deterministic environment. We refer the reader to [20, Theorem 11.2.3] for a general form defined in an one-dimensional inhomogeneous Langevin equation.*

Proof. Due to $D(x, dz) = 1_{\{z=0\}}$, by taking $\phi = 1$ in (5.2.6), we can easily show that there exists a process $(\xi_t)_{t\geq 0}$ defined by $\xi_t := \langle X_t, 1\rangle$ solving the first equation in (5.3.10). Taking $\phi = 1$, from (5.3.4), we have

$$\langle Y_t, A_t 1\rangle = \big(b(x) - d(x) - 2\alpha(x,x)\xi_t\big)\langle Y_t, 1\rangle. \tag{5.3.11}$$

From (5.3.2) and (5.3.5), we have

$$\langle W_., 1\rangle_t = \int_0^t \big(b(x) + d(x) + \alpha(x,x)\xi_s\big)\xi_s ds. \tag{5.3.12}$$

Define

$$B_t = \int_0^t \Big[\big(b(x) + d(x) + \alpha(x,x)\xi_s\big)\xi_s\Big]^{-\frac{1}{2}} d\langle W_s, 1\rangle. \tag{5.3.13}$$

Then, its quadratic variation $\langle B\rangle_t = t$. Thus, $(B_t)_{t\geq 0}$ is a standard Brownian motion. Furthermore, we have

$$d\langle W_t, 1\rangle = \sqrt{\big(b(x) + d(x) + \alpha(x,x)\xi_t\big)\xi_t} \cdot dB_t. \tag{5.3.14}$$

Let $\eta_t := \langle Y_t, 1\rangle$, by plugging (5.3.11) and (5.3.14) back to (5.3.6) when $\phi = 1$, the second equation in (5.3.10) follows. ☐

In the next result, we give the stationary distribution of equations (5.3.10).

Proposition 5.3.7. *Suppose the process* $(\eta_t)_{t\geq 0}$ *is defined as in equations* (5.3.10). *Then, it has a stationary distribution which is Gaussian* $\mathcal{N}(0, \frac{b(x)}{\alpha(x,x)})$.

Remark 5.3.8. *The result is somewhat surprising. As long as* $d(x) < b(x)$, *in a long term, it always has the same fluctuation no matter which value death rate takes.*

Proof. Let

$$\theta_t := -\big(b(x) - d(x) - 2\alpha(x,x)\xi_t\big),$$
$$\sigma_t := \sqrt{\big(b(x) + d(x) + \alpha(x,x)\xi_t\big)\xi_t}.$$

From (5.3.10), it follows that

$$d\eta_t = -\theta_t\eta_t dt + \sigma_t dB_t. \tag{5.3.15}$$

The characteristic function of $(\eta_t)_{t\geq 0}$ has the form

$$\mathbb{E}_{\eta_0}\big[e^{iz\eta_t}\big] = \exp\Big\{ize^{-\int_0^t \theta_u du}\eta_0 - \frac{1}{2}z^2\int_0^t \sigma_u^2 e^{-2\int_u^t \theta_v dv} du\Big\}. \tag{5.3.16}$$

Since ξ_t in (5.3.10) has steady equilibrium $\frac{b(x)-d(x)}{\alpha(x,x)}$, it follows $\lim_{t\to\infty}\theta_t = b(x) - d(x) > 0$ and $\lim_{t\to\infty}\sigma_t^2 = 2b(x)\big(b(x) - d(x)\big)/\alpha(x,x)$.

Then,

$$\begin{aligned}\lim_{t\to\infty}\log\mathbb{E}_{\eta_0}\big[e^{iz\eta_t}\big] &= -\lim_{t\to\infty}\frac{1}{2}z^2\cdot\frac{\int_0^t \sigma_u^2 e^{2\int_0^u \theta_v dv} du}{e^{2\int_0^t \theta_u du}}\\ &= -\frac{1}{2}z^2\cdot\lim_{t\to\infty}\frac{\sigma_t^2}{2\theta_t}\\ &= -\frac{1}{2}z^2\frac{b(x)}{\alpha(x,x)}.\end{aligned} \tag{5.3.17}$$

Finally, we conclude that $(\eta_t)_{t\geq 0}$ has stationary distribution $\mathcal{N}(0, \frac{b(x)}{\alpha(x,x)})$. ☐

5.4 Interchangeability of the long term and the fluctuation limits

In this section, we will discuss the stationary behavior and fluctuation limits of a sequence of particular BPDL processes under some specific conditions. Furthermore, we can show that the stationary limits and fluctuation limits are interchangeable. This section is in parallel with previous section both as applications of Section 5.2.3 and Section 5.3.1.

Denote the following assumptions by (B).

(B1) Trait-independent parameters: $b(x) \equiv b,\ d(x) \equiv 0,\ D(x,z) = D(z)$, $\alpha(x,y) = \alpha(x-y)$ and symmetric property: $\alpha(z) = \alpha(-z),\ D(z) = D(-z)$.

(B2) Suppose $\lambda(x) \overset{\text{def}}{=} \int_{\mathcal{X}} \alpha(x,y)dy$ exists and $\lambda(x) \equiv \lambda$ is trait-independent. Set $\hat{\alpha}(x,y) = \frac{\alpha(x,y)}{\lambda}$, then $\int_{\mathcal{X}} \hat{\alpha}(z)dz = 1$.

(B3) Detailed balanced condition: $\hat{\alpha}(z) = D(z)$ and $\hat{\alpha}(0) = 0$.

In this section, we will only consider a sequence of BPDL processes $\{(\nu_t^n)_{t\geq 0}; n \in \mathbb{N}\}$ defined by (5.2.2) which satisfy the assumptions in (A1) and (B). Let's rewrite its generator under above assumptions:

$$\begin{aligned} L_0^n\Phi(\nu) = &\int_{\mathcal{X}} \nu(dx) \int_{\mathbb{R}^d} \Big(\Phi(\nu+\delta_{x+z}) - \Phi(\nu)\Big) bD(dz) \\ &+ \int_{\mathcal{X}} \nu(dx)\Big(\Phi(\nu-\delta_x) - \Phi(\nu)\Big)\frac{\lambda}{n}\int_{\mathcal{X}} \hat{\alpha}(x,y)\nu(dy). \end{aligned} \tag{5.4.1}$$

Lemma 5.4.1 (Long time behavior of ν_t^n)**.** *For fixed $n \in \mathbb{N}$, let ν_∞^n be a Poisson random measure with intensity $\frac{nb}{\lambda}dx$. Then, $(\nu_t^n)_{t\geq 0}$ starting from ν_∞^n is a stationary BPDL process, i.e. $\mathbb{E}[L_0^n\Phi(\nu_\infty^n)] = 0$, for $\Phi \in B(\mathcal{M}(\mathcal{X}))$.*

The proof of this Lemma can be deduced directly from [22, Proposition 7.9]. In the following Proposition, we would consider the fluctuation limits based on long term behavior limits of ν_t^n.

Proposition 5.4.2 (Fluctuation limit of ν_∞^n)**.** *Let $\Lambda^n(dx) \overset{def}{=} \frac{nb}{\lambda}dx$ be a sequence of biased Lebesgue measures on $\mathcal{X}$. Then, $\frac{\nu_\infty^n - \Lambda^n}{\sqrt{n}}$ converges in law as $n \to \infty$ to a mean zero isonormal Gaussian random measure W_∞ with variance intensity $\frac{b}{\lambda}dx$.*

The above Proposition can be easily proved by martingale central limit theorem. In some sense, the fluctuation limit here extend the convergence result in Theorem 5.2.3 from $[0,\infty)$ to $[0,\infty]$.

