EXTINCTION, FIXATION, AND INVASION IN AN ECOLOGICAL NICHE

by

MattheW Badali

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy
Graduate Department of Physics
University of Toronto

Abstract

Extinction, Fixation, and Invasion in an Ecological Niche

MattheW Badali
Doctor of Philosophy
Graduate Department of Physics
University of Toronto
2018

This thesis is concerned with demographic stochastics. That is, the randomness inherent in systems with a discrete state space. In biology this arises naturally in ecological systems. Stochastics, as applied in the biological context, was first done by Kimura when calculating the dynamics of gene frequencies in a population. Kimura, and most theoretical ecologists since, employed the Fokker-Planck equation, a partial differential equations method which further approximates the system by assuming continuous population sizes In the context of population ecology, the similar Moran model of two species is the cleanest example of two competing species in an ecosystem, with eventually one going extinct and one fixating after some short characteristic time dependent on the system size. However, this model assumes the two species are identical, and that they compete with each other (interspecies) as strongly as they compete with themselves (intraspecies). Some recent researchers have addressed this by noticing that results similar to that of Moran are found in one limit of the famous generalized Lotka-Volterra equations with stochastic fluctuations. They employ various approximate techniques, usually the Fokker-Planck equation, and explore various metrics of this noisy Lotka-Volterra model, which in other limits has a long average extinction time. None, however, have looked at how the system transitions between its slow and fast limits. Most have also restricted themselves to uncontrolled approximations This is where I situate my research. In this thesis I will show that competing species can coexist unless their ecological niches entirely overlap, and that this niche overlap anticorrelates with a species' ability to invade an established ecosystem. To accomplish this I shall first perform a thorough investigation of the various approximation techniques commonly used in stochastic biophysical modelling on a one dimensional Lotka-Volterra toy model. Thence I will investigate the two dimensional version, in particular to characterize the transition between its regular slow dynamics and the fast times limit corresponding to the foundational Moran model. I do this with an arbitrarily accurate technique for calculating mean fixation times. Finally, by regarding the opposite process to fixation, I can comment on the stability of the 2D model with regards to invasion attempts. The obvious consequence of my research is a better null hypothesis for the dynamics of small homogeneous communities like the human microbiomes. More generally the results are of significance in estimations of timescale for paleontology and phylogeny.

Contents

Chapter 0

Ch0-Introduction

0.1 Introduction

0.2 Motivation and Background

"What is biophysics?" This is a question I have been asked by family and friends innumerable times through my doctorate. The short answer is that it is the application of "physics thinking" to biological problems. Classical biophysics involves laser optics and protein radii. The biggest contingent at a modern biophysical society meeting are those who do molecular dynamics simulations of protein folding. I usually say that I have the mathematical and problem-solving skills of a physicist and am interested in biological problems, specifically ecology. In reality my research would be at home in a physics department, or in applied mathematics, mathematical ecology, systems biology. It has been an issue finding conferences to go to, in that my research tends to be more mathematical and technical than what most biologists care for, but with different end goals than the mathematicians, who are more interested in proofs and existence rather than the physical and biological effects and outcomes. This thesis is presented as a requirement for the completion of my doctorate of philosophy in physics - and that is the last time I will mention physics in this thesis. My field is mathematical biology. Physics falls somewhere in between math and biology, being very quantitative but also concerned with the real world, and so thats where my research ended up. But I could have just as easily been housed in applied mathematics, or ecology, or systems biology, or a number of other niche fields. The biology is the motivation; the mathematics is what allows one to get things done. Mathematical modelling can allow for quantitative predictions, something that was missing from the stamp collecting pre-modern biology. Mathematical modelling is also powerful in its qualitative predictions. In setting up a model we start with our basic understanding of the problem, but the act can enlighten us to the system: it may reveal unexpected phases, or tell us what to perturb (and how) to get a desired effect. The qualitative can also inform government policy or life philosophy.

I apply mathematics to biological problems. In particular, I apply stochastic analysis to problems in ecology. The application of mathematics allows for quantitative analysis of problems, which in turn leads to a greater ability to have statistical certainty when interpreting data. Math can also allow you to uncover unforeseen aspects of a problem. You write down a simple model which is biologically inspired and seems sound, then further investigation yields a novel qualitative regime of the system.

1

Mathematical ecology is the oldest discipline of mathematical biology, with its relevance dating back at least since Malthus used a model of exponential growth to argue that overpopulation would lead to widespread famine and disease, and that was more than two hundred years ago. It is certainly older than modern biology, with DNA only being recognized for what it was sixty years ago. About a century ago, Lotka and Volterra extended the logistic equation of Verhulst and applied it to biological systems, arriving at the famous predator-prey equations. Midway through the last century, Wright, Fisher, and Moran proposed urn models that demonstrate fixation in a way that is easily intuited and also treatable mathematically. Around the same time Kimura was revolutionizing genetics by proposing models that could account for the evolution and eventual fixation or extinction of mutant alleles [43, 85]. Ecology benefited from the island biogeography theory of MacArthur and Wilson [101?]. In the last couple decades there has been debate as to the extent of neutral versus niche effects in ecological dynamics, sparked by Hubbells unified neutral theory of biodiversity and biogeography [71]. The history of mathematical and theoretical biology, especially as applied to ecology, is punctuated by significant models like these inspiring deeper investigations of both the quantitative details and qualitative trends that the biological world might contain.

The application of mathematics to ecology opens up the possibility of addressing a variety of problems central to the field. One of the simplest problems, and one treated in this thesis, is: What is the probability of and timescale over which a species will go extinct in an ecosystem [? ?]? There is the related question of: Given two competing species in a system, what is the probability of extinction of either species before the other, and the timescale over which this occurs? In an ecosystem with competing species, when all but one species has gone extinct, that final species is said to have fixated in the system. The lifetime and extinction of species is both of theoretical interest and a pressing concern for humanity, as we exist in an epoch of unprecedented rates of extinction. Conservation biology is a driving motivation for me in both my academic and personal life. It is concerned with managing and maintaining the biodiversity on Earth, to avoid excessive rates of extinction. Biodiversity, simply put, refers to the number of species or genetic strains in an ecosystem. In more detail, biodiversity is sometimes characterized by allele frequency within a species or the abundance distribution of different species. The abundance distribution is the curve that results from binning each species based on its population in the system, such that the first bin indicates the number of species that have a local population of only one organism (or a number falling in the first bin's range), the second bin is the number of species with abundance two (or a population in the second bin's range), and so on.

I would like to highlight the issue of biodiversity, one of the persistent problems in modern ecology [32, 105, 126?]. In 1961 Hutchinson published "The paradox of the plankton [75], in which he speculated about an apparent contradiction: for plankton living in the upper layer of the ocean far from shore there are few different resources on which to live, yet there is an immense diversity of different species of plankton that appear to coexist. Surely those that reproduce the quickest or use the resources most efficiently would outcompete all others such that only the fittest would survive. This principle of competitive exclusion, sometimes called Gauses Law [?] states that "two species cannot coexist if they share a single [ecological] niche." In systems with few resources and therefore few niches, one expects that only few species will persist at any given time. The expectation is that in this homogeneous ecosystem with extreme nutrient deficiency the competition should be severe, and only a few species should persist, many fewer than the number observed. A variety of solutions have been proposed to resolve the paradox of the plankton but there is as yet no consensus [139]. These include: the system is approaching a

steady state of fewer species but very slowly; there exist other limiting factors like resources or toxins overlooked by scientists that help define more niches; environmental fluctuations or oscillations stabilize the system; spatial heterogeneity allows for local extinction but supports the great biodiversity on larger length scales; the system is stabilized by life-history traits of the plankton; the system is stabilized by the presence of predators to the plankton; there is symbiosis or commensalism between the various plankton species. This lack of consensus is a gap in the literature, one that I contribute to by calculating the lifetime of a species, either surviving independently or undergoing competition with another species of varying similarity.

The competitive exclusion principle is sometimes considered tautological [**?**]. To others, it can be derived, as through mathematical models that have the dynamics of two species trending toward the death of one or the other of them [22, 101, 107]. Its veracity and applicability have been debated since before it was named [94, 101**? ? ?**]. Most theories explaining competitive exclusion, especially those which are mathematical in nature, make an argument from limiting factors. These are factors external to the species that affect its birth or death rate. They can be abiotic, like nutrients, toxins, waste products, or living space, or these factors can be biotic, like predators or prey. A series of papers from McGehee and Armstrong [9, 10, 107] showed that, if coexistence is defined as having a stable fixed point with positive population of multiple species in a deterministic differential equations model of species and limiting factors, coexistence of all species is impossible if the number of species is greater than the number of limiting factors. That is, the number of different species that can coexist is limited to the number of different limiting factors. In an ecosystem there are a finite number of different limiting factors; when it is full of its allowed number of species and additional species enters the system it either dies out or will replace one of the existing species. This is exactly what the principle of competitive exclusion predicts. Note that these limiting factors can be ones that affect a species' rate of birth or its rate of death. In either case, two species do not interact with each other directly; rather, the presence of one species modifies the amount of factor existent in the system, which in turn affects the birth rate and/or death rate of the other species, and vice versa. There are some subtleties to coexistence or the absence thereof, which I will be exploring in this thesis, but it suffices the reader to know that the idea of limiting factors is one theory which justifies the competitive exclusion principle, albeit with discrete niches.

The theories dealt with in this thesis have many applications. Most obvious, and arguably most pressing to society, is the realm of conservation biology. Biodiversity is often used as an indicator of the health of an ecosystem [108, 127, 130**? ?**]. A clearer understanding of the forces that maintain biodiversity could provide new and easier metrics for evaluating the health of an ecosystem, and hence the efficacy of various conservation efforts. The mechanisms of species maintenance are related to those of speciation, and an ecosystem losing stability can refer to both its collapse or invasion of a foreign species. Invasion of a new mutant or immigrant strain or species into the system is a problem deeply intertwined with that of biodiversity maintenance [71]. This problem too is of obvious interest in the study of ecosystems. Invasion is also relevant in the domain of health care. We are only recently learning, for example, about the composition of the microbiome in humans and its relation to health [37, 88, 91, 102, 145]. The balance of different species in ones gut seems to be important for avoiding illness. Imbalance of the microbiome composition, or invasion of a new species, can greatly impact a person's wellbeing, and a theory of whether an invasion will be successful and how long it might persist would go a long way toward diagnostics and prognostication. The other end of the process, namely the

extinction of a species, also has a number of applications. Other than the obvious modern ecological ones, extinction times are useful in paleontology. The fossil record shows a number of species in different epochs, and these data make more sense in the light of a consistent theory of species survival and eventual decline. Similarly, extinction and fixation times are already used in the construction of phylogenetic trees [21, 134, 136]. The more accurate a theory of extinction timescales developed, the more precisely we can perform phylogenetic analyses. Mapping existent species to their common ancestors falls under the purview of coalescent theory [87]. This is part of the impact of my research, in that I calculate extinction times to arbitrary accuracy, using a controlled approximation largely ignored in the literature.

0.3   Neutrality

The simplest version of coalescent theory and phylogenetic tree reconstruction is based on neutral models [87, 134]. Neutral models, especially those of Wright, Fisher, and Moran, are minimal models that show random extinction and fixation. These models allow not just for fixation probabilities but also the distribution of times such a random occurrence might take. In fact these models can describe any system where individuals of different species or strains undergo strong but unselective competition in some closed or finite ecosystem, especially those constrained by space. Such ecosystems include any of the microbiomes, of humans [37, 88] or others [25, 121, 135, 145, 150, 155]. These microbiomes have limited space and so any death of an organism is quickly replaced by the birth of another. Immigration is rare due to the closed nature of the system. Other example systems have a limited number of resources hence a finite number of species, and due to a lack of mobility or distance from biodiversity reservoirs do not often see the introduction of new species, as in the soil or the ocean surface [39, 55].

The Moran is a classic urn model used in population dynamics in a variety of ways, yet it is easy to arrive at, requiring only a few simplifying assumptions. Whether these are justified depends on the situation being regarded, so they should be applied judiciously. The misapplication or unthinking application of assumptions is one of the motivations of chapter one of this thesis. The first assumption toward the Moran model is that no individual is better than any other in terms of reproducing faster or living longer; that is, whether an individual reproduces or dies is independent of its species and the state of the system. This makes the Moran model a neutral theory, and any evolution of the system comes from chance rather than from selection. The next assumption is that the population size is fixed, owing to the (assumed) strict competition for resources or space in the system. That is, every time there is a birth the system becomes too crowded and a death follows immediately. Alternately, upon death there is a free space in the system that is filled by a subsequent birth. In the classic Moran model each pair of birth and death event occurs at a discrete time step (cf. the Wright-Fisher model, where each step is longer and involves N of these events, but the limiting dynamics are the same). This assumption of discrete time can be relaxed without a qualitative change in results, as will be reviewed in chapter 3. The Moran model is most appropriate for modeling a small system of asexually reproducing organisms, like bacteria in an enclosed space.

In the Moran model, each time step involves a birth and a death event. For each event the participating species is chosen with a chance proportional to its abundance in the system. Since a species is equally likely to increase or decrease each time step, the model is akin to an unbiased random walk, which is a solved problem. And since each event has an equal probability of happening for a given species, the frequency of that species tends to stay constant on average. However, due to the randomness inherent

to model a species' frequency in fact fluctuates. This fluctuation is not indefinite: there are two states from which the system cannot exit and thus only accumulate in probability of occurrence. These static states are extinction and fixation: the species has no chance of reproducing when at zero population (extinction) and does not change abundance when it is the only species in the system (fixation) as it constantly is both reproducing and dying with unit probability each time step. Both of these cases are absorbing states, so called since once the system reaches either it will stay in that state indefinitely. In this system we can define the first passage time as the time the system takes to reach either fixation *or* extinction. The first passage time can also be calculated, and its mean gives an estimate of the time two species will coexist in a system (or the inverse fixation rate of the system).
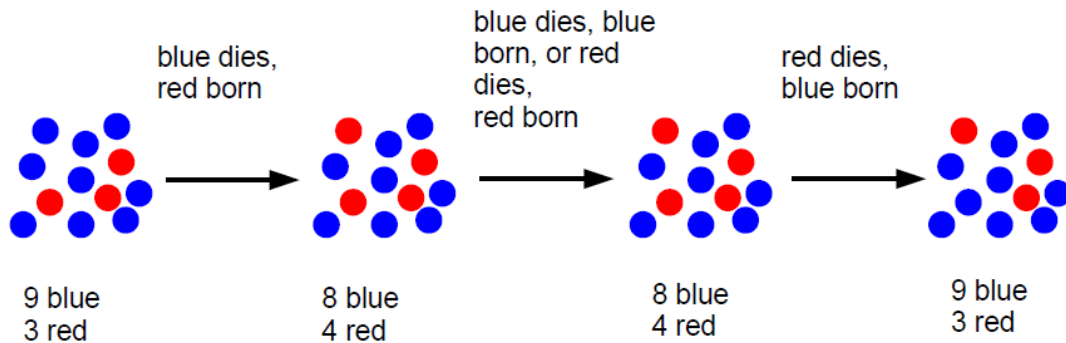
Figure 1: *Example Time Steps of the Moran Model* Here is a sample Moran model with $K = 12$ individuals, initially $n = 3$ of which are red. In the first time step, a red individual is chosen to reproduce (which would happen with probability 3/12) and a blue one dies (probability 9/12). This increases the number of red individuals in the system. Other possibilities each time step are that the number of reds remains the same or decreases. There is a non-zero chance that in as few as three steps a colour will have fixated in the system. Over time the probability of fixation increases such that it is almost certain the system will fixate eventually. Once only one colour remains in the system the chance that a different colour reproduces (and is thus introduced into the system) is zero, since there are none of that different colour around to reproduce.

The unbiased random walk underlying the Moran model suggests that it is a neutral theory. In short, a neutral theory is one for which intraspecies interactions are the same as interspecies interactions. That is, an organism competes equally strongly with members of its own species as with those of other species. No species is distinguished or exceptional in a neutral theory. Thus, unless the whole system's net population is increasing or decreasing, a given organism (and hence its species) is equally likely to reproduce or die, and on average its species abundance is constant. Whether and why different species should regard each other the same as themselves is a matter of debate and apologists [71, 93, 137**?**]. I would like to clarify the difference between neutral theories and those that are simply symmetric. One could formulate a model where intraspecies interactions are different than interspecies interactions, but the intraspecies interactions are the same for each species, as are the interspecies interactions. In a symmetric model a given species behaves an another would in its situation, but not necessarily as another does, given that they are in different situations (namely, who species are typically at different abundances). In a symmetric theory an exchange of labels between two species has the same effect as an exchange of population sizes. For the bulk of this thesis I deal with symmetric theories, with a neutral theory being one limit thereof. Neutral theories are a subset of symmetric theories, since a neutral theory in which each species does not distinguish between self and others automatically allows for an exchange

of species labels with no noticeable effect beyond exchanging abundances.

In the background section I mentioned some of the historic greats in mathematical biology, including Wright, Fisher, and Moran. Kimura and Hubbell also fall under the banner of those who developed neutral theories. Under the approximation of continuous population fraction the Moran model effectively becomes that of Kimura [86]. Kimura was inspired by alleles rather than species, but the rationale is similar. Alleles are the different variants/species of a gene, the segment of DNA that serves a single function. Most non-lethal mutations to an existing allele tend to leave its function entirely unchanged, which clearly makes for a neutral theory. Whereas Moran deals with discrete numbers of individual organisms, Kimura approximates the state space of allele populations as continuous, choosing to deal with allele frequency rather than number. Applying the Fokker-Planck approximation to the Moran model obtains the same probability equations as Kimura, hence the claim that Kimura's results are similar to those of Moran. In each generation each organism provides many copies of its genome, which are chosen indiscriminately (because each organism has two copies of its genome, a factor of two shows up in Kimura's fixation time results when compared those of Moran). Following a few assumptions, Kimura calculates the new mean and variance of the system after one generation of breeding, which are applied in a diffusion equation. Kimura's model can be modified to include many biological effects, like selection. The works of Kimura are well-respected and highly motivated a change in biology to be more quantitative and predictive. Most of Kimura's predictions are numerical by necessity, as no nice analytic forms exist for the solutions. In chapter 3 of this thesis I arrive at some analytical results to describe qualitatively different regimes of a Moran model with immigration, and compare these outcomes to some of Kimura's results.