In the next Proposition, we would firstly apply the fluctuation theorem 5.2.3 and then take the long time limit. By taking $\phi = 1_A$ for any Borel set $A \subseteq \mathcal{X}$ and doing some easy calculations, we write down the differential form of equations (5.2.6) and (5.3.6) as following:

$$\frac{dX_t(A)}{dt} = b\int_A dx \int_{\mathcal{X}} X_t(dy)D(x-y) - \lambda \int_A X_t(dx) \int_{\mathcal{X}} X_t(dy)\hat{\alpha}(x-y), \tag{5.4.2}$$

and

$$\begin{aligned} dY_t(A) = b\int_A dx\int_{\mathcal{X}} Y_t(dy)D(x-y)dt - \lambda\int_A X_t(dx)\int_{\mathcal{X}} Y_t(dy)\hat{\alpha}(x-y)dt \\ - \lambda\int_A Y_t(dx)\int_{\mathcal{X}} X_t(dy)\hat{\alpha}(x-y)dt + dW_t(A), \end{aligned} \tag{5.4.3}$$

where

$$\frac{d}{dt}\langle W_\cdot(A)\rangle_t = b\int_A dx\int_{\mathcal{X}} X_t(dy)D(x-y) + \lambda\int_A X_t(dx)\int_{\mathcal{X}} X_t(dy)\hat{\alpha}(x-y). \tag{5.4.4}$$

Proposition 5.4.3. *Assume condition (B) and the same conditions as in Theorem 5.2.3 hold. Then, $(Y_t)_{t\geq 0}$ converges in law as $t\to\infty$ to a Gaussian random measure W_∞ with mean zero and variance intensity $\frac{b}{\lambda}dx$.*

Proof. For any fixed Borel set A in $\mathcal{X}$, due to the assumption $\hat{\alpha}(x-y) = D(x-y)$, we rewrite (5.4.2) as:

$$\frac{dX_t(A)}{dt} = \int_A \big(bdx - \lambda X_t(dx)\big)\int_{\mathcal{X}} X_t(dy)\hat{\alpha}(x-y). \tag{5.4.5}$$

Let Λ be the Lebesgue measure on $\mathcal{X}$. Then $X_t(A)$ converges to $\frac{b}{\lambda}\Lambda(A)$ as $t\to\infty$. Since A is arbitrary, X_t converges to $\frac{b}{\lambda}\Lambda$ as $t\to\infty$. In the following, we will approximate deterministic measure X_t by $\frac{b}{\lambda}\Lambda$ in (5.4.3) and (5.4.4) when considering the stationary behavior of the process $(Y_t)_{t\geq 0}$.
Therefore, the first two terms on RHS of (5.4.3) can cancel with each other when we take a long time limit. Meanwhile, the third term can be approximated by

$$\lambda\int_A Y_t(dx)\int_{\mathcal{X}} \frac{b}{\lambda}dy\hat{\alpha}(x-y) = bY_t(A). \tag{5.4.6}$$

And the quadratic variation (5.4.4) can be approximated by

$$\begin{aligned} b\int_A dx\int_{\mathcal{X}} \frac{b}{\lambda}dyD(x-y) + \lambda\int_A \frac{b}{\lambda}dx\int_{\mathcal{X}} \frac{b}{\lambda}dy\hat{\alpha}(x-y) \\ = 2\frac{b^2}{\lambda}\Lambda(A). \end{aligned} \tag{5.4.7}$$

Concerned the stationary distribution of $Y_t(A)$ given by (5.4.3) and (5.4.4), due to the above approximations it is equivalent to consider the following real-valued Ornstein-Uhlenbeck process:

$$d\tilde{Y}_t(A) = -b\cdot\tilde{Y}_t(A)dt + \sqrt{2\frac{b^2}{\lambda}\Lambda(A)}dW_t. \tag{5.4.8}$$

Obviously, the stationary distribution of $\big(\tilde{Y}_t(A)\big)_{t\geq 0}$ is Gaussian $\mathcal{N}(0, \frac{b}{\lambda}\Lambda(A))$.
Hence, $(Y_t)_{t\geq 0}$ converges in law as $t\to\infty$ to an isonormal Gaussian random measure W_∞ on $\mathcal{X}$ with mean zero and variance intensity $\frac{b}{\lambda}dx$. □

Remark 5.4.4. *Comparing with Proposition 5.3.7 without a zero death rate assumption, we may hope that a similar result still holds here but without so restrictive assumption (B1). In other words, the natural death term may not take essential role concerning the long term behavior.*

In all, the results in this section suggest us that under a weak interspecific competition, not only the macroscopic limit has equilibrium measure but also its fluctuation limit has a stationary distribution. Moreover, both of the above equilibria are independent for disjoint spatial area.

5.5 Outline of proofs

5.5.1 Moment estimates and tightness

The tightness criterion is established for semimartingales based on the moment estimates (see [18]). Our first two lemmas give the uniform second order moment estimates for a sequence of processes over finite time intervals.

Lemma 5.5.1. *Suppose that a sequence of random variables* (Y_0^n) *in* $\mathcal{M}_F(\mathcal{X})$ *satisfies the same condition as in Theorem 5.2.3. Then,* $X_0^n \overset{in\ law}{\longrightarrow} X_0$ *as* $n \to \infty$ *and*

$$\sup_{n\geq 1} \mathbb{E}\langle X_0^n, 1\rangle^2 < \infty. \tag{5.5.1}$$

Hence, Theorem 5.2.1 holds.

In particular, for any $T < \infty$*, there exists a constant* $C_T^{(1)} > 0$ *such that*

$$\sup_{n\geq 1} \mathbb{E}[\sup_{0\leq t\leq T} \langle X_t^n, 1\rangle^2] \leq C_T^{(1)}. \tag{5.5.2}$$

Proof. In fact, the convergence from (X_0^n) to X_0 in law can be implied by the convergence from (Y_0^n) to γ.

On the other hand, because of the definition of (Y_0^n) as in (5.2.7),

$$\sup_{n\geq 1} \mathbb{E}\langle X_0^n, 1\rangle^2 \leq 2\langle X_0, 1\rangle^2 + 2\sup_{n\geq 1} \frac{1}{n}\mathbb{E}\langle Y_0^n, 1\rangle^2 < \infty.$$

Now concern the proof of the moment estimate (5.5.2).

For fixed $n \in \mathbb{N}$, the rescaled total population $\{\langle X_t^n, 1\rangle; t \geq 0\}$ is a $\mathbb{N}/n$-valued process. It can be bounded by a $\mathbb{N}/n$-valued pure birth process with birth rate $n\bar{b}$. In other words, take $\Phi(\mu) = \langle \mu, 1\rangle^2$ in (5.2.4) and neglect the death terms, one obtains

$$\begin{aligned}\langle X_t^n, 1\rangle^2 =& \langle X_0^n, 1\rangle^2 + \int_0^t \Big(\big(\langle X_{s-}^n, 1\rangle + \frac{1}{n}\big)^2 - \langle X_{s-}^n, 1\rangle^2 \Big) \bar{b} n \langle X_{s-}^n, 1\rangle N(ds) \\ =& \langle X_0^n, 1\rangle^2 + 2\int_0^t ds \Big(\langle X_{s-}, 1\rangle / n + 2\langle X_{s-}^n, 1\rangle^2 \Big) \bar{b} N(ds)\end{aligned} \tag{5.5.3}$$

where $N(ds)$ is a standard Poisson point process, and the term $\sup\limits_{n\geq 1}\mathbb{E}[\sup\limits_{0\leq s\leq T}\langle X_s^n,1\rangle]$ can also be proved to be bounded by some constant $C_T^{(0)}$ in a similar argument.
By taking expectations of supremums over time interval $[0,T]$ on both sides , we get

$$\begin{aligned}\mathbb{E}[\sup_{0\leq t\leq T}\langle X_t^n,1\rangle^2] &\leq \sup_{n\geq 1}\mathbb{E}\langle X_0^n,1\rangle^2 + C_T^{(0)} + 2\bar{b}\mathbb{E}\Big[\sup_{0\leq t\leq T}\int_0^t \langle X_s^n,1\rangle^2 ds\Big] \\ &\leq \sup_{n\geq 1}\mathbb{E}\langle X_0^n,1\rangle^2 + C_T^{(0)} + 2\bar{b}\int_0^T \mathbb{E}\big[\sup_{0\leq u\leq s}\langle X_u^n,1\rangle^2\big]ds\end{aligned} \tag{5.5.4}$$

where $C_T^{(0)}$ changes from line to line but independent of n.

According to Gronwall's inequality, one eventually obtains

$$\mathbb{E}[\sup_{0\leq t\leq T}\langle X_t^n,1\rangle^2] \leq \Big(\sup_{n\geq 1}\mathbb{E}\langle X_0^n,1\rangle^2 + C_T^{(0)}\Big)e^{2\bar{b}T} \overset{\text{def}}{=} C_T^{(1)}.$$

Finally, the independence of $C_T^{(1)}$ w.r.t. n implies (5.5.2). □

Lemma 5.5.2. *Suppose that a sequence of random variables* $Y_0^n \in \mathcal{M}_F(\mathcal{X})$ *satisfies* (5.2.13)*. Then, for any* $T<\infty$*, there exists a constant* $C_T^{(2)}>0$ *such that*

$$\sup_{n\geq 1}\sup_{\|\phi\|_\infty\leq 1}\mathbb{E}[\sup_{0\leq t\leq T}\langle Y_t^n,\phi\rangle^2] \leq C_T^{(2)}. \tag{5.5.5}$$

Proof. From Proposition 5.2.2, by Hölder inequality, one obtains

$$\begin{aligned}\langle Y_t^n,\phi\rangle^2 \leq 2\Bigg\{&\langle Y_0^n,\phi\rangle^2 + t\int_0^t\Big(\int_{\mathcal{X}} Y_s^n(dx)b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)\Big)^2 ds \\ &+ t\int_0^t\Big(\int_{\mathcal{X}} Y_s^n(dx)d(x)\phi(x)\Big)^2 ds \\ &+ nt\int_0^t\Big(\int_{\mathcal{X}} X_s^n(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s^n(dy) \\ &\qquad -\int_{\mathcal{X}} X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s(dy)\Big)^2 ds \\ &+\Big[M_t^n(\phi)\Big]^2\Bigg\}\end{aligned} \tag{5.5.6}$$

$$\begin{aligned}
&\le 2\Big\{\langle Y_0^n,\phi\rangle^2 + t\int_0^t \Big(\int_{\mathcal{X}} Y_s^n(dx)b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)\Big)^2 ds \\
&\quad + t\int_0^t \Big(\int_{\mathcal{X}} Y_s^n(dx)d(x)\phi(x)\Big)^2 ds \\
&\quad + 2t\int_0^t \Big(\int_{\mathcal{X}} Y_s^n(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\Big)^2 ds \\
&\qquad\quad + 2t\int_0^t \Big(\int_{\mathcal{X}} X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)Y_s^n(dy)\Big)^2 ds \\
&\quad + \Big[M_t^n(\phi)\Big]^2\Big\}.
\end{aligned}\tag{5.5.7}$$

For fixed $T<\infty$, firstly take the supremum over time interval $[0,T]$, then take expectation on both sides. It follows that, for any $\phi\in B(\mathcal{X})$ satisfying $\|\phi\|_\infty\le 1$,

$$\begin{aligned}
&\mathbb{E}[\sup_{0\le t\le T}\langle Y_t^n,\phi\rangle^2] \\
&\le 2\mathbb{E}\langle Y_0^n,\phi\rangle^2 + 2T\bar{b}^2\int_0^T \mathbb{E}\sup_{0\le u\le s}\Big(\int_{\mathcal{X}} Y_u^n(dx)\frac{b(x)}{\bar{b}}\int_{\mathcal{X}}\phi(x+z)D(x,dz)\Big)^2 ds \\
&\quad + 2T\bar{d}^2\int_0^T \mathbb{E}\sup_{0\le u\le s}\Big(\int_{\mathcal{X}} Y_u^n(dx)\frac{d(x)}{\bar{d}}\phi(x)\Big)^2 ds \\
&\quad + 4T\bar{\alpha}^2\int_0^T \mathbb{E}\sup_{0\le u\le s}\Big(\int_{\mathcal{X}} Y_u^n(dx)\phi(x)\int_{\mathcal{X}}\frac{\alpha(x,y)}{\bar{\alpha}}X_u^n(dy)\Big)^2 ds \\
&\quad + 4T\bar{\alpha}^2\int_0^T \mathbb{E}\sup_{0\le u\le s}\Big(\int_{\mathcal{X}} X_u(dx)\phi(x)\int_{\mathcal{X}}\frac{\alpha(x,y)}{\bar{\alpha}}Y_u^n(dy)\Big)^2 ds \\
&\quad + 2\mathbb{E}\Big\{\sup_{0\le t\le T}[M_t^n(\phi)]^2\Big\} \\
&\overset{\text{def}}{=} 2\mathbb{E}\langle Y_0^n,\phi\rangle^2 + \mathbf{I}+\mathbf{II}+\mathbf{III}+\mathbf{IV}+\mathbf{V}.
\end{aligned}\tag{5.5.8}$$

To the end, we give estimate of every term in above equation separately.
As for term $\mathbf{V}$, by Doob's maximal inequality and (5.2.9), we have

$$\begin{aligned}
\mathbf{V} &\le 2\times 4\mathbb{E}\Big[M_T^n(\phi)^2\Big] \\
&\le 8\mathbb{E}[\langle M_\cdot^n(1)\rangle_T] \\
&\le 8(\bar{b}+\bar{d})\mathbb{E}\Big\{\int_0^T \sup_{0\le u\le s}\langle X_u^n,1\rangle ds\Big\} + 8\bar{\alpha}\mathbb{E}\int_0^T \sup_{0\le u\le s}\langle X_u^n,1\rangle^2 ds \\
&\le 8(\bar{b}+\bar{d}+\bar{\alpha})T\cdot(C_T^{(0)}+C_T^{(1)}),
\end{aligned}\tag{5.5.9}$$

where the last inequality is due to (5.5.2).
Since $\|\frac{b(x)}{\bar{b}}\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)\|_\infty\le 1$, $\|\frac{d(x)}{\bar{d}}\phi(x)\|_\infty\le 1$, one obtains

$$\mathbf{I}+\mathbf{II}\le 2T^2\big(\bar{b}^2+\bar{d}^2\big)\int_0^T \sup_{\|\phi\|_\infty\le 1}\mathbb{E}[\sup_{0\le u\le s}\langle Y_u^n,\phi\rangle^2]ds. \tag{5.5.10}$$

.
Similarly, **III** and **IV** can be bounded by $4T^2\bar{\alpha}^2C\int_0^T \sup\limits_{\|\phi\|_\infty\le 1}\mathbb{E}[\sup\limits_{0\le u\le s}\langle Y_u^n,\phi\rangle^2]ds$ with some constant C determined by bounded moment estimates in Lemma 5.5.1.
Let $G^n(T) := \sup\limits_{\|\phi\|_\infty\le 1}\mathbb{E}[\sup\limits_{0\le t\le T}\langle Y_t^n,\phi\rangle^2]$, by combining the above estimates and (5.5.8), one obtains

$$\begin{aligned} G^n(T) \le & 2\sup_{n\ge 1}\sup_{\|\phi\|_\infty\le 1}\mathbb{E}\langle Y_0^n,\phi\rangle^2 + 8(\bar{b}+\bar{d}+\bar{\alpha})T\cdot(C_T^{(0)}+C_T^{(1)}) \\ & + \Big(2T^2(\bar{b}^2+\bar{d}^2)+4T^2\bar{\alpha}^2C\Big)\int_0^T G^n(s)ds \end{aligned} \tag{5.5.11}$$

By Gronwall's lemma, we have

$$\begin{aligned} G^n(T) \le & \Big(2\sup_{n\ge 1}\sup_{\|\phi\|_\infty\le 1}\mathbb{E}\langle Y_0^n,\phi\rangle^2 + 8(\bar{b}+\bar{d}+\bar{\alpha})T\cdot(C_T^{(0)}+C_T^{(1)})\Big) \\ & \cdot\exp\Big\{\Big(2T^2(\bar{b}^2+\bar{d}^2)+4T^2\bar{\alpha}^2C\Big)T\Big\} \\ & \overset{\text{def}}{=} C_T^{(2)}. \end{aligned} \tag{5.5.12}$$

Since $C_T^{(2)}$ is a n-independent constant, the lemma follows by taking supremum over $n\in\mathbb{N}$ on both sides of the last inequality. ☐

Proposition 5.5.3. *Consider a sequence of processes $(Y_t^n)_{t\ge 0}$ in $D([0,\infty),\mathcal{M}_F(\mathcal{X}))$ and Y_0^n satisfying (5.2.13). Then, for any $\phi\in\mathcal{S}(\mathcal{X})$, the sequence of laws of the processes $\{\langle Y_\cdot^n,\phi\rangle; n\ge 1\}$ is tight in $D([0,\infty),\mathbb{R})$.*