The seminal work of Hubbell [71] is also similar to that of Moran, but Hubbell is a much more controversial figure than Kimura. Whereas Kimura regarded allele mutations which were often synonymous and therefore neutral, Hubbell argues that different species also follow neutral behaviour and calculates the steady state abundance distribution that follows from such an assumption plus a constant influx of immigrants. Hubbell assumes that each organism from any species competes equally with all others, and therefore as with Moran the species' probability of reproducing or dying is proportional to its fraction of the population. But Hubbell does not predict fixation probabilities and times. Rather, he calculates the distribution of species abundances that should be present within his neutral model, given that there is immigration into (or speciation in) the system and that each immigrant is from a new species. Following the arguments of Hubbell, one can get an estimate of the expected biodiversity of a community, the number of species that should exist in the trophic level (those species which generally consume upon the same set of prey and are preyed upon by the same set of predators). The abundance distribution he predicts matches well with experimental observations in a variety of biological contexts, from trees to birds to microbiomes [71]. The Moran model with immigration analyzed in chapter 3 can be thought of as a modification of Hubbell's theory with recurring immigrants from the same species. While I do discuss abundance distributions I also calculate extinction probability and timescale, something Hubbell's work does not address, as he was motivated entirely by the big picture, indifferent about the average dynamics of an individual species.

As stated previously, Hubbell's neutral theory is contentious. The idea that each species competes with all others to the same degree as intraspecies competition strains credibility. However, slight perturbations from Hubbell's theory do not significantly alter its results. Furthermore, supporters claim that in some sense the different species are equivalent and behave neutrally, which is why Hubbell's theory

seems to work so well in such disparate ecologies [72]. The examples presented in Hubbell's seminal book are compelling, and there may be some truth to these claims. The other side of the debate insists that species differences are the cause of observed abundance distributions. In particular, the environment can be divided into various ecological niches, and it is how these niches are uniquely used by their occupying species that determines the biodiversity of the ecosystem. This is broadly known as niche theory. Niche theory itself has a contentious past mired by confusions. I will do my best to provide a summary here, to contextualize my research.

0.4   Niches

The Moran model shows fixation in a system, so what advantage does niche theory have? Firstly, the concept of a niche is intuitive, certainly more intuitive than neutrality of Hubbell's theory of biodiversity and biogeography. But Hubbell's theory, with its immigration, does not have exclusion, instead predicting a succession of species on a timescale of order inverse immigration rate. And it is not competitive, in that species are not outcompeting each other, being equally matched as they are (this being the quality that makes it a neutral theory). The competitive exclusion principle states that in any given niche one species will eventually dominate. This begs the question, what is an ecological niche?

The concept of niches is an old one, over a century old, and was popularized by Grinnell [?]. There is therefore over a century of debate as to the meaning of a niche, as there is ambiguity in its use. On the theory of niches, Hutchinson [?] says, "Just *because* the theory is analytically true and in a certain sense tautological, we can trust it in the work of trying to find out what has happened" to allow for coexistence of species. In principle, species coexist because they inhabit different niches. Following Leibold [94], I refer to the definition of a niche according its two major uses: as the habitat or requirement niche and the functional or impact niche.

The requirement niche: Grinnell [?] defines a niche as those ecological conditions that a species can live within. These ecological conditions include environmental levels and those organisms on different trophic levels than the species, like their predators and prey, but not those on the same trophic level that might compete with them. Hutchinson [?] has a similar point of view to that of Grinnell, and has provided one of the most enduring conceptualizations of a niche, that of an "n-dimensional hypervolume" in the space of factors that could affect the growth or death of a species. For each factor there is some range at which the species can reproduce faster than it dies out. This is true both for abiotic factors such as temperature, and biotic factors like the concentration of predators. Sometimes these ranges are bounded by zero (eg. cannot survive with no carbon source), sometimes they are unbounded (eg. no amount of prey is too much), and sometimes they depend on the values of the other factors involved (eg. salt is fine for sea creatures so long as there is an appropriate amount of water along with it). But in the space of all these factors, Hutchinson calls the fundamental niche that volume in which the species would have a greater birth rate than death rate. He defines the realized niche as the point or subspace in this high dimensional space that the species effectively experiences, given that it is existing and potentially coexisting in an ecosystem. This also lends a natural definition of niche overlap, as the (normalized) overlap of the fundamental niches of two species [101]. The requirement niche tells us whether the coexistence point of two species is physical, according to simple model of two species [69]. It is inherently linked to the argument of limiting factors as the delimiters of niches as outlined earlier [9, 10, 107].

The other usage of the term niche, that of a functional or impact niche, was popularized by Elton [?] and MacArthur & Levins [101]. Whereas the requirement niche focuses on what factors a species needs to live, the impact niche looks at how the species affects these factors. Their conception of a niche describes how a species influences its environment, or how that species fits in a food web; essentially, what role it plays in an ecosystem. This idea is especially attractive to those who study keystone species [32, 93, 105] (those species that play a disproportionate or critical role in maintaining an ecosystem) but is easily understood from an elementary understanding of what an ecosystem is. By way of example, in every ecosystem with flowers there is something that pollinates them. Whether the pollinator is a bird or an insect species is irrelevant; this role exists in the ecosystem, and so a species evolves to occupy this niche. The niche, in this view point, is the role the species plays in the ecosystem with regards to the other species and the environment; how it impacts the system. As per one simple model of two species, the impact niche tells us whether a coexistence point of two species is stable [?].

Both of these categories of semantics for the word niche have their use. The literature shows attempts to resolve the discrepancies that arise when the two definitions are at odds [93, 94]. This thesis tends to favour the requirement niche definition, based on an argument of limiting factors and explained more fully in chapter 2, but ultimately remains agnostic to the debate. So long as niches exist in some sense, and a niche overlap parameter can be defined, the results I arrive at are sound.

0.5   Deterministic Models

The limiting factor argument for defining the number of niches in an ecosystem is largely taken from Armstrong and McGehee [9, 10, 107]. These arguments are summarized in chapter 2. In the same place I show an example of starting from two limiting factors and arriving at a generalized Lotka-Volterra model. The original Lotva-Volterra model was introduced around a century ago to describe the dynamics of a population of a predator and its prey. It can be seen as an extension of the Verhulst, or logistic, equation, from one to two dimensions. In its modern incarnation the generalized Lotka-Volterra model is typically written as

$$\dot{x}_1 = r_1 x_1 \left(1 - x_1/K_1 - a_{12}x_2/K_1\right)$$
$$\dot{x}_2 = r_2 x_2 \left(1 - a_{21}x_1/K_2 - x_2/K_2\right).$$

The generalized Lotka-Volterra model is the accepted terminology for a dynamical system that depends linearly and quadratically on the populations modelled, with no explicit time dependence. The Verhulst model is one of these equations with its $a = 0$. The classic Lotka-Volterra model is attained by taking the K's to infinity, keeping the a/K ratios positive and finite, choosing r to be negative for the predator and positive for the prey. This predator prey model has oscillating dynamics about a center fixed point. To restrict our investigation to viable species in the same trophic level (treating predators and prey of the species of interest as being some of the limiting factors) we assume K is finite and r is positive. More details of the Lotka-Volterra model will be provided as they become relevant, particularly in chapter 2. Chapter 1 is inspired by the single logistic equation, while chapter 3 further explores the 2D generalized Lotka-Volterra model then considers a Moran model with immigration. Some authors [34, 38, 97, 159] have observed that for certain parameter values that the stochastic 2D generalized Lotka-Volterra model exhibits dynamics similar to those of the Moran model. They did not examine how the effect on the

dynamics as the Moran limit is approached; the transition to this limit is one of the main investigations of this thesis.

The parameters in the Lotka-Volterra equation are easy to understand, albeit hard to measure. The turnover rate r gives the maximum growth rate a species can achieve, specifically when first colonizing an empty system, such that the intraspecific ($1/K$) and interspecific (a/K) competition terms are small. The parameter K is called the carrying capacity of the ecosystem, the maximum population the system can sustain in the absence of competitor species, given the resources available and other limiting factors present in the system. Together, these two parameters, which are the only two that show up in a single logistic equation $\dot{x} = rx(1 - x/K)$, motivate r/K selection theory, coined by MacArthur and Wilson [99]. The theory of r/K selection posits that there is a trade-off between the quantity and the quality of offspring, so most species favour either having many offspring (r-selection) or fewer high-quality offspring (K-selection) that persist close to the carrying capacity of the system. This heuristic fell out of favour in the 1980s as it had ambiguities in interpretation when compared to data. The other parameters in the Lotka-Volterra equations are the a's. These parameters represent the niche overlap between the two species, or the ratio of interspecific to intraspecific competition. They can be derived from limiting factors in at least two ways (see [101] for one example and [100] or chapter 2 of this thesis for a different argument). There is an unresolved debate in the field as to how niche overlap should be measured or defined [1, 2, 31, 33, 74, 90, 94, 128, 129, 140**?**].

The confusion and debate that surrounds niche overlap and other such parameters originates because they are phenomenological parameters rather than strictly physical ones. A phenomenological parameter is one that is consistent with reality without being directly based on physical interactions. In principle these parameters can be derived from physical, measureable ones: the efficiency of a bacterium digesting one molecule of glucose and storing the energy in ATP can be characterized/measured, as can the rate of glucose uptake and its concentration in a system; all of these factors along with a myriad of others combine to generate the carrying capacity of the system. The problem is that there are too many factors, and so many are unknown, that it is easier simply to subsume them all into the one phenomenological parameter of carrying capacity and use that in our modelling and analyses. Phenomenological parameters can be measured: after some time of growing in the sugar water the bacteria will reach a (roughly) steady number; this is the carrying capacity.

To avoid future confusion, I shall provide a quick note about how the term model is used in this thesis. For the first chapter, a model is defined by its specific choice of parameters as well as its mathematical form. This conflicts with how many people use the word model, to mean the mathematical equations with generic parameters unassigned values. I also adopt this usage in chapters 2 and 3, as I explore the effect of taking the niche overlap to unity, such that the interspecies interactions become as strong as the interspecies ones.

The history of mathematical ecology is punctuated by arguments and controversy, large and small. I have already outlined a few, including the interpretation of competitive exclusion, of what is meant by a niche and the difference between requirement and impact niche, what is niche overlap and how to measure it, the applicability of r/K selection theory, and whether neutral theories are appropriate in an ecological context or whether niche theories suffice. There is also a long history of both use and misuse of statistics [**?**]. Within the mathematics side of mathematical biology, other than the statistics necessary to deal with experimental data, the formulation of models can be categorized as either deterministic or stochastic. This is not a debate so much as a choice of utility, of what is most appropriate to model a

given system. It is possible to switch between deterministic and stochastic dynamics.

Deterministic models are characterized by the use of differential equations, or, more rarely, difference equations. Again this comes down to a matter of utility and applicability. Difference equations are equations of the form $x_t = f(x_{t-1})$, where the state at at time t depends in some way on the state at a previous time. They are appropriate for situations where time can be considered discrete, as with species that have seasonal mating or periodic life cycles. Both differential and difference equations can also be "delayed", depending on earlier states, but this is only infrequently used because it results in equations that are difficult to solve.

Deterministic differential equations can be divided into ordinary differential equations (ODEs) and partial differential equations (PDEs). In an ecological context this choice depends on what is being modelled. If the system is well-mixed, if every constituent member organism or moiety could in principle interact with any other, an ODE is appropriate. This well-mixed assumption is one I employ throughout the entirety of this thesis. In contrast, PDEs allow for the modelling of spatially distributed systems, such as chemicals diffusing during embryonic development [70], nutrient distribution in a biofilm [25], or muskrats and minks across Canada [66]. Spatially distributed systems, which allow for local competitive exclusion while globally maintaining biodiversity, are proposed as one resolution to the paradox of the plankton [32, 139].

Another common category of differential equations is stochastic differential equations. These are not deterministic; rather, two realizations of a solution with the same initial conditions can give different results. Stochastic differential equations, and stochastic analyses more generally, are introduced in the next section.

0.6   Stochastic Analysis

As stated before, a stochastic version of the two-dimensional generalized Lotka-Volterra model makes up the bulk of this thesis. Stochasticity is the technical term for randomness or noise in a system. Whereas over time the solution to a logistic differential equation simply increases continuously (and differentiably) toward its asymptote at the carrying capacity, a stochastic version allows for deviations from this trajectory, sometimes decreasing rather than steadily increasing toward the steady state, and thereafter fluctuating about the carrying capacity. See figure 0.6 for a visual example. We need to include stochasticity in our models because of nature's inherent randomness. Especially with the course-graining and phenomenological modelling done in biology for which we cannot account all elements it is necessary to include randomness in our models. Depending on the system of interest, this stochasticity may or may not be relevant: it is usually most important for systems with highly variable environments or small typical population sizes. Beyond biology, there are applications of stochasticity in many disciplines, including linguistics, economics, neuroscience, chemistry, game theory, and cryptography, to name a few [21, 23, 24, 67, 68, 125, 141, 158]. In the biological context Wright and Fisher were pioneers in applying randomness and statistical reasoning. Since then there have been renaissances in the stochastic treatment of genetics due to Kimura and ecology due to Hubbell, and with new mathematical and computational developments it is popular today.

I have made an argument for the use of stochasticity in our modelling to more accurately capture the physical world. The fluctuations caused by this stochasticity also empower us to find new features in our models. In rare cases the fluctuations can bring a system to a population of zero, in which case it does
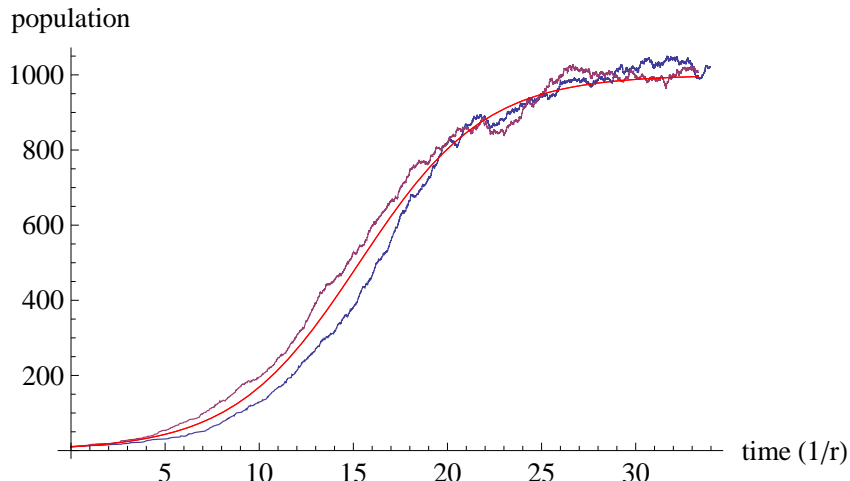
population

Figure 2: *A single logistic system with deterministic and stochastic solutions.* The smooth red line shows the deterministic solution to a one dimensional logistic differential equation with carrying capacity $K = 1000$, which the system asymptotically approaches. The jagged blue and purple lines are each an instantiation of a 'noisy', or stochastic, version of the logistic equation, as simulated using the Gillespie algorithm. Notice that the stochastic versions tend to follow their deterministic analogue but with some fluctuations, sometimes being greater than the deterministic result, sometimes being lesser.

not recover. This arrival at zero population is known as extinction, and is the main phenomenon of study in this thesis. Extinction is not typically seen in the deterministic analogue to these stochastic models. Coupled to extinction is fixation, since if all species but one have gone extinct then the remaining one has fixated. The probability of extinction or fixation can be calculated. For both extinction and fixation, the time before this occurs is distributed, and one can define a mean time. More generally in the field of stochastic analysis this is known as a mean first passage time, the mean time before a system first reaches some predefined state or collection of states. Not only the first passage time is distributed; before the system has gone extinct, its own state is a random variable. Any realization of a stochastic system is of course only in one state at a time, but since different realizations will give different trajectories it is necessary to employ statistical tools like a probability distribution to describe the likelihood of being in a given state at a given time. With a frequentist interpretation, the probability distribution also gives how a population will be distributed within different replicate experiments, or independent measurements. If multiple species are independent or equivalent we can infer the abundance distribution from the probability distribution. And if the the system has attains some sort of steady state then the probability distribution should match the frequency of repeated measurements, with the caveat that the measurements are taken infrequently enough that the system can relax back to steady state after each.

Stochasticity comes in two flavours, originating from different causes. It can arise from the extrinsic fluctuations of the environment [79?], in that limiting factors like resource availability or temperature fluctuate over time. To be clear, I am not talking about the natural dynamics of these quantities due to daily cycles or in response to a species affecting them. Rather, a system at $300K$ might occasionally, and randomly, have one patch warmer than the average, and another part cooler. The more abstracted and phenomenological the model, the less clear the cause of these fluctuations, but the more likely they are to occur. If the sources of noise are independent and many, an invocation of the central limit theorem suggests that a phenomenological parameter will have a Gaussian probability distribution about its

mean value. But even if the environment is entirely controlled, there can be stochasticity in the system. Whereas a deterministic system like the logistic one shown in figure 0.6 has a continuous solution with the population growing smoothly to the carrying capacity, this is not possible in a real biological system, as the number of organisms is quantized. There can be two bacteria or three, but not two and a half. Constraining the system to integer values, and the inherent randomness in the birth and death times of the individuals, leads to demographic noise [13, 34, 38, 48, 52, 56, 60, 97, 159]. Demographic stochasticity is the focus of my thesis. For both environmental and demographic stochasticity it is usually obvious how to recover the deterministic analogue, by taking the noise to zero. Going the other way, from deterministic to stochastic, is obvious for incorporating environmental noise only; the inclusion of demographic fluctuations is less trivial, and is one of the focuses of chapter 1 of this thesis.

It is accepted in the literature that demographic noise in a system whose deterministic analogue has a stable fixed point leads to extinction times scaling exponentially in the system size [42, 48, 80, 92, 160**?**]. That is, if K is the constant or mean system size, then demographic fluctuations lead to:

$$\tau \propto e^{cK}$$

for some constant c. This scaling is most readily observed in the logistic system [7, 13–15, 49, 120**?**], which is also covered in chapter 1. For the record, environmental noise in the logistic system is polynomial [122**?**]:

$$\tau \propto K^d$$

for some constant d. Polynomial dependence is also found when there is no fixed point in the deterministic analogue, or one of neutral stability, like the Moran model [41, 48]. When the deterministic fixed point is unstable extinction happens even in the deterministic limit, and is logarithmic when starting from the fixed point [48, 92]:

$$\tau \propto \ln(K).$$

In all these cases K is the system size, typically taken to be some measure of the magnitude of the fixed point when relevant. Often this fixed point is the carrying capacity. For those systems where the fixed point is stable, the extinction time also does not tend to depend on the initial conditions [34], as the deterministic draw to the fixed point is greater than the destabilizing effects of noise, and it is only a rare fluctuation that leads to extinction. A mean time to extinction that is exponential in the population size is commonly considered to imply stable long term existence for typical biological examples, which have large numbers of individuals [76, 122]. Thus a sub-exponential extinction time implies exclusion of a species, and a reduction of the biodiversity of the ecosystem.