Proof. Since $\{\langle Y_\cdot^n,\phi\rangle; n\ge 1\}$ is a sequence of semimartingale, we verify the tightness criteria given by Aldous [1] and Rebolledo (see, e.g., Etheridge[18, Theorem 1.17]).
For any fixed $t>0$, $\{\langle Y_t^n,\phi\rangle; n\ge 1\}$ is tight due to Lemma 5.5.2.
To the end, we will prove the tightness criterion of finite variation part (say A_t^n) and quadratic variation of martingale part $M_t^n(\phi)$ of $\{\langle Y_\cdot^n,\phi\rangle; n\ge 1\}$, respectively.
For any $\varepsilon>0$ and $T>0$, given a sequence of stopping time τ_n bounded by T. W.O.L.G.,

assume $\|\phi\|_\infty \leq 1$. As for the finite variation part A_t^n of (5.2.8), we have

$$\begin{aligned}
&\sup_{n\geq 1}\sup_{\theta\in[0,\delta]}\mathbb{P}\Big[\Big|A^{(n)}_{\tau_n+\theta}-A^{(n)}_{\tau_n}\Big|>\varepsilon\Big]\\
&\qquad\leq \frac{1}{\varepsilon^2}\sup_{n\geq 1}\sup_{\theta\in[0,\delta]}\mathbb{E}\Big(A^{(n)}_{\tau_n+\theta}-A^{(n)}_{\tau_n}\Big)^2\\
&\qquad\overset{\text{Hölder}}{\leq}\frac{\delta}{\varepsilon^2}\sup_{n\geq 1}\sup_{\theta\in[0,\delta]}\int_{\tau_n}^{\tau_n+\theta}\mathbb{E}\Big\{\int_{\mathcal{X}}Y_s^n(dx)b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)\\
&\qquad\quad-\int_{\mathcal{X}}Y_s^n(dx)d(x)\phi(x)\\
&\qquad\quad-\int_{\mathcal{X}}Y_s^n(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\\
&\qquad\quad-\int_{\mathcal{X}}X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)Y_s^n(dy)\Big\}^2ds\\
&\qquad\leq\frac{2\delta\bar{b}^2}{\varepsilon^2}\cdot\sup_{n\geq 1}\int_0^T\mathbb{E}\sup_{0\leq u\leq T}\langle Y_u^n,,\frac{b(\cdot)}{\bar{b}}\int_{\mathcal{X}}\phi(\cdot+z)D(\cdot,dz)\rangle^2ds\\
&\qquad\quad+\frac{2\delta\bar{d}^2}{\varepsilon^2}\cdot\sup_{n\geq 1}\int_0^T\mathbb{E}\sup_{0\leq u\leq T}\langle Y_u^n,\frac{d(\cdot)}{\bar{d}}\phi(\cdot)\rangle^2ds\\
&\qquad\quad+\frac{2\delta\bar{\alpha}^2C}{\varepsilon^2}\cdot\sup_{n\geq 1}\int_0^T\mathbb{E}\sup_{0\leq u\leq T}\langle Y_u^n,\hat{\phi}\rangle^2ds\\
&\qquad\leq\delta TC_T^{(2)}C,
\end{aligned}\tag{5.5.13}$$

where C changes from line to line and $\|\hat{\phi}\|_\infty\leq 1$. On the other hand, from (5.2.9), we have

$$\begin{aligned}
&\sup_{n\geq 1}\sup_{\theta\in[0,\delta]}\mathbb{P}\Big[\Big|\langle M^n_\cdot(\phi)\rangle_{\tau_n+\theta}-\langle M^n_\cdot(\phi)\rangle_{\tau_n}\Big|>\varepsilon\Big]\\
&\qquad\leq\frac{\delta(\bar{b}+\bar{d})}{\varepsilon}\cdot\sup_{n\geq 1}\mathbb{E}\sup_{0\leq u\leq T}\langle X_u^n,1\rangle ds\\
&\qquad\quad+\frac{\delta\bar{\alpha}}{\varepsilon}\cdot\sup_{n\geq 1}\mathbb{E}\sup_{0\leq u\leq T}\langle X_s^n,1\rangle^2ds\\
&\leq\delta(C_T^{(0)}+C_T^{(1)})C.
\end{aligned}\tag{5.5.14}$$

According to moment estimates results in Lemma 5.5.1 and Lemma 5.5.2, both inequalities (5.5.13) and (5.5.14) can be less than ε if we take δ (which only depends on $T,\varepsilon,\|\phi\|_\infty$) small enough, i.e.,

$$\sup_{n\geq 1}\sup_{\theta\in[0,\delta]}\mathbb{P}\Big[\Big|A^{(n)}_{\tau_n+\theta}-A^{(n)}_{\tau_n}\Big|>\varepsilon\Big]<\varepsilon,$$

$$\sup_{n\geq 1}\sup_{\theta\in[0,\delta]}\mathbb{P}\Big[\Big|\langle M^n_\cdot(\phi)\rangle_{\tau_n+\theta}-\langle M^n_\cdot(\phi)\rangle_{\tau_n}\Big|>\varepsilon\Big]<\varepsilon,$$

which fulfils the Aldous-Rebolledo tightness condition. □

5.5.2 Convergence in the f.d.d. sense

In this section, we prove a weak limit of $\{(Y_t^n)_{t\geq 0}; n\geq 1\}$ in the sense of f.d.d. convergence is a solution of some martingale problem.

Proposition 5.5.4. *Under the conditions given in Theorem* 5.2.3*, the finite dimensional distributions of* $(Y_t^n)_{t\geq 0}$ *converge as* $n\to\infty$ *to those of an* $\mathcal{S}'(\mathcal{X})$*-valued Markov process* $(Y_t)_{t\geq 0}$ *satisfying that for* $\phi\in\mathcal{S}(\mathcal{X})$*, the process*

$$\begin{aligned}M_t(\phi) :=&\langle Y_t,\phi\rangle - \langle \gamma,\phi\rangle - \int_0^t \Big\langle Y_s,\, b(\cdot)\int_{\mathbb{R}^d}\phi(\cdot+z)D(\cdot,dz)\Big\rangle ds\\ &+\int_0^t \Big\langle Y_s,\, d(\cdot)\phi(\cdot)\Big\rangle ds\\ &+\int_0^t \Big\langle Y_s,\, \int_{\mathcal{X}}\alpha(x,\cdot)\phi(x)X_s(dx)\Big\rangle ds\\ &+\int_0^t \Big\langle Y_s,\, \phi(\cdot)\int_{\mathcal{X}}\alpha(\cdot,y)X_s(dy)\Big\rangle ds\\ &.\end{aligned} \tag{5.5.15}$$

is a continuous martingale with quadratic variation

$$\begin{aligned}\langle M_\cdot(\phi)\rangle_t =&\int_o^t ds\int_{\mathcal{X}} X_s(dx)b(x)\int_{\mathbb{R}^d}\phi^2(x+z)D(x,dz)\\ &+\int_0^t ds\int_{\mathcal{X}}X_s(dx)\phi^2(x)\Big(d(x)+\int_{\mathcal{X}}\alpha(x,y)X_s(dy)\Big).\end{aligned} \tag{5.5.16}$$

Proof. By Proposition 5.5.3, we already proved $\{\langle Y_\cdot^n,\phi\rangle; n\geq 1\}$ is tight in $\mathbb{D}([0,\infty),\mathbb{R})$ for any $\phi\in\mathcal{S}(\mathcal{X})$. Following Mitoma [40] (see e.g., Ethier and Kurtz [20, Theorem 3.9.1]), we conclude that the sequence $\{(Y_t^n)_{t\geq 0}; n\geq 1\}$ is tight in $\mathbb{D}([0,\infty),\mathcal{S}'(\mathcal{X}))$. Hence, we can assume there exists a weak limit $(Y_t)_{t\geq 0}$ of a subsequence of $\{(Y_t^n)_{t\geq 0}; n\geq 1\}$.

Firstly, we check that $(Y_t)_{t\geq 0}$ is a.s. continuous. By the construction of (Y_t^n), we have

$$\begin{aligned}\sup_{t\in[0,T]}\sup_{\|f\|\leq 1}|\langle Y_t^n,f\rangle - \langle Y_{t-}^n,f\rangle| &\leq \sup_{t\in[0,T]}\sup_{\|f\|\leq 1}\sqrt{n}\big\{|\langle X_t^n,f\rangle - \langle X_{t-}^n,f\rangle| + |\langle X_t - X_{t-},f\rangle|\big\}\\ &\leq \sqrt{n}\frac{1}{n}+0\\ &=\frac{1}{\sqrt{n}}.\end{aligned} \tag{5.5.17}$$

By letting $n\to\infty$, it implies the continuity of $(Y_t)_{t\geq 0}$, i.e. $(Y_t)_{t\geq 0}\in\mathbb{C}([0,+\infty),\mathcal{S}'(X))$.

To prove $(M_t(\phi))_{t\geq 0}$ is a martingale, it suffices to prove

$$\mathbb{E}[M_t(\phi)] = 0. \tag{5.5.18}$$

Let

$$\begin{aligned}\widetilde{M}^n_t(\phi) :=&\langle Y^n_t, \phi\rangle - \langle Y^n_0, \phi\rangle - \int_0^t ds \int_{\mathcal{X}} Y^n_s(dx) b(x) \int_{\mathbb{R}^d} \phi(x+z) D(x, dz)\\ &+ \int_0^t ds \int_{\mathcal{X}} d(x)\phi(x) Y^n_s(dx)\\ &+ \int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) Y^n_s(dy)\\ &+ \int_0^t ds \int_{\mathcal{X}} Y^n_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) X_s(dy).\end{aligned} \tag{5.5.19}$$

Then, for fixed $t > 0$ and any $n \in \mathbb{N}$, we have

$$|\mathbb{E}[M_t(\phi)]| \le |\mathbb{E}[M^n_t(\phi) - \widetilde{M}^n_t(\phi)]| + |\mathbb{E}[\widetilde{M}^n_t(\phi) - M_t(\phi)]| + |\mathbb{E}[M^n_t(\phi)]|. \tag{5.5.20}$$

According to Proposition 5.2.2, we have $\mathbb{E}[M^n_t(\phi)] = 0$.
Since $\{(Y^n_t)_{t\ge 0}; n \ge 1\}$ converges in law to $(Y_t)_{t\ge 0}$ as $n \to \infty$ and $\big(\widetilde{M}^n_t(\phi) - M_t(\phi)\big)$ is homogeneous w.r.t. $\big(Y^n_t - Y_t\big)$, we get

$$\lim_{n\to\infty} |\mathbb{E}[\widetilde{M}^n_t(\phi) - M_t(\phi)]| = 0. \tag{5.5.21}$$

As for the first term on RHS of (5.5.20),

$$\begin{aligned}&|\mathbb{E}[M^n_t(\phi) - \widetilde{M}^n_t(\phi)]|\\ &= \Bigg|\mathbb{E}\Bigg\{\sqrt{n}\int_0^t ds \int_{\mathcal{X}} \Big(\frac{Y^n_s(dx)}{\sqrt{n}} + X_s(dx)\Big)\phi(x) \int_{\mathcal{X}} \alpha(x,y)\Big(\frac{Y^n_s(dy)}{\sqrt{n}} + X_s(dy)\Big)\\ &\qquad - \sqrt{n}\int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) X_s(dy)\\ &\qquad - \int_0^t ds \int_{\mathcal{X}} X_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) Y^n_s(dy)\\ &\qquad - \int_0^t ds \int_{\mathcal{X}} Y^n_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) X_s(dy)\Bigg\}\Bigg|\\ &\le \frac{1}{\sqrt{n}}\left|\mathbb{E}\int_0^t ds \int_{\mathcal{X}} Y^n_s(dx)\phi(x) \int_{\mathcal{X}} \alpha(x,y) Y^n_s(dy)\right|\\ &\le \frac{1}{\sqrt{n}} \bar{\alpha} T \|\phi\|_\infty C^{(2)}_t\\ &\xrightarrow{n\to\infty} 0,\end{aligned} \tag{5.5.22}$$

where $C^{(2)}_t$ is determined as in Lemma 5.5.2.
By combining the above estimates together, we conclude $|\mathbb{E}[M_t(\phi)]| = 0$.

In the remainder we will justify the quadratic variation of $M_t(\phi)$ has the form (5.5.16). By applying Itô's formula to $\langle Y_t, \phi\rangle^2$, according to the semimartingale decomposition (5.5.15) of $\langle Y_t, \phi\rangle$, we have

$$\begin{aligned}
\langle Y_t, \phi\rangle^2 &= \langle \gamma, \phi\rangle^2 + 2\int_0^t \langle Y_s, \phi\rangle d[\langle Y_s, \phi\rangle] + \langle M_\cdot(\phi)\rangle_t \\
&= \langle \gamma, \phi\rangle^2 + \langle M_\cdot(\phi)\rangle_t \\
&\quad + 2\int_0^t \langle Y_s, \phi\rangle ds\Big\{\int_{\mathcal{X}} Y_s(dx) b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) - \int_{\mathcal{X}} Y_s(dx)d(x)\phi(x) \\
&\quad - \int_{\mathcal{X}} Y_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s(dy) - \int_{\mathcal{X}} X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)Y_s(dy)\Big\} \\
&\quad + \text{martingale}.
\end{aligned} \tag{5.5.23}$$

On the other hand, according to the definition of (Y_t^n), we have

$$\begin{aligned}
\langle Y_t^n, \phi\rangle^2 &= \langle \sqrt{n}(X_t^n - X_t), \phi\rangle^2 \\
&= n\Big[\langle X_t^n, \phi\rangle^2 - 2\langle X_t^n, \phi\rangle\langle X_t, \phi\rangle + \langle X_t, \phi\rangle^2\Big].
\end{aligned} \tag{5.5.24}$$

To simplify the computations, let us introduce new notations:

$$\begin{aligned}
A(s) &:= \int_{\mathcal{X}} X_s(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) - \phi(x)\big(d(x) + \int_{\mathcal{X}}\alpha(x,y)X_s(dy)\big)\Big], \\
B^n(s) &:= \int_{\mathcal{X}} X_s^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) - \phi(x)\big(d(x) + \int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\big)\Big].
\end{aligned} \tag{5.5.25}$$

From (5.2.10), (5.2.11) and (5.2.6), respectively, it follows

$$\begin{aligned}
\langle X_t^n, \phi\rangle &= \langle X_0^n, \phi\rangle + \int_0^t B^n(s)ds + \text{martingale}, \\
\langle X_t^n, \phi\rangle^2 &= \langle X_0^n, \phi\rangle^2 + 2\int_0^t \langle X_s^n, \phi\rangle B^n(s)ds + \frac{1}{n}\int_0^t ds\int_{\mathcal{X}} X_s^n(dx) \\
&\qquad \Big[b(x)\int_{\mathbb{R}^d}\phi^2(x+z)D(x,dz) + \phi^2(x)\big(d(x) + \int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\big)\Big] + \text{martingale}, \\
\langle X_t, \phi\rangle &= \langle X_0, \phi\rangle + \int_0^t A(s)ds.
\end{aligned} \tag{5.5.26}$$

By substituting every term above into (5.5.24), we have

$$\begin{aligned}
\langle Y_t^n, \phi\rangle^2 = n\langle X_0^n, \phi\rangle^2 &+ \int_0^t ds \int_{\mathcal{X}} X_s^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi^2(x+z)D(x,dz) \\
&+ \phi^2(x)\big(d(x) + \int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\big)\Big] + 2n\int_0^t \langle X_s^n,\phi\rangle B^n(s)ds \\
&- 2n\Big[\langle X_0^n,\phi\rangle + \int_0^t B^n(s)ds\Big]\Big[\langle X_0,\phi\rangle + \int_0^t A(s)ds\Big] \\
&+ n\Big[\langle X_0,\phi\rangle + \int_0^t A(s)ds\Big]^2 + \text{martingale.}
\end{aligned} \tag{5.5.27}$$

Set

$$D^{t,n,1} := \int_0^t ds\int_{\mathcal{X}} X_s^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi^2(x+z)D(x,dz)+\phi^2(x)\big(d(x)+\int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\big)\Big]. \tag{5.5.28}$$

By combining all the quadratic term at time 0 in (5.5.27) together, it follows

$$\begin{aligned}
(5.5.27) =& n\langle X_0^n - X_0,\phi\rangle^2 + D^{t,n,1} \\
&+ 2n\int_0^t\langle X_s^n,\phi\rangle B^n(s)ds - 2n\langle X_t,\phi\rangle\int_0^t B^n(s)ds \\
&- 2n\langle X_0^n,\phi\rangle\int_0^t A(s)ds + 2n\langle X_0,\phi\rangle\int_0^t A(s)ds \\
&+ n\Big[\int_0^t A(s)ds\Big]^2 + \text{martingale}
\end{aligned} \tag{5.5.29}$$

$$\begin{aligned}
=&\langle Y_0^n,\phi\rangle^2 + D^{t,n,1} \\
&+ 2n\int_0^t\langle\frac{Y_s^n}{\sqrt{n}} + X_s,\phi\rangle B^n(s)ds - 2n\langle X_t,\phi\rangle\int_0^t B^n(s)ds \\
&- 2\sqrt{n}\langle Y_0^n,\phi\rangle\int_0^t A(s)ds + n\Big[\int_0^t A(s)ds\Big]^2 + \text{martingale}
\end{aligned} \tag{5.5.30}$$