Demographic fluctuations are best modelled using the master equation. The master equation describes the evolution of a probability distribution function. It is a differential equation in time and a difference equation in the population size, which accounts for the integer number of organisms. Stochastic analysis of systems with environmental noise is done using the Kolmogorov equations, the forward equation of which is more commonly known in the physics community as the Fokker-Planck equation. Equivalent to the Fokker-Planck equation is the Langevin equation, which is the easiest formulation of a stochastic equation to envision. A Langevin equation is also known as a stochastic differential equation (SDE) and is a regular differential equation or series of equations with a random noise term added. The solution is therefore a random variable. Simulating a particular realization of the solution gives a different trajectory every time. Instead, for all random variables, to solve a system means something

different. Typically what is meant by solving is either finding the probability distribution function, or its moments, or just the first moment. When referring to the extinction time, as I do throughout this thesis, I imply the mean time to extinction (MTE), or more generally the mean first passage time. For this reason both the master equation and the Kolmogorov equations describe the evolution of the probability distribution function. Using the Kramers-Moyal expansion one can approximate the master equation as a Fokker-Planck equation. Both are hard to solve: a solution can be found for one dimensional systems, but in general not for higher dimensions. The dimensionality, in an ecological context, is given by the number of distinct species or strains being modelled. I will provide more details throughout the thesis, but especially in chapter 1 where I investigate various approximations to the master equation. For most of my research I calculate the extinction time exactly, following a textbook formulation, or at least to arbitrary accuracy [119, 120]. There also exist many approximation techniques to deal with stochastic problems, as I briefly outline below.

SDEs can be simulated similarly to regular DEs, with a smaller time step giving a more accurate solution. Particular realizations of solutions to the master equation are found via the Gillespie algorithm, also knows as the stochastic simulation algorithm [26, 58]. The probability distribution associated with these particular solutions is found by aggregating many simulations, can be used to verify the aptitude of various approximations. The existence of a system size parameter K raises the possibility of approximation to the master equation. The aforementioned Fokker-Planck equation is an expansion of the master equation in $1/K$ to continuous populations, going from a difference-differential equation to a partial differential equation. The results tend to look Gaussian distributed about the deterministic dynamics and near stable fixed points. As stated previously, extinction originates from a rare fluctuation away from the fixed point to zero population, so the Fokker-Planck approximation is expected to perform poorly. It nevertheless does better than expected, and its ease of use makes it a popular choice in the literature. Another popular approximation is the WKB expansion. Rather than just expanding about the fixed point as is the case for Fokker-Planck, WKB expands about the most probable trajectory. Most of my own approximations are more accurate, though I occasionally make use of the Fokker-Planck approximation as a supporting technique to allow for analytic intuition. The main technique employed in this thesis is related to the formal solution to the master equation. The MTE comes from inverting the matrix of transition rates, which in principle is semi-infinite, accounting for population values between zero and infinity. By introducing a cutoff to the matrix I can calculate the MTE. Varying the cutoff allows for arbitrary accuracy. In this way I find the extinction times for two species systems more accurately than any other approximation approach employed in the literature. This in turn allows me to capture not just the exponential dependence on carrying capacity that dominates the MTE, but also the prefactor, which becomes relevant as the Lotka-Volterra system transitions to the Moran limit.

0.7   Structure of Thesis

The major questions of this thesis are: What are the probability and timescale of a single species extinction in an ecosystem? How should the probability and MTE be calculated? Inspired by problems of biodiversity, what is the mean time to fixation of two competing species? Conversely, what is the probability and timescale of invasion of a second species into an ecosystem occupied by a first? The structure of the thesis is as follows.

First, I use the exact techniques mentioned above and introduced more completely in sections 2.3

and 2.4 to investigate a one dimensional logistic system, comparing the influence of the linear and quadratic terms to the quasi-steady state distribution and the MTE. I find that those species with high birth and death rates, and those for whom competition acts to increase death rate rather than reduce their birth rate, tend to go extinct more rapidly. With the simplicity of this test system I explore the applicability of various common approximation techniques. I conclude the Fokker-Planck approximation works well close to the deterministic fixed point, but incorrectly estimates the scaling of the extinction time with system size. The WKB approximation performs better, but misidentifies the prefactor to the exponential scaling. The exact techniques and the approximations together make up chapter 1, regarding a one dimensional system. This chapter is being prepared as a paper for publication [**?**].

The natural extension from a one dimensional logistic is to couple two such systems together; this arrives at the two dimensional generalized Lotka-Volterra system and is the subject of the next chapter, chapter 2. The mean time to fixation is used as a tool to diagnose the longevity of the two interacting species. The overlap of their ecological niches is the parameter that controls the transition between effective coexistence and rapid fixation. I determine that two species will effectively coexist unless they have complete niche overlap, even if they have only a slight niche mismatch. Along with the MTE, my analysis uncovers a typical route to fixation, or rather a lack of a typical route, the discussion of which wraps up this chapter.

The final chapter introducing novel research, chapter 3, extends the scope of this thesis to invasion of a new species into an already occupied niche. I calculate the probability of a successful invasion as a function of system size and niche overlap. Then the MTE conditioned on the success of the invasion is analyzed. I discover that the closer the invader is to having complete niche overlap with the established species, the less likely it is to successfully invade, and the longer an invasion attempt will take before it is resolved. Once these timescales are developed, I regard the Moran model modified to account for repeated invasions of the same species. I identify the critical value of the immigration rate above which a species will have a moderate population size and below which the population is either large or largely absent in its contribution to the abundance distribution. Chapter 2 and half of chapter 3 together form another paper being reviewed for publication [16].

The conclusions chapter covers a variety of topics: I explore applications and extensions of the results arrived at in this thesis; I address the central problems introduced in this preliminary chapter and draw some conclusions informed by my results; and I suggest next steps for this research, both continuations and implementations to novel situations.

Chapter 1

Ch1-SingleLogistic

pre-intro note

This chapter is based on a paper written by myself, Jeremy Rothschild, and our supervisor Anton Zilman, which is currently be prepared for submission [**?**].

1.1   Introduction

As deterministic dynamics have a longer history of being applied in a biological context than their stochastic analogues, and as deterministic mathematics are easier to solve, many researchers start with a deterministic approach to their problem of choice. This is not a bad thing; it allows them to get a sense of the problem if noise is minimal or negligible, which is often the case. One problem with going from deterministic to stochastic dynamics is that the mapping is not unique; many stochastic problems give the same deterministic limit as noise becomes small. For this reason I argue that the stochastic description should be explicitly chosen and motivated, for any analysis which involves stochastics even in part. In this chapter I will justify my argument in two ways, both of which use the ubiquitous example of the Verhulst, or logistic, model. Using the metrics of the quasi-stationary probability distribution function (QPDF) and the mean time to extinction (MTE) I will show that the allocation of linear and nonlinear contributors between the birth and death rate has a drastic effect, and I will evaluate the validity of various commonly employed approximation techniques.

The deterministic equation I consider is the logistic equation, one of the most common models to describe a biological system [5, 6, 15, 62, 116, 118, 120, 122**?**]. It shows up in epidemiology [12**?**], biodiversity [**?** **?**], and generally as a default for modelling a population that grows to a constant value [**?** **?**]. For a population of n individuals, I will be dealing with stochastic equations that give the deterministic limit

$$\frac{dn}{dt} = r\,n\left(1 - \frac{n}{K}\right),\tag{1.1}$$

where r is a rate constant and K is a carrying capacity, a phenomenological measure of the system size. The deterministic equation arises as a large population limit of a stochastic system [57, 119, 138]; namely it is the difference of the stochastic birth and death rates. Therefore when starting from only a deterministic equation there is some freedom to choose the stochastic rates for birth (b_n) and death (d_n). As the choice of birth and death rates contains ambiguity, researchers have leeway in making

their decision, resulting in a variety of similar but distinct models [5, 6, 15, 62, 116, 118, 120, 122**?**]. These models, despite showing the same limit when fluctuations are small, are not equivalent for the stochastic measures chosen in this chapter. The logistic equation includes interactions between organisms by introducing a nonlinearity into the birth or death rates [5, 6, 15, 62, 116, 118, 120, 122**?**]. Biologically this means the per capita birth rate is reduced by the presence of competitors, for instance if the competitors reduce the resource abundance and growth is slowed [114, 152]. Alternatively, the per capita death rate can be increased by neighbours, perhaps due to secreted factors like toxins or waste products introduced by those neighbours [62, 133, 149]. The biological reality determines how this shows up in a mathematical model that captures the growth and decay of the population. I include the parameter δ to account for the stochastic relevance of the absolute values of the per capita birth and death rates, but in the deterministic limit only their difference r affects the dynamics of the system. Parameter q describes where the intraspecies inhibition acts: a high q near unity implies competition for resources and a decreased effective birth rate, whereas a low q near zero reflects more direct conflict, with intraspecies interactions resulting in greater death rates of organisms. Only the difference of the stochastic birth and death rates is observed in the deterministic dynamics, so anything that acts to commensurately change both the birth and death rates is undetected [116, 120]. A systematic exploration of the effect of these "hidden parameters has not been undertaken.

There is a choice to be made when modelling a particular biological system as to how much intraspecies interactions should affect a species birth rate, death rate, or both. The objective of this work is to investigate the impact of this choice on one measurable quantity, the mean time to extinction (MTE). Generally a community is made up of many species; mathematically the dimensionality of the problem is constrained to the number of species [10]. This will be elaborated upon in the next chapter. In most cases, only the one dimensional MTE can be solved exactly [120]. In more complicated situations an approximation is necessary, and there exist many such techniques [57, 119]. These techniques tend to rely on a system size expansion and assume that the population is typically large, a reasonable assumption in most biological systems.

First, I will introduce the model in more detail, motivating it and presenting the parameters associated with the ambiguity of the deterministic equation. Then both the steady state population distribution and the MTE will be calculated under different biological assumptions. Various common approximation techniques will be investigated and compared to the exact results. Finally, a discussion of the results will conclude that increasing the birth and death rates commensurately leads to greater population variance and lesser MTEs, and that the choice of model is of critical importance when establishing a system from which to draw conclusions.

1.2   Model

The simplest model of an isolated population has linear birth and death terms (that is, the per capita birth and death rates are constant: $b_n/n = \beta$, $d_n/n = \mu$). The difference between per capita birth and death gives some rate constant r, the Malthusian or exponential growth rate, such that the deterministic per capita growth would be $\frac{1}{n}\frac{dn}{dt} = r$. This model is a classic but gives the outcome of population explosion. Even in the stochastic case, there is a finite probability of population explosion, and the mean diverges [119]. To mathematically curb this infinite growth, and to biologically allow for intraspecies interactions, a non-linear term is required. A quadratic is the easiest non-linearity to handle, such that

$\frac{1}{n}\frac{dn}{dt} = r\left(1 - \frac{n}{K}\right)$. The rate constant is inhibited by the density of the population, hence a decrease by n/K, giving the desired quadratic term. This is exactly the logistic equation 1.1.

Extinction occurs at $n = 0$, with flux from small populations. In this thesis I consider only birth-death processes, so in fact extinction would only occur from the last individual organism dying before reproducing. This adds additional motivation to the choice of a quadratic equation. For any per capita dynamics $r f(n/\tilde{K})$ with some large system parameter $\tilde{K}$ that gives exponential growth at small population we can write an expansion $f(n/\tilde{K}) \approx f(0) + f'(0)n/\tilde{K}$. Defining $K \equiv -\tilde{K}/f'(0)$ we recover the logistic equation for small populations. For example, if exponential growth is inhibited by Michaelis-Menten kinetics such that $\frac{1}{n}\frac{dn}{dt} = r\left(1 - \frac{n}{n+\tilde{K}}\right)$, at small population the dynamics are the same as the logistic equation, with $K = \tilde{K}$. Since I concern myself with extinction, it is exactly the dynamics of small populations that interests me; the population can have different behaviour at large n, but my MTE results should still hold validity. Note that the QPDF will in general be different, and that most approximation techniques considered in this chapter work best near the mean of the QPDF rather than near the small population sizes relevant to extinction.

Extinction occurs at $n = 0$, an unstable fixed point of the logistic equation, whereas there is a stable fixed point at $n = K$. Common practice in dynamical systems analysis is to rescale variables to remove parameters and simplify the system. Since we are dealing with continuous time we can remove the rate constant from our equation; I do so by rescaling the time by r. Similarly, in the deterministic equation 1.1 we could rescale n by K and have no remaining parameters. However, in the stochastic version we cannot apply this latter rescaling, because of the implicit population scale of ± 1 organism for each birth/death event. The integer number of organisms in systems with demographic noise has an implicit population scale of 1.

The system being constrained to integer populations gives a clear example of why the deterministic analysis is insufficient. Instead of the continuous, fractional populations of equation 1.1, I must define birth and death rates. I assume that each birth event is independent and distributed exponentially with a probability $b_n dt$ of occurring in each infinitesimal time interval dt, and this is similarly assumed for death events. In this chapter/to this end I use the birth rate

$$b_n = (1+\delta)\, r\, n - \frac{r\, q}{K} n^2 = r\, n \left(1 + \delta - q\frac{n}{K}\right) \tag{1.2}$$

and death rate

$$d_n = \delta\, r\, n + \frac{r(1-q)}{K} n^2 = r\, n \left(\delta + (1-q)\frac{n}{K}\right). \tag{1.3}$$

Note that I introduce two new parameters in the equations 1.2 and 1.3: $q \in [0,1]$ shifts the nonlinearity between the death term and the birth term, whereas the parameter $\delta \in [0,\infty)$ establishes a scale for the contribution of linear terms in both the birth and death rates. I include the parameter δ to account for the stochastic relevance of the absolute values of the per capita birth and death rates; in the deterministic limit only their difference r affects the dynamics of the system. Parameter q describes where the intraspecies inhibition acts: a q near unity implies competition for resources and a decreased effective birth rate, whereas a low q near zero reflects more direct conflict, with intraspecies interactions resulting in greater death rates of organisms. It can be readily checked that $b_n - d_n$ recovers the right-hand side of equation 1.1 where, as per design, the new parameters q and δ do not appear. The choice of these parameters specifies a particular model and has consequences on the QPDF and MTE.

The model described above has one other notable feature. Except at $q = 0$, there is a population at

which the competition brings the effective birth rate to zero. This is the maximum size the population can achieve, and I define this cutoff as

$$N = \frac{1+\delta}{q}K. \tag{1.4}$$

Therefore I limit my calculations to the biologically relevant range $n \in [0, N]$. Already it is evident that the "hidden" parameters of δ and q have an effect on the system, as different values of the parameters will naturally define a range of states accessible in the model. Note that so long as $q \leq 1$ the death rate is positive semi-definite for the domain of interest. At $n = 0$ both the birth and death rates go to zero: this is the stochastic absorbing state.

1.3   Quasi-stationary Probability Distribution Function

A probability distribution function is a useful mathematical tool to describe the state of a dynamical system. I denote $P_n(t)$ as the probability that the population is composed of n organisms at time t. The evolution of the distribution in the single birth and death process is captured in the master equation

$$\frac{dP_n}{dt} = b_{n-1}P_{n-1}(t) + d_{n+1}P_{n+1}(t) - (b_n + d_n)P_n(t). \tag{1.5}$$

Note that ultimately at large times the probability of being at population size $n \neq 0$ decays to zero, as more and more of the probability gets drawn to the absorbing state. Although this is an important property of this model, it is difficult to describe any dynamics of our model with such a distribution. The approach to the true steady state is slow (on the order of the MTE; see next section below). Prior to reaching extinction, the system tends toward a quasi-stationary distribution. That is, after some decorrelation time, the system reaches a state that only changes very little as it leaks probability into the absorbing state at extinction.

I am interested in this conditional probability distribution function P_n^c: the probability distribution of the population conditioned on not being in the steady extinct state.

$$P_n^c = \frac{P_n}{1 - P_0}.$$

The dynamics of this conditional distribution are described in a slightly different master equation than equation 1.5:

$$\frac{dP_n^c}{dt} = b_{n-1}P_{n-1}^c + d_{n+1}P_{n+1}^c - \left(b_n + d_n - P_1^c d_1\right)P_n^c. \tag{1.6}$$

After an initial transient period, this conditional probability will stabilize to some steady $\tilde{P}_n^c$ for which $d/dt\,\tilde{P}_n^c = 0$. The steady state of this distribution is referred to as the quasi-stationary distribution (qpdf), not to be confused with the true stationary distribution of the population which is the state where $\tilde{P}_n(t \to \infty) = \delta_{n,0}$.

One way to obtain the quasi-stationary distribution is to exploit equation 1.6 in an algorithm which iteratively calculates the change in the distribution ΔP_n^c in an arbitrarily small time interval Δt until all change in the distribution is negligible [16]. Decreasing the time step Δt increases both the accuracy and the runtime, such that an arbitrarily accurate distribution takes a prohibitively long time to calculate. Instead I employ a different algorithm [119]: at steady state equation 1.6 can be rearranged to relate $\tilde{P}_{n+1}$ to $\tilde{P}_n$ and $\tilde{P}_{n-1}$; given that $b_0 = 0$ there is a lower cutoff and so the whole distribution can be
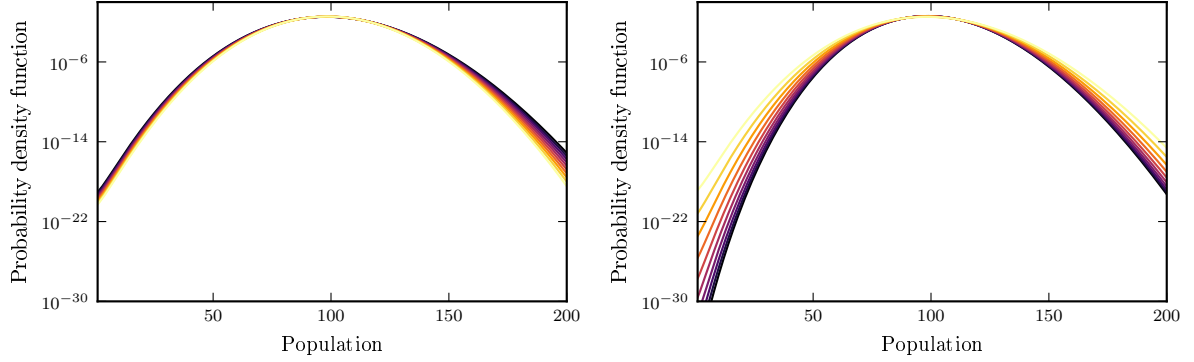
(a) *Probability distribution with $\delta = 1.00$ and $K = 100$*   (b) *Probability distribution with $q = 0.06$ and $K = 100$*

Figure 1.1: *Probability distribution of the population* The conditional probability distribution functions as found using the quasi-stationary distribution algorithm. Note that for each curve, the population cutoff N is outside the domain presented here. In 1.1a increasing lightness indicates an increase in q. Similarly, the lightness increase in 1.1b corresponds to an increase in δ

written in terms of $\tilde{P}_1$, which is then solved by normalization. The former technique is shown in figure 1.1, for different values of q and δ. Increasing the value of δ shifts the mode toward the anterior of the distribution and spreads the distribution out, increasing the variance. Decreasing q has a similar but lesser effect to increasing δ.