$$\begin{aligned}
=&\langle Y_0^n,\phi\rangle^2 + D^{t,n,1} \\
&+ 2\sqrt{n}\int_0^t\langle Y_s^n,\phi\rangle B^n(s)ds \\
&+ 2n\int_0^t\langle X_s,\phi\rangle B^n(s)ds - 2n\langle X_t,\phi\rangle\int_0^t B^n(s)ds \\
&- 2\sqrt{n}\langle Y_0^n,\phi\rangle\int_0^t A(s)ds + n\Big[\int_0^t A(s)ds\Big]^2 + \text{martingale}
\end{aligned} \tag{5.5.31}$$

$$
\begin{aligned}
\overset{\text{Integration by parts}}{=} & \langle Y_0^n, \phi\rangle^2 + D^{t,n,1} \\
& + 2\sqrt{n}\int_0^t \langle Y_s^n, \phi\rangle B^n(s)ds \\
& - 2n\int_0^t dsA(s)\int_0^s B^n(r)dr - 2\sqrt{n}\langle Y_0^n, \phi\rangle \int_0^t A(s)ds \\
& + n\Big[\int_0^t A(s)ds\Big]^2 + \text{martingale}
\end{aligned} \tag{5.5.32}
$$

$$
\begin{aligned}
\overset{\text{Replace } B^n(s) \text{ by(5.5.25)}}{=} & \langle Y_0^n, \phi\rangle^2 + D^{t,n,1} \\
& + 2\int_0^t \langle Y_s^n, \phi\rangle ds\Big\{\int_{\mathcal{X}} Y_s^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) - \phi(x)\Big(d(x) \\
& + \int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\Big)\Big] - \int_{\mathcal{X}} X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)Y_s^n(dy)\Big\} \\
& + 2\sqrt{n}\int_0^t \langle Y_s^n, \phi\rangle A(s)ds \\
& - 2n\int_0^t dsA(s)\int_0^s B^n(r)dr - 2\sqrt{n}\langle Y_0^n, \phi\rangle \int_0^t A(s)ds \\
& + n\Big[\int_0^t A(s)ds\Big]^2 + \text{martingale.}
\end{aligned} \tag{5.5.33}
$$

Set

$$
\begin{aligned}
D^{t,n,2} := & 2\int_0^t \langle Y_s^n, \phi\rangle ds\Big\{\int_{\mathcal{X}} Y_s^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) - d(x)\phi(x) \\
& - \phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s^n(dy)\Big] - \int_{\mathcal{X}} X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)Y_s^n(dy)\Big\}.
\end{aligned} \tag{5.5.34}
$$

Replace $\langle Y_s^n, \phi\rangle$ and $B^n(r)$ by (5.2.8) and (5.5.25) respectively, one obtains

$$
\begin{aligned}
(5.5.33) = & \langle Y_0^n, \phi\rangle^2 + D^{t,n,1} + D^{t,n,2} \\
& + 2\sqrt{n}\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} Y_r^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) - d(x)\phi(x)\Big] \\
& - 2n\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} X_r^n(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_r^n(dy) \\
& + 2n\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} X_r(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_r(dy) \\
& - 2n\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} X_r^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz) \\
& \quad - \phi(x)\big(d(x) + \int_{\mathcal{X}}\alpha(x,y)X_r^n(dy)\big)\Big] + n\Big[\int_0^t A(s)ds\Big]^2 + \text{martingale}
\end{aligned} \tag{5.5.35}
$$

$$\begin{aligned}
=&\langle Y_0^n,\phi\rangle^2 + D^{t,n,1} + D^{t,n,2}\\
&+2\sqrt{n}\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} Y_r^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)-d(x)\phi(x)\Big]\\
&+2n\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} X_r(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_r(dy)\\
&-2n\int_0^t dsA(s)\int_0^s dr\int_{\mathcal{X}} X_r^n(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)-d(x)\phi(x)\Big]\\
&+n\Big[\int_0^t A(s)ds\Big]^2+\text{martingale}.
\end{aligned} \tag{5.5.36}$$

Combine X_r^n, X_r and Y_r^n, it thus follows

$$\begin{aligned}
\langle Y_t^n,\phi\rangle^2 =&\langle Y_0^n,\phi\rangle^2 + D^{t,n,1} + D^{t,n,2}\\
&-2n\int_0^t dsA(s)\int_0^s A(r)dr\\
&+n\Big[\int_0^t A(s)ds\Big]^2+\text{martingale}\\
\overset{\text{Integration by parts}}{=}&\langle Y_0^n,\phi\rangle^2 + D^{t,n,1} + D^{t,n,2}+\text{martingale}.
\end{aligned} \tag{5.5.37}$$

Obviously, both $D^{t,n,1}$ and $D^{t,n,2}$ converge as $n\to\infty$.

Finally, we get

$$\begin{aligned}
\langle Y_t,\phi\rangle^2 = \langle\gamma,\phi\rangle^2 &+ \int_o^t ds\int_{\mathcal{X}} X_s(dx)b(x)\int_{\mathbb{R}^d}\phi^2(x+z)D(x,dz)\\
&+\int_0^t ds\int_{\mathcal{X}} X_s(dx)\phi^2(x)\Big(d(x)+\int_{\mathcal{X}}\alpha(x,y)X_s(dy)\Big)\\
&+2\int_0^t\langle Y_s,\phi\rangle ds\Big\{\int_{\mathcal{X}} Y_s(dx)\Big[b(x)\int_{\mathbb{R}^d}\phi(x+z)D(x,dz)-d(x)\phi(x)\\
&-\phi(x)\int_{\mathcal{X}}\alpha(x,y)X_s(dy)\Big]-\int_{\mathcal{X}} X_s(dx)\phi(x)\int_{\mathcal{X}}\alpha(x,y)Y_s(dy)\Big\}\\
&+\text{martingale}.
\end{aligned} \tag{5.5.38}$$

By comparing the representations of (5.5.23) and (5.5.38), we conclude

$$\begin{aligned}
\langle M_\cdot(\phi)\rangle_t = &\int_o^t ds\int_{\mathcal{X}} X_s(dx)b(x)\int_{\mathbb{R}^d}\phi^2(x+z)D(x,dz)\\
&+\int_0^t ds\int_{\mathcal{X}} X_s(dx)\phi^2(x)\Big(d(x)+\int_{\mathcal{X}}\alpha(x,y)X_s(dy)\Big).
\end{aligned} \tag{5.5.39}$$

□

A Stability of a Lotka-Volterra system

Consider a Lotka-Volterra system $(n(x), n(y))$ satisfying the following equations.

$$\left\{ \begin{array}{l} \dot{n}_t(x) = (b(x) - d(x) - \alpha(x,x)n_t(x) - \alpha(x,y)n_t(y))\, n_t(x) \\ \dot{n}_t(y) = (b(y) - d(y) - \alpha(y,x)n_t(x) - \alpha(y,y)n_t(y))\, n_t(y). \end{array} \right. \tag{A.0.1}$$

Suppose that $n_0(x), n_0(y) > 0$ and $f(y,x) := b(y) - d(y) - \alpha(y,x)\bar{n}(x) > 0,\ \bar{n}(x) = \frac{b(x)-d(x)}{\alpha(x,x)}$, and its symmetric form $f(x,y) < 0$. Then we conclude that $(0, \bar{n}(y))$ is the only stable point.

In fact, there are four fixed points of above system, namely, $(0,0)$, $(\bar{n}(x), 0)$, $(0, \bar{n}(y))$, and $(n^*(x), n^*(y))$, where $(n^*(x), n^*(y))$ is such that

$$\left\{ \begin{array}{l} b(x) - d(x) - \alpha(x,x)n_t(x) - \alpha(x,y)n_t(y) = 0 \\ b(y) - d(y) - \alpha(y,x)n_t(x) - \alpha(y,y)n_t(y) = 0. \end{array} \right.$$

By simple calculation, we obtain that

$$\left\{ \begin{array}{l} n^*(x) = \frac{\alpha(y,y)f(x,y)}{\alpha(x,x)\alpha(y,y) - \alpha(x,y)\alpha(y,x)} \\ n^*(x) = \frac{\alpha(x,x)f(y,x)}{\alpha(x,x)\alpha(y,y) - \alpha(x,y)\alpha(y,x)}. \end{array} \right.$$

To make sense of the solution as a population density (which must be non-negative), one needs $f(x,y) \cdot f(y,x) > 0$. It contradicts the assumption $f(x,y) < 0, f(y,x) > 0$. We thus exclude the solution $(n^*(x), n^*(y))$.