1.4   Exact Mean Time to Extinction

As described earlier, the system ultimately goes to the absorbing extinct state. The time in which this happens is a random variable, the mean of which is the mean time to extinction τ_e. In fact, in many cases the MTE nicely characterizes the distribution of exit times, which is typically observed to look roughly exponential. Because the absorbing point is deterministically repelling and, as the QPDF shows, the system spends most of its time near the deterministic fixed point, extinction events are rare, as are trajectories that get close to extinction. These extinction attempts can be considered as almost independent, since the decorrelation time is so much shorter than the time between attempts. The system has repeated, independent events that occur with at a constant rate; it is Poissonian, hence the distribution of extinction times is exponential and described by its mean, the MTE.

For one-species systems it is well known how to exactly solve the MTE for a birth-death process [119? ?]. The mean time of extinction, for a population of size n, is

$$\tau(n) = \frac{1}{d(1)} \sum_{i=1}^{n} \frac{1}{R(i)} \sum_{j=i}^{N} T(j) \tag{1.7}$$

where

$$R(n) = \prod_{i=1}^{n-1} \frac{b(i)}{d(i)} \quad \text{and} \quad T(n) = \frac{d(1)}{b(n)} R(n+1).$$

Combining equations 1.2 and 1.3 with the solution for the mean time to extinction 1.7 obtains a complicated analytical expression in the form of a hypergeometric sum. Little intuition can be gained from
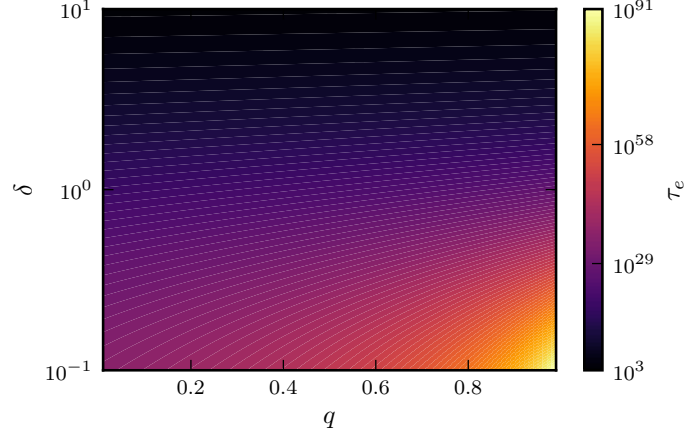
Figure 1.2: *Exploring the mean time to extinction in the parameter space* Recall that q shifts the non-linearity between the birth and death rates: for $q = 0$ the nonlinearity is purely in the death rate, for $q = 1$ nonlinearity appears only in birth. The birth and death rates are increased simultaneously with δ.

the mathematical expression, but the numerical results of the MTE, as shown in figure 1.2, are more interpretable.

(a) *Varying δ*
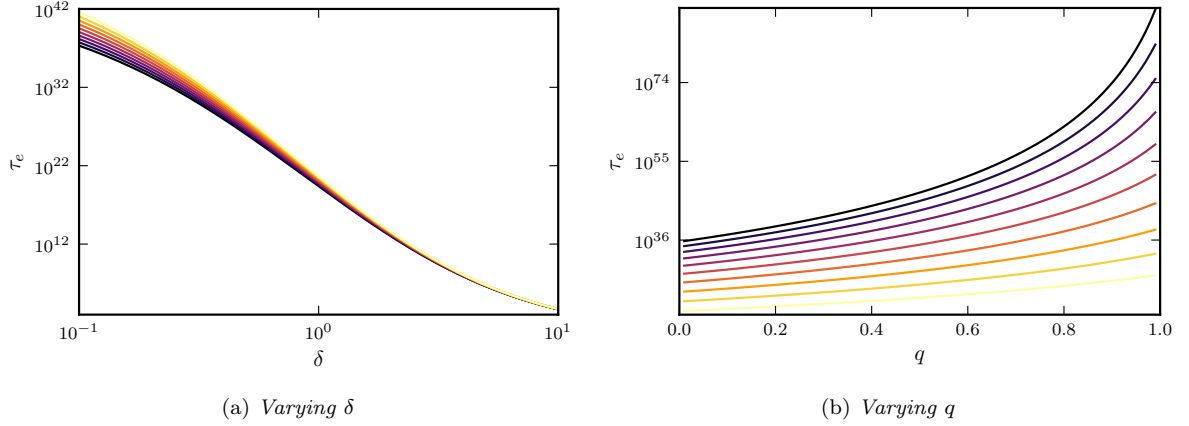
(b) *Varying q*

Figure 1.3: *Mean time to extinction for varying δ and q* Each line represents a slice in Figure 1.2: Figure 1.3a are vertical slices which show how, for different values of q, the δ affects τ. Similarly Figure 1.3b are horizontal slices which show how, for different values of δ, the q affects τ. As in Figure 1.1, lightness of the line indicates an increase of 1.3a q and 1.3b δ

A typical trajectory starting from n goes first to the deterministic fixed point K and fluctuates about that point before a large fluctuation leads to its extinction. Since the time for the population to reach carrying capacity is insignificant compared to the extinction time, the MTE is largely independent of the initial population, and I write $\tau(n) \approx \tau(K) \equiv \tau_e$ for all n. This approximation only fails for small n. It is well known that τ_e goes like e^K [122] and this is indeed what I observe. What is less well known is the dependence on the hidden parameters. It is evident that the MTE depends on the values of δ and q, parameters that appear in the births and deaths but do not appear in the deterministic equation. Increasing the scaling of the linear terms δ in birth and death rates has a tendency to decrease τ_e. On

the other hand, shifting the nonlinearity from the death to the birth rate, in other words increasing q, causes an increase in τ_e. Note however that the effect of q is magnified for smaller values of δ and weaker for larger values of δ: see figure 1.3.

1.5   Approximations

As shown above, for a one-species model it is possible to write down a closed form solution for the MTE τ_e. However, finding a general solution for the mean time to extinction given multiple populations, and therefore higher dimensions, is not as trivial. Nor can a nice analytic expression be found, even for a single species. Models of stochastic processes away from equilibrium are also difficult to study. Many approximations have been developed to accommodate these complications. These approximations make the calculations possible or reduce the computing runtime significantly; therefore it is important to know which of these tools to use and when they are applicable. Unfortunately the regime of parameter space in which each approximation is valid is not very well understood. Using the same model system of a single logistic species I shall evaluate these approximations and compare them to the above exact results, in order to gain insight into their utility.

The first approximation I will regard is the Fokker-Planck (FP) equation, which approximates the discrete populations as continuous, while still maintaining stochasticity. It is equivalent to writing a Langevin equation [?]. Starting from the master equation 1.5 and expanding the ± 1 terms as $P_{n\pm 1} \approx P_n \pm \partial_n P_n + \partial_n^2 P_n$ we arrive at the popular Fokker-Planck equation:

$$\partial_t P_n(t) = -\partial_n\big((b_n - d_n)P_n(t)\big) + \frac{1}{2}\partial_n^2\Big((b_n + d_n)P_n(t)\Big). \tag{1.8}$$

In going from the master equation to FP I have used the Kramers-Moyal expansion [57]. Pawula theorem says that when employing the Kramers-Moyal expansion the only valid orders to stop the expansion are at two terms or infinite [?]. I have been cavalier; more precisely, we need a large parameter, which I denote K, to do the expansion. Typically in bio the large parameter is volume. Then $x = n/K$ is something like density, and is the parameter about which we expand, requiring that $|1/K| \ll 1$. Furthermore, the expansion requires that the rates W_n can be written in a form $W_n/K = w(x)$. The more pedagogical Fokker-Planck equation is then

$$\partial_t P(x,t) = -\partial_x\big((b(x) - d(x))P(x,t)\big) + \frac{1}{2K}\partial_x^2\Big((b(x) + d(x))P(x,t)\Big), \tag{1.9}$$

which is equivalent to equation 1.8 above for $b(x) = b_n/K$ and similarly for d. Instead of a difference differential equation for the probability, equation 1.8 is a partial differential equation for the probability density. The first term on the right hand side is called the drift term and corresponds to the dynamical equation at the deterministic limit, when fluctuations are neglected. The second term is the diffusion term and describes the magnitude of the effect of stochasticity on the system. A quasi-steady state can be calculated when the time derivative $\partial_t P_n(t)$ is small. That is,

$$\ln P_n^{ss} \propto \frac{2(b_n - d_n) - \partial_n(b_n + d_n)}{(b_n + d_n)}. \tag{1.10}$$

By analogy with Boltzmann statistics, the right-hand side of the above equation is sometimes referred to as a pseudo-potential [?]. Equation 1.8 can also be solved directly to give a MTE [57]. For ease of

reference, I leave the equation here:

$$\tau(n) = 2 \int_0^n dy \frac{1}{\phi(y)} \int_y^\infty dz \frac{phi(z)}{B(z)} \tag{1.11}$$

In the above, $phi(x) = \exp\left[\int_0^x dn 2A(n)/B(n)\right]$, $A(n) = b_n - d_n$, and $B(n) = b_n + d_n$.

To simplify the situation further the birth and death rates can be linearized about a stable fixed point, which implies a Gaussian solution to the FP equation. More specifically, the drift term is replaced with $(n - n^*)\partial_n(b_n - d_n)|_{n=n*}$ (since $(b_n - d_n)|_{n=n*} = 0$) and the diffusion with $(b_n + d_n)|_{n=n*}$. We only expect this further approximation to hold near the fixed point. If we extend the domain to all space the solution is Gaussian, peaked near the fixed point.

Another method frequently utilized is the WKB approximation []. Generally, the WKB method involves approximating the solution to a differential equation with a large parameter (such as K) by assuming an exponential solution (an ansatz) of the form

$$\P_n \backsim e^{K \sum_i \frac{1}{K^i} S_i(n)}. \tag{1.12}$$

Starting again from the masters equation 1.5, one can immediately apply the ansatz in the probability distribution and solve the subsequent differential equations to different orders in $1/K$[14?].To leading order, only $S_0(n)$ is needed. This method is commonly referred to as the real-space WKB approximation, wherein we obtain a solution for the quasi-stationary probability distribution. Another method, known as the momentum-space WKB, is to write an evolution equation of the generating function of P_n, the conjugate of the master equation, and then apply the exponential ansatz [?]. The momentum-space WKB has been shown to differ from real-space WKB [?], and has gone out of favour. Real-space WKB remains popular. The quantity $S_0(n)$ can be interpreted as the action in a Hamilton-Jacobi equation. Its solution is $S_0(n) = \int_{n=0}^{n^*} \ln\left(\frac{b_n}{d_n}\right)$. From the pdf, the MTE is calculated via $\tau = \frac{1}{d_1 P_1}$ [?].

Rather than approximating the probability distribution function near the fixed point, a different approximation can be done to estimate the probability distribution function near the absorbing state $n = 0$. If the bulk of the probability mass is centered on K then the probability of being close to the absorbing state is small (note that this condition is similar to the quasi-stationary approximation, since the flux out of the system is proportional to the probability of being at a state close to 0). Furthermore, it is assume that the probability distribution function grows rapidly, away from the absorbing state, such that $P_{n+1} \gg P_n$, whereas neighbouring birth and death rates are of the same order [?]. Rewriting the master equation 1.5 as $\partial_t P_n = (b_{n-1} P_{n-1} - b_n P_n) + (d_{n+1} P_{n+1} - d_n P_n)$ one approximates the left hand side as zero and the right hand side as $(-b_n P_n) + (d_{n+1} P_{n+1})$. Rearranging this gives $P_n = \frac{b_{n-1}}{d_n} P_{n-1} = \prod_{i=2}^n \frac{b_{i-1}}{d_i} P_1$. P_1 can be found by ensuring the probability is normalized; despite the sum extending beyond the region for which $P_{n+1} \gg P_n$ is valid, the probability distribution generated from this small n approximation is qualitatively reasonable and is used, in conjunction with approximations that work near the fixed point, to verify numerical solutions [?].

As seen above, certain of these approximation methods permit the calculation of a quasi-stationary distribution. Figure 1.4 gives an instance of the quasi-stationary distribution as a function of n (or x) for a choice of the parameters q, δ and K for each technique. Note that with this scale of the figure the WKB approximation and the quasi-stationary algorithm are not distinguishable by eye, though there are indeed slight differences. In general, the ability of the techniques to successfully approximate the
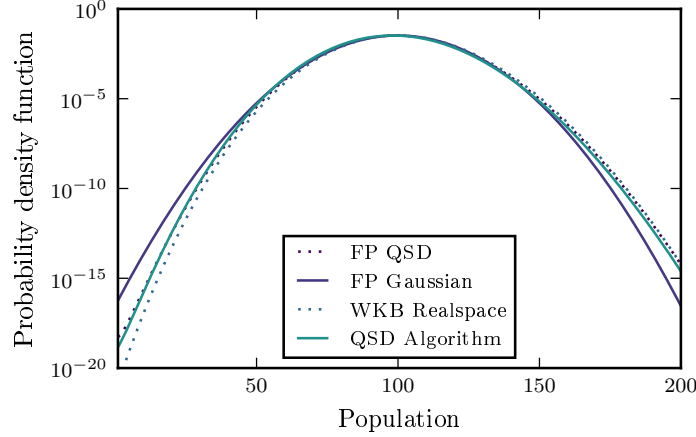
Figure 1.4: *Techniques for calculating a probability distribution function* A comparison of the different probability distribution approximations show how the described dynamics at equilibrium may differ for various techniques.

quasi-stationary distribution depends heavily on the region of parameter space. What I observe is that for large K the celebrated FP approximation is valid in all cases except for low δ and q. For small K it is a poor approximation except low δ and high q. The WKB method fares better, appearing to be reasonable everywhere in q, δ, K parameter space. All other approximations match near the fixed point but fail elsewhere.

It is also possible to obtain the mean time to extinction from these distributions. As described earlier, the quasi-stationary probability distribution leaks from P_n, a non-extinct population, to P_0. As since this is a single step process (with the population only changing by one individual per event), the only transition from which it can reach the absorbing state is through a death at P_1: all population extinctions must go through this sole state. The flux of the probability to the absorbing state is thus given by the expression $d(1)P_1$, hence the approximation [? ?]

$$\tau_e \approx \frac{1}{d(1)P_1}. \tag{1.13}$$

This same equation can be applied to different methods and algorithms that have produced quasi-stationary distributions.

Having calculated the MTE using each approximation, I can now compare the results to the exact solution and verify their accuracy in the parameter space of q, δ, and K. These results are summarized in Figure 1.5. As with the probability distribution function, the difference between the solution of the approximations and equation 1.7 is dependent on q, δ and K. Most of the solutions tend to converge at very low K, though this is unsurprising as the techniques should all approach zero as K decreases and the initial condition of K approaches the final absorbing state at 0. With increasing K the divergence of each approximation becomes more evident. I find that while no approximation works well for large δ, many of them recover the correct scaling in K, albeit off by a factor. For all other parameter regimes, small n and the WKB approximation are both reasonable approximations for the exact results.
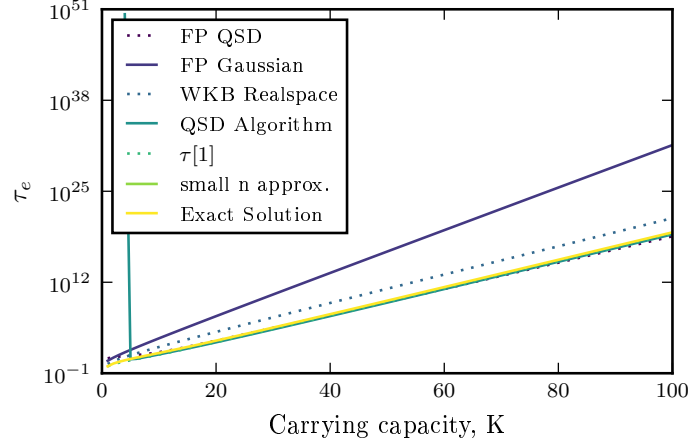
Figure 1.5: *Techniques for calculating the mean time to extinction* Plotted as a function of the carrying capacity, a comparison of the ratio of the MTE of different techniques to that of the 1D sum reveals the ranges for which they are more accurate for approximating τ_e.

1.6   Discussion

What is the justification for the use of q and δ in the birth and death rates? These parameters are products of the assumptions made in constructing the mathematical framework that change the behaviour of the models without affecting the deterministic dynamics. The parameter δ gives a scale for the linear per capita birth and death rates individually (rather than their difference r). This scaling differentiates systems with low birth and death rates from those with high turnover, even when the two models would have the same average dynamics. Whereas the deterministic dynamics includes only the difference of birth and death and will not distinguish between populations with high or low turnover, the size of the birth and death rates is still of relevance. For example, the probability of extinction of a system with linear birth and death rates starting from population n_0 goes as $(b_n/d_n)^{n_0}$ [119]. In this context, the model with high turnover will differ from one with low turnover as the ratio b_n/d_n will depend on the scale, that is to say, on δ.

Additionally, q determines where the quadratic dependence lies, whether more in the birth or death rate. By having the nonlinearity in the birth, one is assuming that the competition, modelled by the quadratic term, slows down the birth rate, for instance in the form of quorum sensing [114]. If it were present in the death the supposition is that the competition instead kills off individuals, for example with illness spreading more rapidly in denser populations []. Although I have worked with a $q \in [0, 1]$, there is no mathematical reason why q could not take values outside this range. Negative values of q would increase both rates, meaning that the density dependence would in fact be beneficial for the birth rates, as in the Allee effect [], and of further significance for the death rates. All $q > 1$ has the opposite effect of negatively impacting both rates, signifying that population density would reduce both the birth and death rates.

Figure **??** summarizes the numerical results of equation 1.7 into a heat map of the mean time to extinction as a function of the two hidden parameters q and δ. For the range of q and δ explored, I find that the MTE changes similarly upon decreasing q and increasing δ, although it depends more sensitively on δ. It is not trivially apparent from the form of the exact analytical solution that the linear

contribution to the rates should be more significant. However, one can get an intuition for why this might be the case. At small populations, the linear term is of order δ (for $\delta \geq 1$) and the quadratic terms is of order q/K, hence the linear term dominates the small population end of the distribution. It is exactly this portion of the distribution that affects the MTE, as seen in equation 1.13.

These qualitative results can be readily intuited by considering the effect each of these parameters has on the distribution, see Figure **??**. A broader probability distribution function corresponds to a shorter MTE, as probability more readily leaks from the quasi-steady state solution to extinction. I find that decreasing δ sharpens the peak of the distribution and slightly shifts the mode posteriorly. As before, the reverse is true for varying q: similar tendencies are observed when q is instead decreased. As the population is allowed more variance about the carrying capacity, states further from the fixed point will be explored more frequently, increasing the probability that the system can stochastically go extinct. Varying the parameters has another effect on the probability distribution, as the parameters determine the maximum population size, restricting the possible states to those less than the population cutoff. It is readily checked, however, that this change in maximum population size has little to no effect on the MTE, by setting manual cutoffs in the numerical analysis and comparing the results to the true MTE. In any case, the mean extinction time of a single, self-interacting logistic species is now known, for all coefficients for first and second order terms.

Regarding the approximations, the best candidate in most regimes appears to be the WKB approximation. It generalizes to multiple dimensions without conceptual difficulty. Mathematically, at higher dimensions WKB necessitates solving a Hamiltonian system, in order to find a likely route to extinction along which to integrate; an analytic solution cannot be derived in general, and a symplectic integrator is necessary to find the numeric trajectory []. Nevertheless, in this one-dimensional case I find an analytic expression for the mean time to extinction using the WKB approximation to be

$$(1.14)$$

We can contrast this with the Gaussian approximation to the Fokker-Planck equation, which always gives an analytic result:

$$\tau_{\text{FP gaussian}} = \frac{2}{1 + 2\delta} \sqrt{2\pi K (1 + \delta - q)} \tag{1.15}$$

Both formulae are dominated by their exponential dependence on K. This is to be expected, as the true solution with $\delta, q = 0$ increases as $\frac{1}{K} e^K$ [92]. It is the prefactor multiplying with the carrying capacity in the exponential that is of critical importance in determining the qualitative behaviour of the MTE. Lacking an analytical form for this prefactor in the exact solution, we can compare these prefactors in each approximation. The Gaussian Fokker-Planck solution has a prefactor $\frac{1}{2(1+\delta-q)}$, so we expect it to underestimate τ_e in the low parameter regime. To lowest order in δ and q the WKB approximation has a prefactor of $1 - 1/K$, which nicely matches the expected limit of 1. The WKB approximation matches well for all δ and q values, as seen in Figures 1.4 and 1.5. The other technique that successfully approximates the true probability distribution and mean extinction time is the small n approximation. It assumes the probability distribution grows rapidly, which is justified for small n and large K. Since the mean extinction time depends only on P_1 this technique gives a reasonable approximation for τ_e. I also note that inverting the transition matrix, as will be done in the following chapters, gives results practically identical to the analytic results.

How can the MTE be probed in a lab setting? For experimentalists the difficulty of measuring a

birth or death rate alone, as it changes with something like population density, varies with the system of interest. However, as previously discussed, measuring the average dynamics alone is insufficient. It is possible experimentally to corroborate some of the claims made in this chapter. For a bacterial species the birth rate could be inferred by the amount of reproductive byproduct present in a sample, for instance factors involved in DNA replication or cell division. The death rate is easily inferred from the birth rate and the average dynamics, or it can be measured using radioisotopes [**?**]. With these two rates and a couple of 96 well plates it should be a routine procedure to probe the quasi-steady state population probability distribution. A patient experimentalist could also verify the dependence of the MTE on δ and q.

The use of approximations is widespread and necessary when using mathematics to model real systems. One takeaway message of this chapter is that one must be mindful in their modelling. I find that FP, WKB, and small n are largely suitable in the models considered here, in that they recover the correct exponential scaling of the MTE with carrying capacity. The biggest caveat is that FP fails for low values of δ and q, and at low K. I will nevertheless use the Fokker-Planck equation to motivate some results in the next chapter, despite selecting a model with $\delta, q = 0$. Other techniques do not fare so well. It is common knowledge that various approximations have situations in which they are more or less applicable. Mindfulness in modelling not only refers to the method of solution, but to the choice of model itself. Historically the choice of model seemed to be one of taste or mathematical convenience [5, 6, 15, 62, 116, 118, 120, 122**?**]. So long as the correct deterministic results were given, the choice of model was not discussed. I have shown that stochastic assumptions have significant qualitative and quantitative effects on at least two metrics of interest in mathematical biology, including the MTE. This does not invalidate the results of past research, but it does imply that the results are valid only for the hidden parameters chosen at that time, and anyone looking to extend or generalize the results should be wary. Going forward armed with the knowledge that not all stochastic models are created equal, I argue that one should give careful regard to the biology of relevance when selecting a model. Our decisions must be informed by the real world if we are to make models that properly capture this biology.

Chapter 2

Ch2-SymmetricLogistic

pre-intro note

This chapter is based on a paper written by me and my supervisor Anton Zilman, currently submitted to Royal Society Interfaces journal [16].

2.1 Introduction

Remarkable biodiversity exists in biomes such as the human microbiome [37, 91, 123], the ocean surface [39, 75], soil [?], the immune system [46, 144, 153] and other ecosystems [115, 146]. Quantitative predictive understanding of long term population behavior of complex populations is important for many practical applications in human health and disease [37, 88, 123], industrial processes [155], maintenance of drug resistance plasmids in bacteria [59], cancer progression [11], and evolutionary phylogeny inference algorithms [21, 87, 134]. Nevertheless, the long term dynamics, diversity and stability of communities of multiple interacting species are still incompletely understood.

The competitive exclusion principle postulates that due to abiotic constraints, resource usage, inter-species interactions, and other factors, ecosystems can be divided into ecological niches, with each niche supporting only one species in steady state, and that species is said to have fixated [65, 106, 113?]. However, the exact definition of an ecological niche varies and is still a subject of debate [2, 4, 28, 32, 52, 75, 94], and maintenance of biodiversity of species that occupy similar niches is still not fully understood [105, 126, 131]. Commonly, the number of ecological niches can be related to the number of limiting factors that affect growth and death rates, such as metabolic resources or secreted molecular signals like growth factors or toxins, or other regulatory molecules [9, 10, 107, 131]. Observed biodiversity can also arise from the turnover of transient mutants or immigrants that appear and go extinct in the population, as in Hubbell's model [29, 45, 71].

Deterministically, ecological dynamics of mixed populations has been commonly described as a dynamical system of the numbers of individuals of each species and the concentrations of the limiting factors [9, 10, 107]. Steady state co-existence typically corresponds to a stable fixed point in such dynamical system, and the number of stably co-existing species is typically constrained by the number of limiting factors. In some cases, deterministic models allow co-existence of more species than limiting factors, for instance when the attractor is a limit cycle rather than a point [10, 143]. Particularly pertinent for this

chapter is the case when the interactions of the limiting factors and the target species have a redundancy that results in the transformation of a stable fixed point into a marginally stable manifold of fixed points. Then the stochastic fluctuations in the species numbers become important [8, 10, 22, 31, 131, 151]. I will return to the mathematical formulation of these concepts later.

Stochastic effects, arising either from the extrinsic fluctuations of the environment [79?], or the intrinsic stochasticity of individual birth and death events within the population [13, 34, 38, 48, 52, 56, 60, 97, 159], modify the deterministic picture. As in the previous chapter, I focus on this latter type of stochasticity, known as demographic noise. Demographic noise causes fluctuations of the populations abundances around the deterministic steady state until a rare large fluctuation leads to extinction of one of the species [34, 97?]. In systems with a deterministically stable co-existence point, the mean time to extinction is typically exponential in the population size [15, 80, 120, 122], as was seen in the previous chapter. Exponential scaling is commonly considered to imply stable long term co-existence for typical biological examples with relatively large numbers of individuals [76, 122].

By contrast, in systems with a neutral manifold that restore fluctuations off the manifold but not along it, mean extinction timescales as a power law with the population size, indicating that the co-existence fails in such systems on biologically relevant timescales [35, 84, 97, 111]. This type of stochastic dynamics parallels the stochastic fixation in the classical Moran-Fisher-Wright model that describes strongly competing populations with fixed overall population size [28, 54, 111, 134, 136, 144, 156?].

As is reviewed below, a broad class of dynamical models of multi-species populations interacting through limiting factors can be mapped onto the generalized Lotka-Volterra model. Lotka-Volterra models allow one to conveniently distinguish between various interaction regimes, such as competition or mutualism, and which have served as paradigmatic models for the study of the behavior of interacting species [8, 22, 31, 34, 38, 48, 52, 56, 83, 97, 151, 159]. Remarkably, the stochastic dynamics of LV type models is still incompletely understood, and has recently received renewed attention motivated by problems in bacterial ecology and cancer progression [28, 34, 52, 81, 144, 149]. There has also been the observation that for certain parameter values that the stochastic 2D generalized Lotka-Volterra model exhibits similar dynamics to the Moran model [34, 38, 97, 159]. How the system transitions from the typical LV results of MTE scaling exponentially with system size to the algebraic times of the Moran model is the main result of this chapter. The results outline the conditions of niche overlap and carrying capacity that allow two species to co-exist (and conversely, those that will lead to relatively quick fixation).

In this chapter, I analyse a model of two competing species with the emphasis on the transition from deterministic co-existence to stochastic fixation. I use the master equation and first passage formalism that enables numerically solution to arbitrary accuracy in all regimes. First I will provide a definition of ecological niche [do I?!?] and a derivation of the competitive LV model, and examine its regimes of deterministic stability. Then I will introduce the stochastic description of the LV model and analyse fixation times as a function of the niche overlap between the two species. These results will be compared to known analytic limits, included here for completeness. I will make further comparisons to the Fokker-Planck and WKB approximations before concluding with a general discussion of the results.

2.2   Long-term stability of deterministic interacting populations

Quite generally, the dynamics of a system of N asexually reproducing species that interact with each other only through M limiting factors (such as food, soluble signaling and growth/death factors, toxins, metabolic waste) and experience no immigration can be described by the following system of equations for the species $x_1, ..., x_N$ and the limiting factor densities $f_1, ..., f_M$ [9, 10, 107]:

$$\dot{x}_i = \beta_i(\vec{f})x_i - \mu_i(\vec{f})x_i, \tag{2.1}$$

where $\vec{f}$ is the state of all factors that might affect the per capita birth rate $\beta_i(\vec{f})$ and the death rate $\mu_i(\vec{f})$ of the species i.

The density of a factor j in the environment, f_j, follows its own dynamical production-consumption equation

$$\dot{f}_j = g_j(\vec{f}, \vec{x}) - \lambda_j(\vec{f}, \vec{x})f_j \tag{2.2}$$

where g_j is a production-consumption rate that includes both the secretion and the consumption by the participating species as well any external sources of the factor f_j, and λ_j is its degradation rate. Alternatively, for some abiotic constrains such as physical space or amount of sunlight, the concentration of the factor f_j can be set through a conservation equation of a form [10, 107] $f_j = c_j(\vec{f}, \vec{x})$.

The fixed points of the $N + M$ equations (2.1) and (2.2) determine the steady state numbers of each of the N species and the corresponding concentrations of the M limiting factors. However, the structure of equations (2.1) imposes additional constraints on the steady state solutions: at a fixed point $\beta_i(\vec{f}) = \mu_i(\vec{f})$ for each of the N species, which determines the steady state concentrations of the M limiting factors $\vec{f}$. However, if $N > M$, the system (2.1) of N equations is over-determined and typically does not have a consistent solution, unless the fixed point populations of $N - M$ of the species are equal to zero [9, 10, 53, 107, 131]. This reasoning provides a mathematical basis for the competitive exclusion principle, whereby the number of independent niches is determined by the number of limiting factors, and a system with M resources can sustain at most M species in steady state.

Nevertheless, as mentioned in the introduction, the number of species at the steady state can exceed the number of limiting factors, when the N equations for the species are not independent and thus provide less than N constraints on the solutions. In this case, at steady state the populations of the non-independent species typically converge onto a marginally stable manifold on which each point is stable with respect to off-manifold perturbations but is neutral within the manifold [8, 30, 48, 97, 107]. I return to this point in the following sections within the discussion of the Lotka-Volterra model.

2.3   Minimal model of interacting species and the derivation of 2D LV model

As a minimal example, in this section we introduce a model of two interacting species whose dynamics is constrained by two secreted factors. Each species x_i has basal per capita birth rate β_i, death rate μ_i, and each generates the secreted soluble factors t_j at rates g_{ji}. Each factor t_i is degraded at a rate λ_i, and affects the death rate of each bacterium linearly with the efficacy e_{ij}. Positive e_{ij} may correspond to metabolic wastes, toxins or anti-proliferative signals [78, 103, 133, 149, 157?], while negative e_{ij} would
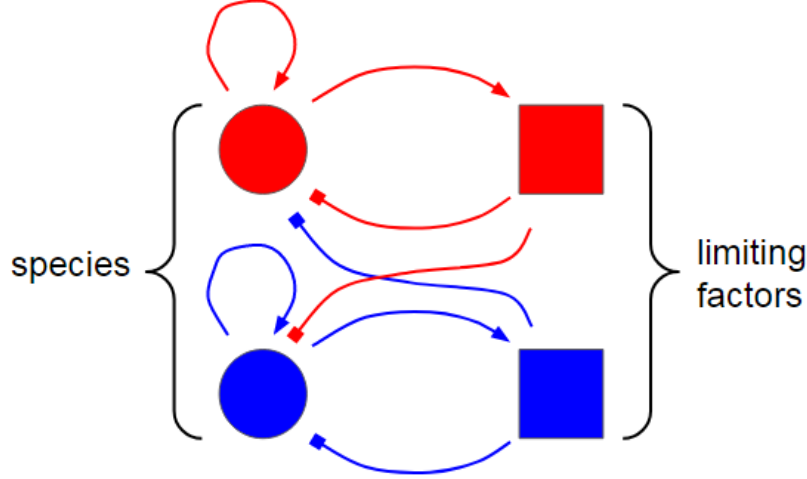
Figure 2.1: *A simple two species two resource model that obtains the Lotka-Voltera model* Each of the two species (here, red and blue circles) reproduces (arrows to self) and produces a toxin (arrows to limiting factors, respectively red and blue squares) which inhibits its own growth (square-ending lines to self) and the growth of the other (square-ending lines to other colour).

describe growth factors or secondary metabolites [103, 154**?**]. The model kinetics is encapsulated in the following equations for the turnover of the species numbers:

$$\dot{x}_1 = \beta_1 x_1 - \mu_1 x_1 - e_{11} t_1 x_1 - e_{12} t_2 x_1$$
$$\dot{x}_2 = \beta_2 x_2 - \mu_2 x_2 - e_{21} t_1 x_2 - e_{22} t_2 x_2, \tag{2.3}$$

and the equations for the production and the degradation of the secreted factors:

$$\dot{t}_1 = g_{11} x_1 + g_{12} x_2 - \lambda_1 t_1$$
$$\dot{t}_2 = g_{21} x_1 + g_{22} x_2 - \lambda_2 t_2. \tag{2.4}$$

It is useful to recast Equations (2.3), (2.4) defining vectors $\vec{x} = (x_1, x_2)$ and $\vec{t} = (t_1, t_2)$, so that

$$\dot{\vec{x}} = \hat{R}\hat{X}\left(\vec{1} - \hat{E}\vec{t}\right) \quad \text{and} \quad \dot{\vec{t}} = \hat{L}\left(\hat{G}\vec{x} - \vec{t}\right), \tag{2.5}$$

where we have the matrices $\hat{X} = \begin{pmatrix} x_1 & 0 \\ 0 & x_2 \end{pmatrix}$, $\hat{L} = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}$, $\hat{R} = \begin{pmatrix} r_1 & 0 \\ 0 & r_2 \end{pmatrix} \equiv \begin{pmatrix} \beta_1 - \mu_1 & 0 \\ 0 & \beta_2 - \mu_2 \end{pmatrix}$,
$\hat{G} = \begin{pmatrix} g_{11}/\lambda_1 & g_{12}/\lambda_1 \\ g_{21}/\lambda_2 & g_{22}/\lambda_2 \end{pmatrix}$, and $\hat{E} = \begin{pmatrix} e_{11}/r_1 & e_{12}/r_1 \\ e_{21}/r_2 & e_{22}/r_2 \end{pmatrix}$.

In many experimentally relevant systems, such as communities of microorganisms and cells, the timescale of production, diffusion, and degradation of secreted factors is on the order of minutes [19], whereas cell division and death occurs over hours [95, 132], and the dynamics of the turnover of the secreted factors can be assumed to adiabatically reach a steady state $\vec{t}^*$ given by $\vec{t}^* = \hat{G}\vec{x}$ [14, 35, 131]. In this approximation the dynamical equations for the species number reduce to

$$\dot{\vec{x}} = \hat{R}\hat{X}\left(\vec{1} - (\hat{E}\hat{G})\vec{x}\right). \tag{2.6}$$

Written explicitly, this becomes the familiar generalized two-species competitive Lotka-Volterra system [22, 34, 38, 44, 48, 96, 100, 159]:

$$\dot{x}_1 = r_1 x_1 \left(1 - \frac{x_1 + a_{12}x_2}{K_1} \right)$$
$$\dot{x}_2 = r_2 x_2 \left(1 - \frac{a_{21}x_1 + x_2}{K_2} \right),$$

(2.7)

where $\frac{1}{K_i} = \frac{e_{ii}g_{ii}}{r_i\lambda_i} + \frac{e_{ij}g_{ji}}{r_i\lambda_j}$ and $\frac{a_{ij}}{K_i} = \frac{e_{ii}g_{ij}}{r_i\lambda_i} + \frac{e_{ij}g_{jj}}{r_i\lambda_j}$. The turnover rates r_i set the timescales of the birth and death for each species, and K_i are known as the carrying capacities. The interaction parameters a_{ij} provide a mathematical representation of the intuitive notion of the niche overlap between the species [2, 33, 101, 140]. When $a_{ij} = 0$, species j does not affect the species i, and they occupy separate ecological niches. At the other limit, $a_{ij} = 1$, the species j compete just as strongly with species i as species i does within itself, and both species occupy same niche. We refer to the a_{ij} as the niche overlap parameters.

The number of deterministically viable species is typically constrained by the number of limiting factors [10], as described in the previous section. Namely, if both matrices $\hat{E}$ and $\hat{G}$ are non-singular and invertible, the solutions to Equation (2.5) are that one (or both) of the species is zero or else $\vec{x}^* = (EG)^{-1}\vec{1}$. The latter solution corresponds to the co-existence of the two species.

When the matrix $(\hat{E}\hat{G})$ is singular ($a_{12}a_{21} = 1$), the co-existence fixed point $\vec{x}^* = (EG)^{-1}\vec{1}$ does not exist, and the Equations (2.5) are satisfied only if the population of one (or both) of the species is zero. Biologically, this condition corresponds to the complete niche overlap between two species, whereby only one species can survive in the niche. (Of note, exclusion of one species by the other can also occur in non-singular cases, as discussed in the next section.) Nevertheless, even in the complete niche overlap case, multiple species can deterministically coexist within one niche if the matrix $(\hat{E}\hat{G})$ possesses a further degeneracy, $K_1/K_2 = a_{12} = 1/a_{21}$, corresponding to an additional symmetry in the interactions of the species with the constraining factors, as illustrated in the next paragraph.