The Jacobian matrix for the system (A.0.1) at point $(0,0)$ is

$$\begin{pmatrix} b(x) - d(x) & 0 \\ 0 & b(y) - d(y) \end{pmatrix}.$$

Obviously its eigenvalues are both positive. Thus $(0,0)$ is unstable.

The Jacobian matric at point $(\bar{n}(x), 0)$ is

$$\begin{aligned} &\begin{pmatrix} -(b(x) - d(x)) & -\alpha(x,y)\bar{n}(x) \\ 0 & b(y) - d(y) - \alpha(y,x)\bar{n}(x) \end{pmatrix} \\ &= \begin{pmatrix} -(b(x) - d(x)) & -\alpha(x,y)\bar{n}(x) \\ 0 & f(y,x) \end{pmatrix}. \end{aligned}$$

Since one of its eigenvalue $-(b(x) - d(x))$ is negative whereas the other one is $f(y,x) > 0$, the equilibrium $(\bar{n}(x), 0)$ is unstable.

The Jacobian matric of system (A.0.1) at point $(0, \bar{n}(y))$ is

$$\begin{pmatrix} b(x) - d(x) - \alpha(x,y)\bar{n}(y) & 0 \\ -\alpha(y,x)\bar{n}(y) & -(b(y) - d(y)) \end{pmatrix} = \begin{pmatrix} f(x,y) & 0 \\ -\alpha(y,x)\bar{n}(y) & -(b(y) - d(y)) \end{pmatrix},$$

whose eigenvalues are both negative because of the condition $f(x,y) < 0$. Thus $(0, \bar{n}(y))$ is the only stable equilibrium of the system (A.0.1).

B R-programming for TST

Figure 3.4 is generated by the following R-programming functions:

```
output<- TST-4(4*10^7, 10^3, 2*10^2, 2.5, c(3,6,8,10), c(0,0,0,0), 1, 0.8, 1.5)
output<- TST-5(4*10^7, 10^3, 4*10^2, 2.5, c(3,6,8,10,12), c(0,0,0,0,0), 1, 0.8,
               1.5).
```

Take the second line as an example, the output for the last stored vector output[1001] is as follows:

```
> output[1001,6]
[1] 11.5825
> output[1001,5]
[1] 0.2475
> output[1001,4]
[1] 7.5675
> output[1001,3]
[1] 0.1025
> output[1001,2]
[1] 3.0775
```

We enclose the code of the function TST-5 as follows.

```
sim.BDCM125type <- function(iter, plot.points, scale.K, init.density=2.5,
                   B=c(3,6,8,10,12), D=c(0,0,0,0,0), c=1, alpha=.5, beta=1.5)
{
  # The function sim.BDCM simulates a birth-death process with competition
  # and mutation between the two species.
  #
  # Input:
  # iter         - number of iterations
  # plot.points  - number of points stored in the vector N and M
  # scale.K      - scaling parameter
  # init.density - initial density
  # B            - vector of birth rates
  # D            - vector of death rates
  # c            - scalar of competition kernel
  # alpha        - exponent of mutation given in the form scale.Ksim^(-alpha)
  #
  # Output:
  # data.frame   - containing the history of N and M

  # Initializing needed arrays
  output.T <- vector(length=plot.points);

  output.N <- vector(length=plot.points);
  output.M <- vector(length=plot.points);
  output.Q <- vector(length=plot.points);
  output.R <- vector(length=plot.points);
```

```
output.S <- vector(length=plot.points);

# Set initial values
store.dist  <- floor(iter/plot.points);
store.break <- store.dist;

N <- scale.K * init.density;
M <- 0;
Q <- 0;
R <- 0;
S <- 0;

t <- 0;
j <- 2;

output.N[1] <- N;
output.M[1] <- M;
output.Q[1] <- Q;
output.R[1] <- R;
output.S[1] <- S;

output.T[1] <- t;

# Set migation probability of a birth
p <- scale.K^(-alpha);
# Set mutation probability of a birth
q <- scale.K^(-beta);
# Simulation loop
for (i in seq(1,iter))
{
  # Computation of the corresponding rates
  frakN <- (B[1] + D[1] + c * N/scale.K  + c * M/scale.K );
  frakM <- (B[2] + D[2] + c * M/scale.K  + c * N/scale.K + c * Q/scale.K);
  frakQ <- (B[3] + D[3] + c * Q/scale.K  + c * M/scale.K + c * R/scale.K );
  frakR <- (B[4] + D[4] + c * R/scale.K  + c * Q/scale.K + c * S/scale.K );
  frakS <- (B[5] + D[5] + c * S/scale.K  + c * R/scale.K );

  # Simulate the length of the timestep
  t <- t + rexp(1, rate= N * frakN + M * frakM +
                Q * frakQ + R * frakR + S * frakS );

  # Generate random events
  v1 <- rbinom(1, 1, N/(N + M + Q + R + S));
  v2 <- rbinom(1, 1, M/(N + M + Q + R + S));
  v3 <- rbinom(1, 1, Q/(N + M + Q + R + S));
  v4 <- rbinom(1, 1, R/(N + M + Q + R + S));
  v5 <- rbinom(1, 1, S/(N + M + Q + R + S));

  u1 <- rbinom(1, 1, B[1]/frakN);
  u2 <- rbinom(1, 1, B[2]/frakM);
  u3 <- rbinom(1, 1, B[3]/frakQ);
  u4 <- rbinom(1, 1, B[4]/frakR);
  u5 <- rbinom(1, 1, B[5]/frakS);

   w1 <- rbinom(1, 1, q);
   w2 <- rbinom(1, 1, q);
   w3 <- rbinom(1, 1, q);
   w4 <- rbinom(1, 1, q);

  # Update rules#

  if (v1==1&&u1==1&&w1==1)
  {
   ##mutation from type1 to type2##
   M <- M+1;
  }
```

```
    else if (v1==1&&u1==1&&w1==0)
    {
      ##birth from type1 non-mutation##
      m = rbinom(1, 1, p);
      if (m==1&&M>0)
      {
        ### migration birth ###
        M <- M+1;
      }
      else
      {
        ### copy birth ###
        N <- N+1;
      }
    }

    else if (v1==1&&u1==0)
    {
      ##death of type1##
      N <- N-1;
    }
    else if (v2==1&&u2==1&&w2==1)
    {
     ##mutation from type2 to type3##
     Q <- Q+1;
    }
    else if (v2==1&&u2==1&&w2==0)
     {
      ####birth from type2 non-mutation####
      m <- rbinom(1, 1, p);
      m1 <- rbinom(1, 1, 1/2);

      if (m==1&&Q==0)
      {
        ###migation birth to left-truncate right###
        N <- N+1;
      }
      else if (m==1&&Q>0&&m1==1)
      {
        ###migration birth to left###
        N <- N+1;
      }
      else if (m==1&&Q>0&&m1==0)
      {
      ###migration birth to right###
        Q <- Q+1;
      }
      else
      {
        ###copy birth###
        M <- M+1;
      }
     }

    else if (v2==1&&u2==0)
    {
      ##death of type2##
      M <- M-1;
    }
    else if (v3==1&&u3==1&&w3==1)
    {
     ##mutation from type3 to type4##
      R<- R+1;
    }
    else if (v3==1&&u3==1&&w3==0)
    {
```

```
      ##birth from type3 non-mutation##
      m <- rbinom(1, 1, p);
      m1 <- rbinom(1, 1, 1/2);

      if (m==1&&R==0)
      {
        ###migation birth to left-truncate right###
        M <- M+1;
      }
      else if (m==1&&R>0&&m1==1)
      {
        ###migration birth to left###
        M <- M+1;
      }
      else if (m==1&&R>0&&m1==0)
      {
        ###migration birth to right###
        R <- R+1;
      }
      else
      {
        ###copy birth###
        Q <- Q+1;
      }
     }
    else if (v3==1&&u3==0)
     {
      ##death of type3##
      Q <- Q-1;
     }
    else if (v4==1&&u4==1&&w4==1)
    {
     ##mutation from type4 to type5##
      S<- S+1;
    }
    else if (v4==1&&u4==1&&w4==0)
    {
      ##birth from type4 non-mutation##
      m <- rbinom(1, 1, p);
      m1 <- rbinom(1, 1, 1/2);

      if (m==1&&S==0)
      {
        ###migation birth to left-truncate right###
        Q <- Q+1;
      }
      else if (m==1&&S>0&&m1==1)
      {
        ###migration birth to left###
        Q <- Q+1;
      }
      else if (m==1&&S>0&&m1==0)
      {
        ###migration birth to right###
        S <- S+1;
      }
      else
      {
        ###copy birth###
        R <- R+1;
      }
     }
    else if (v4==1&&u4==0)
     {
      ##death of type4##
      R <- R-1;
```

```
     }
    else if (v5==1&&u5==1)
    {
      ##birth of type5##
      m <- rbinom(1, 1, p);
      if (m==1)
      {
        ###migration birth to left###
        R <- R+1;
      }
      else
      {
        ###copy birth###
        S <- S+1;
      }
     }
    else if (v5==1&&u5==0)
     {
      ##death of type5##
      S <- S-1;
     };

    # Store output
    if (i == store.break)
    {
      output.T[j] <- t;
      output.N[j] <- N;
      output.M[j] <- M;
      output.Q[j] <- Q;
      output.R[j] <- R;
      output.S[j] <- S;
      j <- j+1;

      store.break <- store.break + store.dist
    }
  }

  # Graphical output
  plot(output.T, output.N/scale.K, type="s", col="red", ylim=range(0,13))
  lines(output.T, output.M/scale.K, type="s", col="blue", ylim=range(0,13))
  lines(output.T, output.Q/scale.K, type="s", col="green", ylim=range(0,13))
  lines(output.T, output.R/scale.K, type="s", col="black", ylim=range(0,13))
  lines(output.T, output.S/scale.K, type="s", col="yellow", ylim=range(0,13))


  # Output
  data.frame(time=output.T, N=output.N/scale.K, M=output.M/scale.K,
    Q=output.Q/scale.K, R=output.R/scale.K, S=output.S/scale.K)
}
```