These mathematical notions can be understood in a biologically illustrative example, when the matrix $\hat{E}$ is singular, so that $\det(\hat{E}) = 0$. Any singular 2×2 real matrix can be written in the general form $\hat{E} = \begin{pmatrix} \alpha & \alpha\beta \\ \alpha\gamma & \alpha\beta\gamma \end{pmatrix}$, where α, β and γ are arbitrary real numbers [?]. In this case Equation (2.3) becomes

$$\dot{x}_1 = r_1 x_1 \big(1 - \alpha \, (t_1 + \beta t_2) \big)$$
$$\dot{x}_2 = r_2 x_2 \big(1 - \gamma\alpha \, (t_1 + \beta t_2) \big),$$

(2.8)

so that both secreted factors effectively act as one factor with concentration $t \equiv t_1 + \beta t_2$. With $\gamma \neq 1$ this corresponds to the classic notion of two species and only one limiting factor. The two equations cannot be simultaneously satisfied and the only solution of Equations (2.8) is either $x_1 = 0$ or $x_2 = 0$ (or both). This is one example of competitive exclusion due to competition within a single niche. Finally, when $\gamma = 1$ (corresponding to $a_{12} = 1/a_{21} = K_1/K_2$), both the species and the secreted factors are functionally identical, and the Equations (2.8) allow multiple solutions lying on the line in phase space defined by $\vec{x}^* = \hat{G}^{-1}\vec{t}^*$ and $1 = \alpha \, (t_1^* + \beta t_2^*)$ [38, 107]; in this case many different mixtures of the two species can be deterministically stable, depending on the initial conditions. However, as discussed in the next section, this line of fixed points is unstable with respect to perturbations along the line, and

stochastic effects become important. These derivations above provide a mathematical definition and a biological illustration of the niche overlap between two interacting species, and can be extended to a general case of N species interacting via M factors, as shown in the Supplementary Information. In the next two sections, we study how the niche overlap affects the stability of the species co-existence in deterministic and stochastic cases.

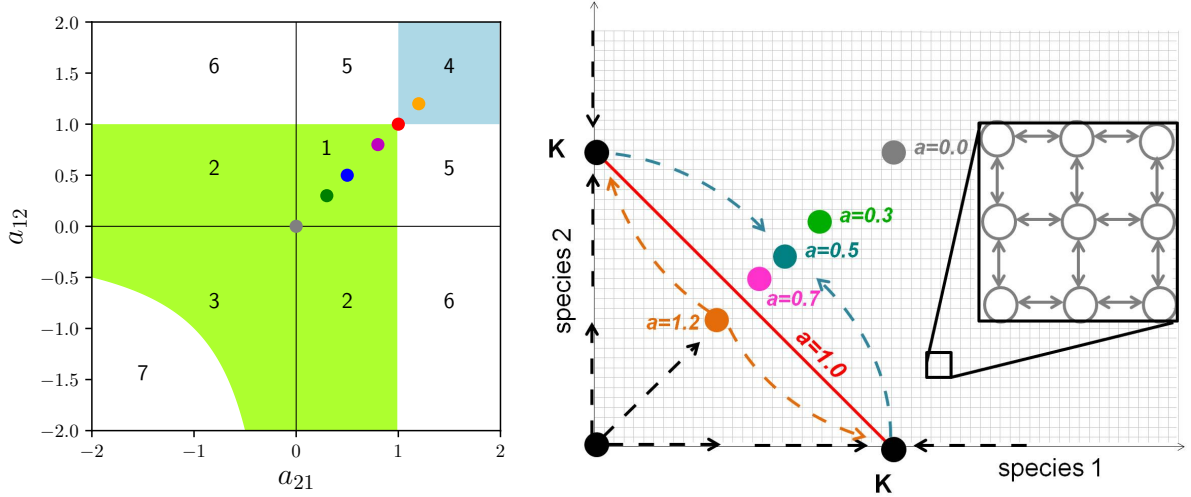## 2.4 Deterministic stability of the Lotka-Volterra model

Figure 2.2: *Left: stability phase diagram of the co-existence fixed point for $K_1 = K_2 = K$. The co-existence fixed point $C = \left(\frac{K_1 - a_{12} K_2}{1 - a_{12} a_{21}}, \frac{K_2 - a_{21} K_1}{1 - a_{12} a_{21}} \right)$ is stable in the green region and unstable in the blue region; in the white regions it is non-biological. Colored dots indicate the parameter range studied in the paper. The numbered regions correspond to different biological different regimes; see text. For the degenerate case $a_{12} = a_{21} = 1$, indicated by the red dot, the co-existence fixed point is replaced by a line of marginal stability, shown in the Right Panel. Right: phase space of the coupled logistic model. Colored dots show C at the indicated values of the niche overlap a. The fixed point is stable for $a < 1$. At $a = 0$ the two species evolve independently. As a increases, the deterministically stable fixed point moves toward the origin. At $a = 1$ the fixed point degenerates into a line of marginally stable fixed points, corresponding to the Moran model. The dashed lines illustrate the deterministic flow of the system: black is for $a = 0.5$, and orange for $a = 1.1$. The zoom inset illustrates the stochastic transitions between the discrete states of the system. Fixation occurs when the system reaches either of the axes. See text for details.*

In this section, I examine the behavior of the deterministic Equations (2.7), which have four fixed points:

$$O = (0,0) \quad A = (0, K_2) \quad B = (K_1, 0) \quad C = (\frac{K_1 - a_{12} K_2}{1 - a_{12} a_{21}}, \frac{K_2 - a_{21} K_1}{1 - a_{12} a_{21}}). \tag{2.9}$$

The origin O is the fixed point corresponding to both species being extinct, and is unstable with positive eigenvalues equal to r_1 and r_2 along the corresponding on-axis eigendirections. The single species fixed points A and B are stable on-axis (with eigenvalues $-r_1$ and $-r_2$, respectively), but are unstable with respect to invasion if point C is stable, reflected in the positive second eigenvalue equal to $r_2(1 - a_{21} K_1 / K_2)$ and $r_1(1 - a_{12} K_2 / K_1)$, respectively. Fixed point C corresponds to the co-existence of the two species and is stable in the green shaded region in the left panel of Figure 2.2, which shows the

stability diagram of the system for $K_1 = K_2$. [34, 40, 117].

The different regions of the phase space in Figure 2.2 have different biological interpretations [1?]. Parasitism, or predation/antagonism, occurs in regions 2 and 6 of (a_{12}, a_{21}) space, where $a_{12}a_{21} < 0$, with one species gaining from a loss of the other. In the strong parasitism regime (region 6), where the positive a_{ij} is greater than one, the parasite/predator drives the prey to extinction deterministically, and the only stable point is the predator's fixed point (A or B). Conversely, weak parasitism (region 2) allows co-existence of both species despite the detriment of one to the benefit of the other [34?].

The regions with both $a_{ij} < 0$ correspond to mutualistic/symbiotic interactions between the species [34, 40, 117?]. Weak mutualism (region 3) is mathematically similar to weak competition in that it results in stable co-existence. Strong mutualism (region 7) results in population explosion. Detailed study of this regime lies outside of the scope of the present work (but see [110]).

The quadrant with both $a_{12} > 0$ and $a_{21} > 0$ corresponds to the competition regime. At strong competition with either a_{12} or a_{21} greater than one (regions 4 and 5 in the left panel in Figure 2.2), either one of the species deterministically outcompetes the other (region 5) or the system possesses two single-species stable fixed points A and B with separate basins of attraction (region 4). The complete niche overlap regime of the underlying model of Equations (2.5) and defined by $\det[\hat{E}\hat{G}] = 0$ is contained within region 4, and is given by the line $a_{12}a_{21} = 1$. These regimes correspond to the classical competitive exclusion theory, together with the strong parasitism case (region 6). By contrast, weak competition (region 1) where both $0 < a_{ij} < 1$ results in the stable co-existence at the mixed point C. In the special case $a_{12} = a_{21} = 1$ (shown by the red dot) the stable fixed point degenerates into a neutral line of stable points, defined by $x_2 = K - x_1$, as shown in the right panel of Figure 2.2. Each point on the line is stable with respect to perturbations off line, but any perturbations along the line are not restored to their unperturbed position [30, 107]. This line correspond to the singular case, discussed in the previous section, where the two species are functionally identical with respect to the action of the secreted factors (eg. $e_{11}/r_1 = e_{12}/r_1$ and $e_{22}/r_2 = e_{21}/r_2$ in Equations (2.5)). The stochastic dynamics along this line correspond to the classical Moran model as discussed below, and in the following we refer to this line as the Moran line.

The right panel of Figure 2.2 shows the phase portrait of the system, in the symmetric case of $K_1 = K_2 \equiv K$, $r_1 = r_2 \equiv r$, and $a_{12} = a_{21} \equiv a$, where neither of the species has an explicit fitness advantage. This equality of the two species, also known as neutrality, serves as a null model against which systems with explicit fitness differences can be compared. In this thesis, I focus on species co-existence in the weak competition regime, finding the scaling of the mean time to fixation due to stochasticity. The asymmetric case is also treated, with results qualitatively similar to the symmetric case.

2.5   The stochastic Lotka-Volterra model

Stochasticity naturally arises in the dynamics of the system from the randomness in the birth and death times of the individuals - commonly known as the demographic noise [13, 50, 124, 148]. Competitive interactions between the species can affect either the birth rates (such as competition for nutrients) or the death rates (such as toxins or metabolic waste), and in general may result in different stochastic descriptions [7, 16], as was discussed in the previous chapter. In this chapter, I follow others [38, 56, 97] in considering the case where the inter-species competition affects the death rates, so that the per capita

birth and death rates b_i and d_i of species i are:

$$b_i/x_i = r_i$$
$$d_i/x_i = r_i \frac{x_i + a_{ij}x_j}{K_i}. \tag{2.10}$$

In terms of the previous chapter, this corresponds to choosing $\delta = 0$ and $q = 0$. In the deterministic limit of negligible fluctuations the model recovers the mean field competitive Lotka-Volterra Equations (2.7) [97].

The system is characterized by the vector of probabilities $P(s,t|s^0)$ to be in a state $s = \{x_1, x_2\}$ at time t, given the initial conditions $s^0 = (x_1^0, x_2^0)$: $\vec{P}(t) \equiv (\ldots, P(s,t|s^0), \ldots)$ [112]. The forward master equation describing the time evolution of this probability distribution is [148]

$$\frac{d}{dt}\vec{P}(t) = \hat{M}\vec{P}(t), \tag{2.11}$$

where $\hat{M}$ is the (semi-infinite) transition matrix. I do not include the absorbing states in my transition matrix, and the master equation 2.11 as written does not preserve probability, as some of it leaks into fixation.

Because the approximate analytical and semi-analytical solutions of the master Equation (2.11) often do not provide correct scaling in all regimes ([14, 16, 49]; see also the previous chapter), I analyse the master equation numerically in order to recover both the exponential and polynomial aspects of the mean time to fixation. To enable numerical manipulations, I introduce a reflecting boundary condition at a cutoff population size $C_K > K$ for both species to make the transition matrix finite [27, 112] and enumerate the states of the system with a single index [112] via the mapping of the two species populations (x_1, x_2) to state s as

$$s(x_1, x_2) = (x_1 - 1)C_K + x_2 - 1, \tag{2.12}$$

where s serves as the index for our concatenated probability vector, uniquely enumerating all the states. In this representation, the non-zero elements of the sparse matrix $\hat{M}$ are $\hat{M}_{s,s} = -b_1(s) - b_2(s) - d_1(s) - d_2(s)$ along the diagonal, $\hat{M}_{s,s+1} = d_2(s+1)$ and $\hat{M}_{s+1,s} = b_2(s)$ at ± 1 off the diagonal, and $\hat{M}_{s,s+C_K} = d_1(s+C_K)$ and $\hat{M}_{s+C_K,s} = b_1(s)$ off-diagonal at $\pm C_K$. Some diagonal elements are modified to ensure the reflecting boundary at $x_i = C_K$.

2.6 Comparison with the Gillespie algorithm

Numerical results obtained from the Gillespie algorithm are accurate, assuming a sufficient number are averaged over [58]. Unfortunately even for a system size as small as $K = 20$ some of the simulations took over ten million steps before fixating. A tau-leaping implementation helps [26], but the problem remains that this fixation is a slow process and simulations of large K will be prohibitively long. As shown above, the distribution of fixation times is roughly exponential. Any simulations that do not finish will be from the tail end of the distribution but will have the largest contribution to the mean time, hence cannot be ignored.

Inverting the truncated transition matrix, as has been done in this chapter, is a much faster compu-
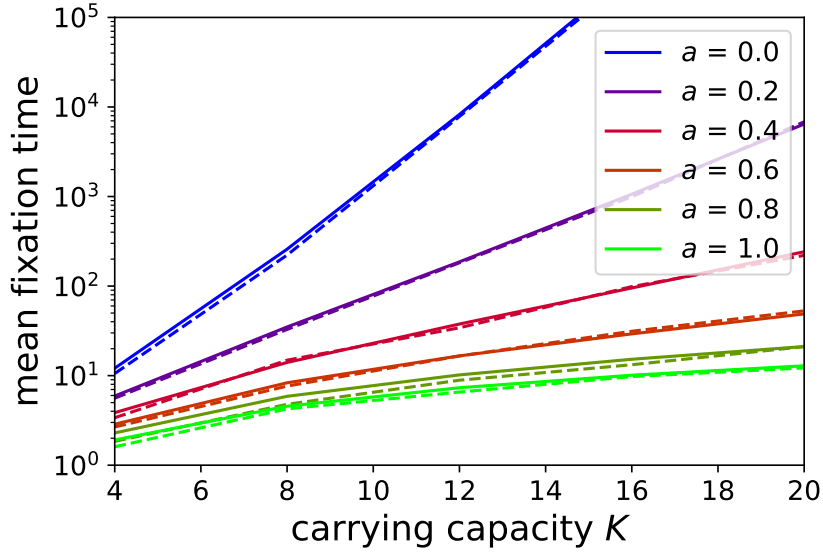
Figure 2.3: *Directly solving the (truncated) master equation agrees with Gillespie simulations.* Solid lines come from directly solving the backwards master equation by inverting the transition matrix, after a cutoff has been applied to the matrix to make it finite. Dashed lines are each an average of a hundred realizations of the stochastic process, as simulated using the Gillespie algorithm.

tational problem, and is hindered by insufficient RAM rather than interminable runtimes. Changing the cutoff means that the solution can be arbitrarily precise. In figure 2.3, the direct solution from inverting the truncated transition matrix compares favourably with the Gillespie simulations.

To ensure accuracy of the mean times to 0.1% or better I choose $C_K = 5K$. This is largely excessive and even $C_K = 2K$ is sufficient for all but the smallest carrying capacities, for which it is least important to be accurate. The sparse matrix LU decomposition algorithm is implemented with the C++ library Eigen [?].

2.7   Mean fixation time in the classical Moran model

Here I derive the mean fixation time for the Moran model [111], which will be used later as a limiting case of the Lotka-Volterra dynamics. Previous authors have already shown that this is the limit [34, 38, 97], and the fixation time for the Moran model is already known [111]. I only include it here for completeness. The Wright-Fisher model gives similar results for large K, but is less intuitable, dealing as it does with a whole generation at a time, rather than one birth and one death.

In the classical Moran model, at each time step, an individual is chosen at random to reproduce. In order to keep the population constant, another one is chosen at random to die. The probabilities that the number of individuals of species 1 increases or decreases by one in one time step are [111]:

$$b_M(n) = f(1 - f) = (1 - f)f = d_M(n) = \frac{n}{K}\left(1 - \frac{n}{K}\right) = \frac{1}{K^2}n(K - n), \qquad (2.13)$$

where n is the number and f is the fraction of species 1 in the system (of total system size K). The mean fixation time, $\tau(n)$, starting from some initial number n of species 1 is described by the following
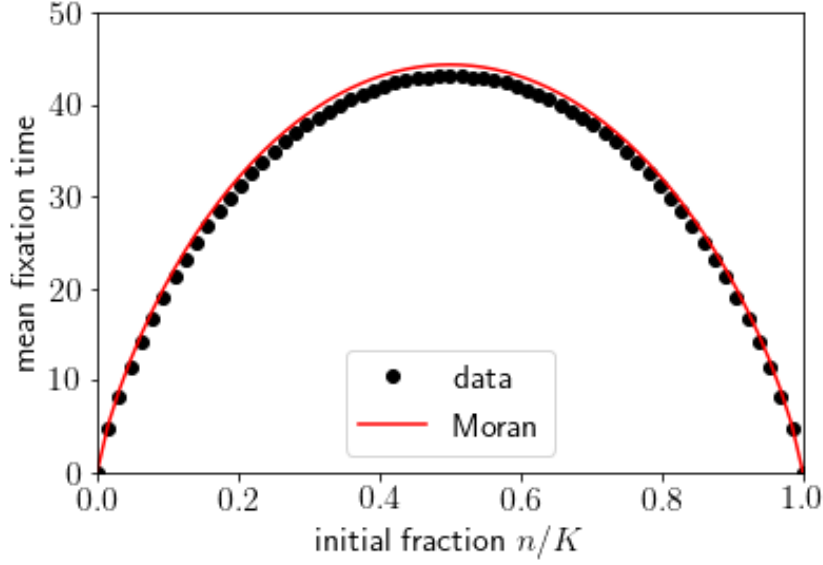
Figure 2.4: *The coupled logistic model agrees with the Moran model in the limit of complete niche overlap,*
$a = 1$. *Fixation time varies with initial fraction of the species in the population. The fixation time for
the Moran model is in red and the coupled logistic model for $a = 1$ is in black. The population size
of the Moran model is set equal to the carrying capacity $K = 64$ of the corresponding coupled logistic
model.*

backward master equation [119]:

$$\tau(n) = \Delta t + d_M(n)\tau(n-1) + (1 - b_M(n) - d_M(n))\,\tau(n) + b_M(n)\tau(n+1),$$

where Δt is the time step. Substituting the values of the 'birth' and 'death' probabilities of species 1
from equation (2.13) we get

$$\tau(n+1) - 2\tau(n) + \tau(n-1) = -\frac{\Delta t}{b_M(n)} = -\Delta t \frac{K^2}{n(K-n)}.$$

At $K \gg 1$, the Kramers-Moyal expansion in $1/K$ results in

$$\frac{\partial^2 \tau}{\partial n^2} = -\Delta t\, K\left(\frac{1}{n} + \frac{1}{K-n}\right).$$

Integrating, using the boundary conditions $\tau(0) = \tau(K) = 0$, gives

$$\tau(n) = -\Delta t\, K^2\left(\frac{n}{K}\ln\left(\frac{n}{K}\right) + \frac{K-n}{K}\ln\left(\frac{K-n}{K}\right)\right). \tag{2.14}$$

For the initial condition analogous to the co-existence point, $n = K/2$, this gives

$$\tau = \Delta t\, K^2 \ln(2).$$

Recall that the Moran model counts one time unit Δt every birth and death pair of events. The

correspondence between the Moran model and the related Wright-Fisher model occurs when the Moran model has undergone a number of births and deaths equal to the (fixed) population size of Wright-Fisher, often called the generation time [?]. The time scale regarded is important. Similarly, for comparison between Moran and the coupled logistic model, one needs to match the time scales. First, note that the birth and death events can be treated as independent under the assumption that a single birth or death does not change the rates significantly, an assumption which is valid far from the axes. Then the probability of the next birth event happening around time t is $f(t)dt = b\,e^{-bt}dt$, and similarly $g(t)dt = d\,e^{-dt}dt$. Then define the probability of the birth happening by by time t is $F(t) = \int_0^t f(t')dt'$, and similarly $G(t) = 1 - e^{-dt}$. Assuming independence, a birth and a death event have happened by time t with probability $F(t)G(t)$, and the distribution of this time is given by $\partial_t[F(t)G(t)]$. The average time of a birth and death is then

$$\Delta t = \int_0^\infty t\partial_t[FG]dt = \int_0^\infty t\partial_t[FG] = \int_0^\infty t\left(f(t) + g(t) - (b+d)e^{-(b+d)t}\right)$$

$$notag$$

$$= \frac{1}{b} + \frac{1}{d} - \frac{1}{b+d} = \frac{b^2 + bd + d^2}{b(b+d)d}.$$

In principle this should be averaged over all combinations of the two species giving birth and dying, weighted by the probabilities of each pairing. It should also account for the probability of being at a particular state, as the state affects the rates. To simplify this I provide a lower bound by assuming that most of the time is spent near the initial state for the Moran limit, $(K/2, K/2)$, such that $b_i\,(K/2, K/2) = d_i\,(K/2, K/2) = K/2$. This gives

$$\Delta t = \frac{3}{2}\frac{1}{b} = \frac{3}{2}\frac{1}{K/2} = \frac{3}{K} \tag{2.15}$$

and therefore

$$\tau = 3\ln(2)\,K. \tag{2.16}$$

The fixation time of the Moran model agrees well with the results of the coupled logistic model for complete niche overlap, as shown in Figure 2.4.

2.8   Fixation time of the coupled logistic model in the independent limit

Here we calculate the mean fixation time in the independent limit of the coupled logistic model given a distribution of extinction times for a single logistic model. The fixation occurs when either of the species goes extinct. Denoting the probability distribution of the extinction times for either of independent species as $p(t)$ and its cumulative as $F(t) = \int_{s=0}^t p(s)ds$, the probability that *either* of the species goes extinct in the time interval $[t, t + dt]$, is the probability of species 1 going extinct while species 2 has not plus the probability that species 2 goes extinct while species 1 has not, since these are the two possibilities. That is,

$$p_{min}(t)dt = \left(p(t)\,(1 - F(t)) + (1 - F(t))\,p(t)\right)dt = 2p(t)\,(1 - F(t))\,dt. \tag{2.17}$$
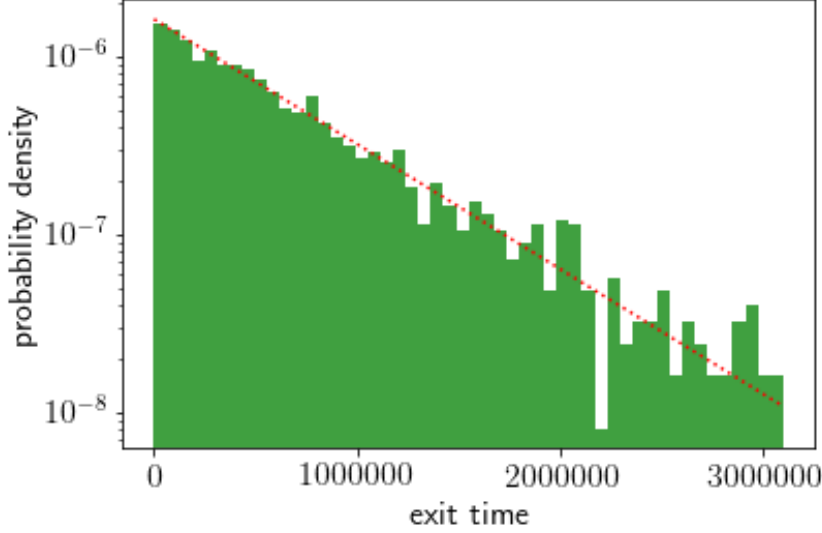
Figure 2.5: *Extinction time distribution of the logistic model is dominated by a single exponential tail.* The bulk of the probability density is modelled by an exponential distribution with the same mean, shown in the red dotted line. Data are generated using using the Gillespie algorithm for $K = 16$. For higher carrying capacities the assumption of exponentially distributed times becomes even more accurate.

The mean time to fixation $\langle t \rangle$ is

$$\langle t \rangle = \int_0^\infty dt \, t \, p_{min}(t). \tag{2.18}$$

As shown in Figure 2.5, the probability distribution of fixation times of a single species is dominated by the exponential tail. It can be approximated as

$$p(t) = \alpha e^{-\alpha t}, \quad F(t) = 1 - e^{-\alpha t} \tag{2.19}$$

with $\frac{1}{\alpha} \simeq \frac{1}{K} e^K$ from the previous section. Finally, we obtain for the mean time to fixation

$$\langle t \rangle = \int_0^\infty dt \, t \, 2\alpha e^{-2\alpha t} = \frac{1}{2\alpha}. \tag{2.20}$$

which leads to the equation $\tau \simeq \frac{1}{2K} e^K$.

2.9 Fixation time as a function of the niche overlap

In this section I calculate the first passage times to the extinction of one of the species and the corresponding fixation of the other, induced by demographic fluctuations, starting from an initial condition of the deterministically stable co-existence point. The master Equation (2.11) has a formal solution obtained by the exponentiation of the matrix: $\vec{P}(t) = e^{\hat{M}t} \vec{P}(0)$. However, direct matrix exponentiation, as well as direct sampling of the master equation using the Gillespie algorithm [26, 58], are impractical since the fixation time grows exponentially with the system size. However, the moments of the first passage times can be calculated directly without explicitly solving the master equation [63]. The mean

residence time in any state s during the system evolution is

$$\langle t(s^0) \rangle_s = \int_0^\infty dt \, P(s,t|s^0) = \int_0^\infty dt \, (e^{\hat{M}t})_{s,s^0} = -(\hat{M}^{-1})_{s,s^0}. \tag{2.21}$$

The final equality in the previous equation is obtained integration by parts and requires that all the eigenvalues of the transition matrix $\hat{M}$ are negative, a fact that is evident by its construction: since the master equation conserves probability (which is bounded by one), none of the eigenvalues can be positive; since the steady state absorbing states have been removed, there are no zero eigenvalues. Thus, the mean time to fixation starting from a state s^0 is [77]

$$\tau(s^0) = -\sum_s \langle t(s^0) \rangle_s = -\sum_s \left(\hat{M}^{-1} \right)_{s,s^0}. \tag{2.22}$$

This expression can be also derived using the backward equation formalism [77]. The matrix inversion was performed using LU decomposition algorithm implemented with the C++ library Eigen [?], which has algorithimic complexity of the calculation scaling algebraically with K. Increasing the cutoff C_K enables calculation of the mean fixation times to an arbitrary accuracy.
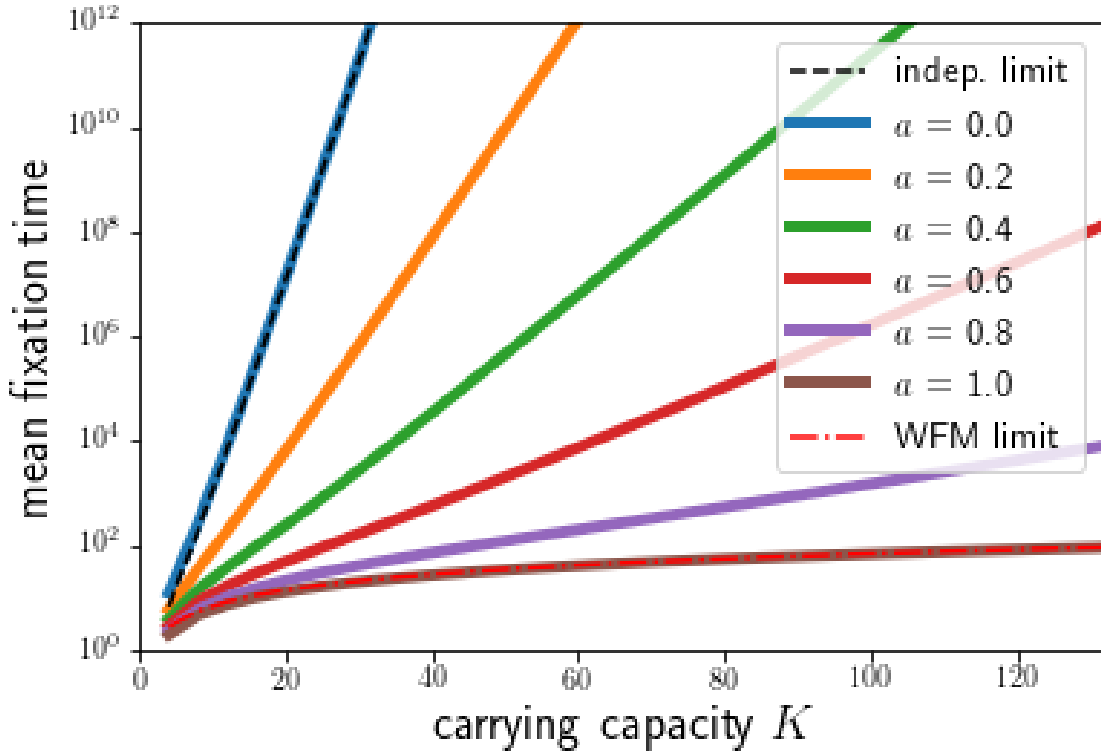
Figure 2.6: *Dependence of the fixation time on carrying capacity and niche overlap.* The lowest line, $a = 1$, recovers the Moran model results with the fixation time algebraically dependent on K for $K \gg 1$. For all other values of a, the fixation time is exponential in K for $K \gg 1$.

The left panel of Figure 2.6 shows the calculated fixation times with the initial condition at the deterministically stable co-existence fixed point as a function of the carrying capacity K for different values of the niche overlap a. In the limit of non-interacting species ($a = 0$), each species evolves

according to an independent stochastic logistic model, and the probability distribution of the fixation times is a convolution of the extinction time distributions of a single species, which are dominated by a single exponential tail [64, 120, 122]. Mean extinction time of a single species can be calculated exactly as in the previous chapter, and asymptotically for $K \gg 1$ it varies as $\frac{1}{K}e^K$ [92] giving for the overall fixation time in the two species model $\tau \simeq \frac{1}{2K}e^K$ as in Equation (2.20). This analytical limit is shown in Figure 2.6 in a black dashed line and agrees well with the numerical results of Equation (2.22). From the biological perspective, for sufficiently large K, the exponential dependence of the fixation time on K implies that the fluctuations do not destroy the stable co-existence of the two species.

In the opposite limit of complete niche overlap, $a = 1$, any fluctuations along the line of neutrally stable points are not restored, and the system performs diffusion-like motion that closely parallels the random walk of the classic Moran model [8, 34, 38, 48, 52, 82, 97]. The Moran model shows a relatively fast fixation time scaling algebraically with K [97, 111], $\tau \simeq \ln(2)K^2\Delta t$; see Equation (2.16). The fixation time predicted by the Moran model is shown in Figure 2.6 as a red dotted line and shows excellent agreement with our exact result. The relatively short fixation time in the complete niche overlap regime implies that the population can reach fixation on biologically realistic timescales.

The exponential scaling of the fixation time with K persists for incomplete niche overlap described by the intermediate values of $0 < a < 1$. However, both the exponential and the algebraic prefactor depend on the niche overlap a. The exponential scaling is expected for systems with a deterministically stable fixed point [14, 49, 52, 56, 122], as indicated in [34, 48, 97] using Fokker-Planck approximation and in [56] using the WKB approximation. However, the Fokker-Planck and WKB approximations, while providing the qualitatively correct dominant scaling, do not correctly calculate the scaling of the polynomial prefactor and the numerical value of the exponent simultaneously [16, 82, 122], as was shown in the previous chapter. For large population sizes and timescales, effective species co-existence will be typically observed experimentally whenever the fixation time has a non-zero exponential component.

To quantitatively investigate the transition from the exponentially stable fixation times to the algebraic scaling in the complete niche overlap regime, we use the ansatz

$$\tau(a, K) = e^{h(a)} K^{g(a)} e^{f(a)K}. \tag{2.23}$$

In the Moran limit, $a = 1$, we expect $f(1) = 0$, $g(1) = 1$, and $h(1) = \ln\left(\ln(2)\right)$ as follows from equation (2.16). In the independent species limit with zero niche overlap, $a = 0$, equation (2.20) suggests $f(0) = 1$, $g(0) = -1$, and $h(0) = -\ln(2)$. Figure 2.7 shows the ansatz functions $f(a)$, $g(a)$, and $h(a)$, obtained by numerical fit to the fixation times as a function of K shown in Figure 2.6. The numerical results agree well with the expected approximate analytical results for $a = 0$ and $a = 1$ with small discrepancies attributable to the approximate nature of the limiting values. Notably, $f(a)$, which quantifies the exponential dependence of the fixation time on the niche overlap a, smoothly decays to zero at $a = 1$: only when two species have complete niche overlap ($a = 1$) does one expect rapid fixation dominated by the algebraic dependence on K. In all other cases the mean time until fixation is exponentially long in the system size [64, 122]. Even two species that occupy *almost* the same niche ($a \lesssim 1$) effectively co-exist for $K \gg 1$, with small fluctuations around the deterministically stable fixed point.
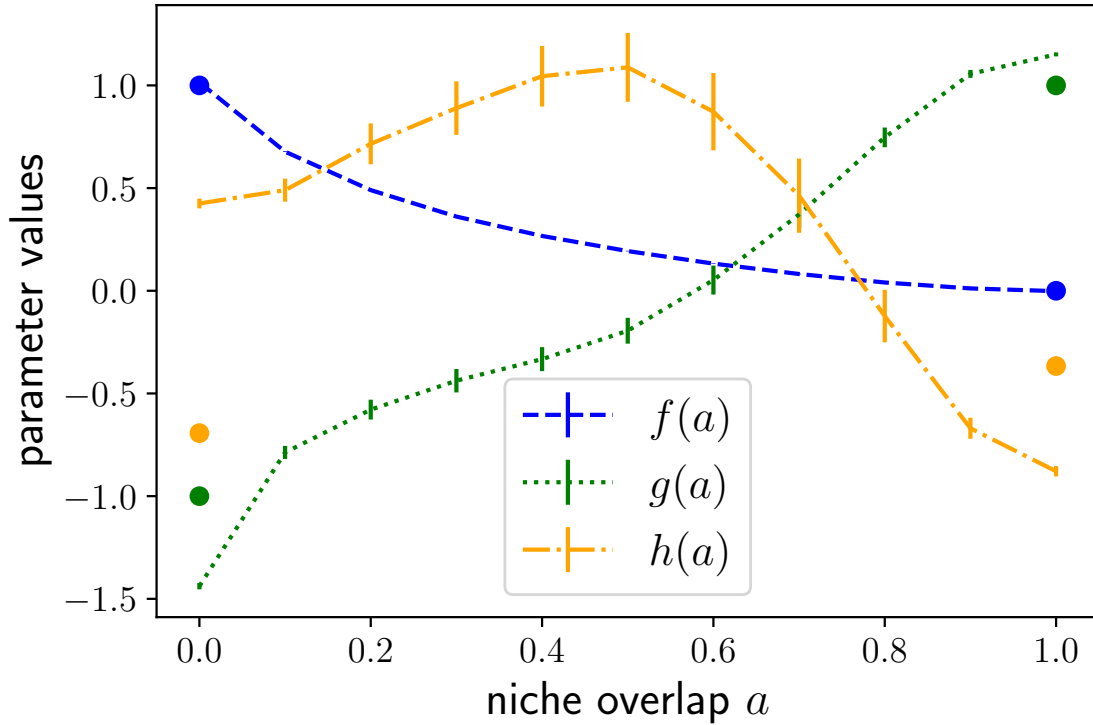
Figure 2.7: *Right: Niche overlap controls the transition from co-existence to fixation.* Blue line: $f(a)$ from the ansatz of Equation (2.23) characterizes the exponential dependence of the fixation time on K; it smoothly approaches zero as the niche overlap reaches its Moran line value $a = 1$. Green line: $g(a)$ quantifies the scaling of the pre-exponential prefactor $K^{g(a)}$ with K. Yellow line: $h(a)$ is the multiplicative constant. Dashed bars represent a 95% confidence interval. The dots at the extremes $a = 0$ and $a = 1$ are the expected asymptotic values.

2.10   Co-existence versus fixation in parameter space

I make the claim that when biological system sizes are large, a fixation time that scales exponentially with carrying capacity effectively implies co-existence. This is typically the case. However, some systems have only a few competing members, as in nascent cancers or plasmids in a single cell. I want to get a better sense of when the exponential scaling is relevant, especially since for those systems with almost complete niche overlap the exponential scaling is slow. To this end I compare the expected mean fixation time with that of the Moran model. The ansatz $e^{h(a)}K^{g(a)}e^{f(a)K}$ is fit to the data and then used to estimate the fixation time at a variety of parameter values. This time is compared to the fixation time of a Moran model with the same carrying capacity. In figure 2.8 the shaded region represents those parameter combinations for which the estimated fixation is faster than the corresponding Moran model. For example, a carrying capacity of about 35 organisms is sufficient to allow for effective co-existence of two species which are not more than 99% identical in their niches. This is a small population size in most biological contexts. Even for systems with a smaller carrying capacity, unless the two species are more similar they are expected to co-exist for long times before fixation.
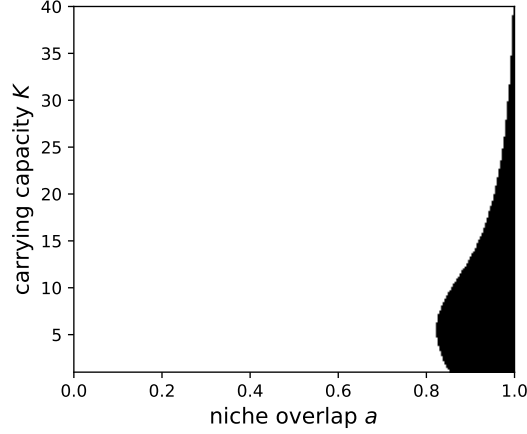
Figure 2.8: *Parameter space in which fixation is fast.* The white area shows where two species are expected to effectively co-exist, while the black shading identifies the regime where fixation is faster than a similar Moran model. Fixation is estimated by extrapolating the ansatz parameter fits to the a, K parameter space.

2.11 Analysis of the Fokker-Planck approximation in this context

The most common approximation to the master equation is Fokker-Planck, which assumes the state space is continuous. I attempt its use here to get an analytic estimate of the dependence of fixation time on K and a. We shall see that its utility is only marginal, though with some further approximations and an application of Kramers' theory I get my desired estimate.