Bibliography

[1] D. Aldous. Stopping times and tightness. *Ann. Probab.*, 6:335–340, 1978.

[2] J. Blath, N. Etheridge, and M. Meredith. Coexistence in locally regulated competing populations and survival of branching annihilating random walk. *Ann. Appl. Probab.*, 17:1474–1507, 2007.

[3] T. Bojdecki and L. G. Gorostiza. Inhomogeneous infinite dimensional langevin equations. *Stoch. Anal. Appl.*, 6(1):1–9, 1988.

[4] B. Bolker and S. Pacala. Using moment equations to understand stochastically driven spatial pattern formation in ecological systems. *Theor. Popul. Biol.*, 52:179–197, 1997.

[5] A. Bovier and N. Champagnat. Time scales in adaptive dynamics: directional selection, fast and slow branching. 2009. Preprint.

[6] A. Bovier and S. D. Wang. Trait substitution tree on two timescales analysis. 2011. Preprint.

[7] P. Cattiaux, P. Collet, A. Lambert, S. Martínez, S. Méléard, and J. San Martín. Quasi-stationary distributions and diffusion models in population dynamics. *Ann. Probab.*, 37(5):1926–1969, 2009.

[8] N. Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. *Stoch. Proc. Appl.*, 116:1127–1160, 2006.

[9] N. Champagnat and A. Lambert. Evolution of discrete populations and the canonical diffusion of adaptive dynamics. *Ann. Appl. Probab.*, 17:102–155, 2007.

[10] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. *Probab. Theor. and Relat. Field.*, 148, 2010.

[11] A. Clayton and S. N. Evans. Mutation-selection balance with recombination: convergence to equilibrium for polynomial selection costs. *SIAM J. Appl. Math*, 69:1772–1792, 2009.

[12] P. Collet, S. Méléard, and J.A.J. Metz. Random modeling of adapive dynamics for sexual populations. *Preprint*, 2011.

[13] J. A. Coyne, N. H. Barton, and M. Turelli. Perspective: A critique of sewall wright's shifting balance theory of evolution. *Evolution*, 51(3):643–671, 1997.

[14] C. Darwin. *On the origin of species by means of natural selection.* John Murray, London, 1859.

[15] D. A. Dawson. *Measure-valued Markov processes, in École d'Été de Probabilités de Saint Flour XXI*, volume 1541 of *Lecture Notes in Mathematics* ***1541***. Springer, Berlin, 1993.

[16] D. A. Dawson and A. Greven. Multiscale analysis: Fisher-wright diffusions with rare mutations and selection, logistic branching system. 2010.

[17] I. Eshel. On the changing concept of evolutionary population stability as a reflection of a changing point of view in the quantitative theory of evolution. *J. Math. Biol.*, 34:485–510, 1996.

[18] A. M. Etheridge. *An introduction to superprocesses*, volume 20 of *University Lecture Series.* AMS, Rhode Island, 2000.

[19] A. M. Etheridge. Survival and extinction in a locally regulated population. *Ann. Appl. Probab.*, 14:188–214, 2004.

[20] S. N. Ethier and T. G. Kurtz. *Markov proccesses: characterization and convergence.* John Wiley and Sons, New York, 1986.

[21] S. N. Evans, D. Steinsaltz, and K. W. Wachter. A mutation-selection model for general genotypes with recombination. 2007.

[22] N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximation. *Ann. Appl. Probab.*, 14:1880–1919, 2004.

[23] M. I. Freidlin and A. D. Wentzell. *Random perturbations of dynamical systems.* Springer, New York, 1984.

[24] L. G. Gorostiza and Z. H. Li. High density fluctuations of immigration branching particle systems. *Stochastic Models, CMS conference proceedings*, 26:159–171, 1999.

[25] T. E. Harris. *The theory of branching processes.* Springer, 1963.

[26] M. Hutzenthaler. The virgin island model. *Elect. Journ. Probab.*, 14:1117–1161, 2009.

[27] M. Hutzenthaler and A. Wakolbinger. Ergodic behavioer of locally regulated branching populations. *Ann. Appl. Probab.*, 17:474–501, 2007.

[28] G. Kalliianpur and V. Perez-Abreu. Stochastic evolution equations driven by nuclear space valued martingales. *Appl. Math. Optim.*, 17:237–272, 1988.

[29] N. Kaplan. A continuous time markov branching model with random environements. *Adv. Appl. Prob.*, 5:37–54, 1973.

[30] N. Keiding. Extinction and exponential growth in random environments. *Theor. Popul. Biol.*, 8:49–63, 1975.

[31] N. Keiding and J. Nielsen. The growth of supercritical branching processes with random environments. *Ann. Prob.*, 1(6):1065–1067, 1973.

[32] T. G. Kurtz. Limit theorems for sequences of jump markov processes approximating ordinary differential processes. *J. Appl. Probab*, 8:344–356, 1971.

[33] A. Lambert. The branching process with logistic growth. *Ann. Appl. Prob*, 15:1506–1535, 2005.

[34] A. Lambert. Quasi-stationary distributions and the continuous-state branching process conditioned to be never extinct. *Elect. Journ. Probab.*, 12:420–446, 2007.

[35] R. Law and U. Dieckmann. Moment approximations of individual-based models. *The Geometry of Ecological Interactions: Simplifying Spatial Complexity*, pages 252–270, 2002.

[36] Z. H. Li. Asymtotic behavior of continuous time and state branching processes. *Austral. Math. Soc. Ser. A*, 68:68–84, 2000.

[37] Z. H. Li. *Measure-valued branching Markov processes.* Springer, 2010.

[38] S. Méléard and V. C. Tran. Trait substitution sequence process and canonical equation for age-structured populations. *Journal of Math. Biol*, 58:881–921, 2009.

[39] J. A. J. Metz, S. A. H. Geritz, G. Meszéna, F. A. J. Jacobs, and J. S. Van Heerwaarden. Adaptive dynamics: a geometrical study of the consesequences of nearly faithful reproduction. *Stochastic and Spatial Structures of Dynamical System*, pages 183–231, 1996.

[40] I. Mitoma. Tightness of probabilities on $C([0,1],S')$ and $D([0,1],S')$. *Ann. Probab.*, 11:989–999, 1983.

[41] C. M. Newman, J. E. Cohen, and C. Kipnis. Neo-darwinian evolution implies punctuated equilibria. *Nature*, 315:400–401, 1985.

[42] K. A. Schneider. Long-term evolution of polygenic traits under frequency-dependent intraspecific competition. *Theor. Popul. Biology*, 71:342–366, 2007.

[43] D. Villemonais. Approximation of quasi-stationary distributions for 1-dimensional killed diffusions with unbounded drifts. 2009. arXiv:0905.3636v1.

[44] J. B. Walsh. *An introduction to stochastic partial differential equations*, volume 1180 of *Lect. Notes in Math.* Verlag-Springer, 1986.

[45] S. D. Wang. Trait substitution sequence in accelerated population dynamics. 2011. Preprint.