The Fokker-Planck approximation to the coupled logistic system studied herein takes its traditional form [119]:

$$\frac{dP}{dt} = -\partial_1[(b_1 - d_1)P] - \partial_2[(b_2 - d_2)P] + \frac{1}{2}\partial_1^2[(b_1 + d_1)P] + \frac{1}{2}\partial_2^2[(b_2 + d_2)P]$$
$$= -\sum_i \partial_i F_i P + \sum_{i,j} \partial_i \partial_j D_{ij} P \tag{2.24}$$

where F is the force vector and D is the diffusion matrix (in this case diagonal). Here, under symmetric conditions and nondimensionalization by r, $F_1 = \frac{x_1}{K}(K - x_1 - ax_2)$ and $D_{11} = \frac{x_1}{K}(K + x_1 + ax_2)$, with similar terms for species 2.

In general, Equation (2.24) cannot be reduced to diffusion in a potential $U(\vec{x})$ with an equilibrium distribution function $P(\vec{x}) \sim \exp(U(\vec{x}))$. The condition of zero flux at equilibrium, $J_i = F_i P - 1/2 \sum_j \partial_j D_{ij} P = 0$, would require [57]

$$\partial_i \log P = \sum_k (D^{-1})_{ik}\left(2F_k - \sum_j \partial_j D_{kj}\right) \equiv -\partial_i U,$$

However, for consistency it also requires $\partial_j(-\partial_i U) = \partial_i(-\partial_j U)$ [57]. It is easy to show that this is not upheld for the two directions unless $a_{12} = a_{21} = 0$ and the system can be decomposed into two one-dimensional logistic systems.

Instead, I define the pseudo-potential as:

$$U(x_1, x_2) \equiv -\ln\left[P_{ss}(x_1, x_2)\right]. \tag{2.25}$$

where $P_{ss}(x_1, x_2)$ is a quasi-stationary probability distribution function [161]. I calculate $P_{ss}(x_1, x_2)$ in the approximation to the Fokker-Planck Equation (2.24) linearized about the deterministic co-existence fixed point. The linearized equation is

$$\partial_t P = -\sum_{i,j} A_{ij}\partial_i(x_j - x_j^*)P + \frac{1}{2}\sum_{i,j} B_{ij}\partial_i\partial_j(x_i - x_i^*)(x_j - x_j^*)P \tag{2.26}$$

where $A_{ij} = \partial_j F_i|_{\vec{x}=\vec{x}^*}$ and $B_{ij} = D_{ij}|_{\vec{x}=\vec{x}^*}$. The quasi-equilibrium solution to Equation (2.26) is $P_{ss} = \frac{1}{2\pi}\frac{1}{|C|^{1/2}}\exp[-(\vec{x}-\vec{x}^*)^T C^{-1}(\vec{x}-\vec{x}^*)/2]$, a Gaussian centered on the co-existence point and with a variance given by the covariance matrix $C = B \cdot A^{-1}/2$ in the symmetric case $a_{12} = a_{21} = a$, $K_1 = K_2 = K$ [148]. In this case the diagonal term of C is $\frac{1}{1-a^2}K$ and gives the variance of a species about its mean value. The off-diagonal, which corresponds to the covariance between the two species, is $-\frac{a}{1-a^2}K$. Thus the Pearson correlation coefficient between the two species is $-a$. That is, they are maximally anti-correlated when $a = 1$, lying along the line $x_1 + x_2 = K$ - the Moran line.

For the initial condition at the co-existence fixed point and assuming that the system escapes towards fixation once it reaches one of the axial fixed points $(0, K)$ or $(K, 0)$, from Equation (2.25) the well depth is proportional to carrying capacity K, being

$$\Delta U = \frac{(1-a)}{2(1+a)}K. \tag{2.27}$$

In a Kramers' type approximation, the escape time from the pseudo-potential well scales as $\sim \exp(\Delta U)$ [64], reproducing the exponential scaling of the extinction time with K, observed numerically. Moreover, the Fokker-Planck approximation also shows that the exponential scaling disappears as niche overlap a approaches unity, in accord with the numerical results above. The correlation between the two species goes to negative one in this parameter limit, such that they are entirely anti-correlated. Whereas the well has a single lowest point at the co-existence fixed point for partial niche overlap, at $a = 1$ the potential shows a trough of constant depth going between the two axial fixed points. This is the Moran line, along which diffusion is unbiased; diffusion away from the Moran line is restored, as the system is drawn toward the bottom of the trough. Because everywhere along the Moran line is equally likely, the probability cannot be normalized, and the linearization approximation fails. This is to be expected, as it is an expansion about a fixed point, but the fixed point is replaced by the Moran line in the Moran limit of $a = 1$.

2.12   Breaking the parameter symmetries

I have addressed the symmetric case of $K_1 = K_2 \equiv K$ and $a_{12} = a_{21} \equiv a$. The result of exponential scaling of the fixation time except when the Moran line exists is true even when some symmetries are broken. However, the evidence is not as clear as in the symmetric case.

Figure 2.9 shows the dependence of the fixation time on the niche overlap a_{21} while keeping $a_{12} = 0.5$ for $K_1 = K_2 \equiv K$; using the similar ansatz, I apply the same $\tau = e^{h(a_{21})}K^{g(a_{21})}e^{f(a_{21})K}$. As the niche
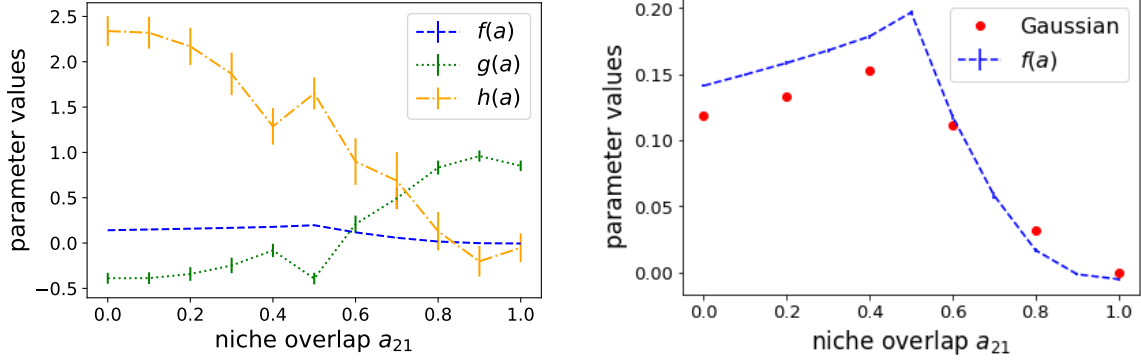
Figure 2.9: *Breaking the symmetry in a*. As in Figure 2 in the main text, lines come from fitting the ansatz to generated data. The exponential dependence is non-zero except at $a_{21} = 1$, at which point the "co-existence" fixed point is coincident with the fixed point on the x-axis. The right panel compares this ansatz fit with the Gaussian well depth at the same parameter values.

overlap a_{21} changes from 0 to 1, the location of the co-existence fixed point shifts from $(K/2, K/4)$ to $(K, 0)$. Accordingly, the fixation time starting from the fixed point maintains its exponential scaling with carrying capacity up until $a_{21} = 1$, where the fixed time is equal to zero, as reflected in the shape of the of $h(a_{21})$. Notably, in this asymmetric case the exponential scaling function $f(a_{21})$ is much weaker compared to the symmetric case, partially because the fixed point is located closer to an axis that in the symmetric case even at $a_{21} = 0$. The right panel of Figure 2.9 shows the comparison of the results of the ansatz fit with the estimates of the exponential part of the fixation time using Kramers'/Fokker-Planck pseudo-potential described in the previous section that explains the observed trends of $f(a_{21})$.
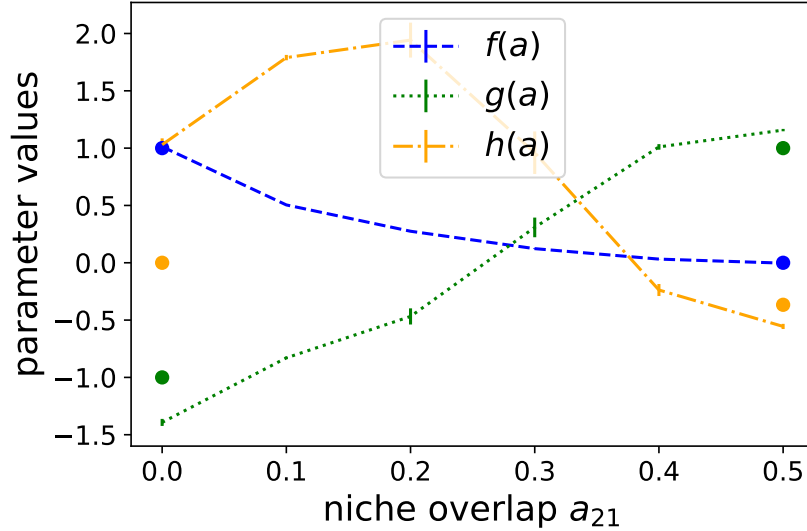
Figure 2.10: *Breaking the symmetry in K*. As in Figure 2 in the main text, lines come from fitting the ansatz to generated data. The exponential dependence is non-zero except at the appearance of the Moran line at $a_{21} = 1/2$. The extreme points are the expected asymptotic values.

Next let us consider breaking the symmetry such that the Moran line can still be recovered. The carrying capacity symmetry is broken, such that $K_2 = 2K_1$. The two species are still independent when

$a_{12} = a_{21} = 0$, but in this case the Moran line exists when $a_{12} = 2$ and $a_{21} = 1/2$. Figure 2.10 shows the results when the symmetry is broken both in the carrying capacity and the niche overlap. It shows the change in the fixation time as a function of the niche overlap a_{21} for $K_2 = 2K_1 \equiv K$ while the niche overlaps change along the line where $a_{12} = 4a_{21}$, starting from the independent case $a_{12} = a_{21} = 0$ to $a_{12} = 2$ and $a_{21} = 1/2$ where the system reaches its corresponding Moran line. The observed behaviour is very similar to that shown in the symmetric case, with the exponential dependence transitioning smoothly to zero at the Moran line.

I uphold the conclusion that at the Moran line will fixation be fast; when the system parameters are even slightly off those niche overlap values which balance the carrying capacities and allow for the Moran line to exist, the fixation is exponential in the carrying capacity, to the point that the two species effectively co-exist. I include the caveat that fixation will also be fast when the co-existence fixed point is close to one axis, as evinced with the broken niche overlap symmetry above.

2.13   Route to Fixation
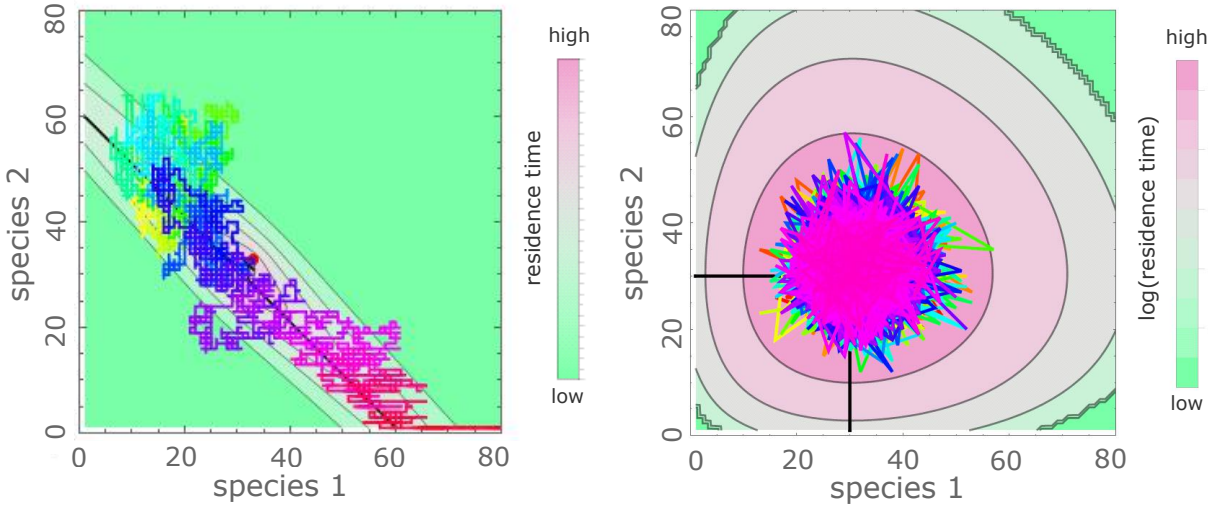
Figure 2.11: *The system samples multiple trajectories on its way to fixation.* Contour plot shows the average residency times at different population states of the system, with pink indicating longer residence time, deep green indicating rarely visited states. The colored line is a sample trajectory the system undergoes before fixation; color coding corresponds to the elapsed time with orange at early times, purple at the intermediate times and red at late stages of the trajectory. The red dot shows the deterministic co-existence point. See text for more details. *Left*: Complete niche overlap limit, $a = 1$, for $K = 64$. *Right*: Independent limit with $a = 0$ and $K = 32$.

To gain deeper insight into the fixation dynamics, in this section I calculate the residency times in each state during the fixation process [63], given by Equation (2.21), reproduced here:

$$\langle t(s^0) \rangle_s = \int_0^\infty dt P(s, t | s^0, 0) = \hat{M}^{-1}_{s,s^0}.$$

This gives the mean total time the system resides in state s, given that it starts in state s^0. As with the fixation times, the dependence on the initial state is weak, so long as that initial state is not near an axis. Whereas I previously analyzed the fixation time scaling, the residency times themselves proffer some

insight into the system. The results are shown as a contour plot in Figure 2.11, where pink corresponds to the high occupancy sites and green to the rarely visited ones, for two different niche overlaps, one at and the other far from the Moran limit. The set of states lying along the steepest descent lines of the contour plot, shown as the black line, can be thought of as a "typical" trajectory [56, 82, 104]. However, even for two species close to complete niche overlap the system trajectory visits many states far from this line. This departure is even greater for weakly competing species, where the system covers large areas around the fixed point before the rare fluctuation that leads to fixation occurs [60]. These deviations from a "typical" trajectory are related to the inaccuracy of the WKB approximation in calculating the scaling of the pre-exponential factor [14, 60, 92]; see also the previous chapter.

This occupancy landscape can be qualitatively thought of as an effective Lyapunov function/effective potential of the system dynamics [161], although the LV system does not possess a true Lyapunov function - an issue that also arises in the Fokker-Planck approximation [34, 161]. One way to deal with this issue is the linearization in the Fokker-Planck section above (section 2.11). This allows one to easily solve the Fokker-Planck equation in any system with an attractive deterministic fixed point. More pertinently, linearizing the Fokker-Planck equation, as in Equation 2.26 described above, allows one to get an estimate of the depth of the pseudo-potential: $\frac{(1-a)}{2(1+a)}K$. This provides an intuitive underpinning for the general exponential scaling in the incomplete niche overlap regime: the fixation process can be thought of as the Kramers'-type escape from a pseudo-potential well [?]. The Kramers result is dominated by $\tau \simeq \exp(\text{well depth})$, corresponding to the dominant scaling $\tau \simeq \exp(f(a)K)$. As a increases and the species interact more strongly, the potential well becomes less steep, resulting in weaker exponential scaling. In the complete niche overlap limit, the pseudo-potential develops a soft direction along the Moran line that enables relatively fast escape towards fixation. This is what is seen in the residence time graph; in the Moran limit, the states along the line connecting the co-existence fixed point to the two axial fixed points are visited much more frequently than those off it. Outside of the Moran limit, it is rather the states in a cloud around the co-existence point that enjoy long residence times.

In linearizing the FP equation I also arrived at the Pearson correlation coefficient between the two species: $-a$. They are anti-correlated, and this anti-correlation becomes complete as niche overlap approaches one. In state space this corresponds to the system lying on the Moran line. Thus we expect the pseudo-potential to become less steep as a increases, eventually developing a level trough along the Moran line that enables relatively fast escape toward fixation. Though I am unaware of any direct connection, this disappearance is also mimicked in the deterministic co-existence point eigenvalue associated with the $(1, -1)$ direction, which goes to zero as niche overlap goes to one, as $-\frac{1-a}{1+a}$.

2.14 Discussion

Maintenance of species biodiversity in many biological communities is still incompletely understood. The classical idea of competitive exclusion postulates that ultimately only one species should exist in an ecological niche, excluding all others. Although the notion of an ecological niche has eluded precise definition, it is commonly related to the limiting factors that constrain or affect the population growth and death. In the simplest case, one factor corresponds to one niche, which supports one species, although a combination of factors may also serve as a niche, as discussed above. The competitive exclusion picture has encountered long-standing challenges as exemplified by the classical "paradox of the plankton" [32, 75] in which many species of plankton seem to co-habit the same niche; in many

other ecosystems the biodiversity is also higher than appears to be possible from the apparent number of niches [32, 71, 98, 105, 142].

Competitive exclusion-like phenomena can appear in a number of popular mathematical models, for instance in the competition regime of the deterministic Lotka-Volterra model, whose extensive use as a toy model enables a mathematical definition of the niche overlap between competing populations [2, 33, 101, 140]. Another classical paradigm of fixation within an ecological niche is the Moran model (and the closely related Fisher-Wright, Kimura, and Hubbell models) that underlies a number of modern neutral theories of biodiversity [3, 71, 87, 97, 106, 111?]. Unlike the deterministic models, in the Moran model fixation does not rely on deterministic competition for space and limiting factors but arises from the stochastic demographic noise. Recently, the connection between deterministic models of the LV type and stochastic models of the Moran type has accrued renewed interest because of new focus on the stochastic dynamics of the microbiome, immune system, and cancer progression [8, 11, 14, 34, 38, 97, 131, 150].

Much of the recent studies of these systems employed various approximations, such as the Fokker-Planck approximation [34, 38, 48, 52, 97], WKB approximation [56, 82] or game theoretic [8] approach. The results of these approximations typically differ from the exact solution of the master equation, especially for small population sizes [14, 16, 49, 82, 122], as was discussed in more detail in the previous chapter. In this chapter, I have interrogated stochastic dynamics of a system of two competing species using a numerically arbitrarily accurate method based on the first passage formalism in the master equation description. The algorithmic complexity of this method scales algebraically with the population size rather than with the exponential scaling of the fixation time, (as is the case with the Gillespie algorithm [58]) enabling us to capture both the exponential behaviour and the algebraic prefactors in the fixation/extinction times for both small and large population sizes. This accuracy is needed in order to observe the transition from slow, exponentially dominated processes to the algebraically fast fixation of the Moran limit.

Stochastic fluctuations allow the system to escape from the deterministic co-existence fixed point towards fixation. If the escape time is exponential in the (typically large) system size, in practice it implies effective co-existence of the two species around their deterministic co-existence point. If the time is algebraic in K, as in degenerate niche overlap case (closely related to the classical Moran model), the system may fixate on biological timescales [85, 111]. For those biological systems with small characteristic population sizes, exponential scaling does not dominate the fixation time; power law and prefactor become more relevant. Figure 2.8 shows that a niche overlap as low as 0.8, for a carrying capacity around 6, has rapid fixation, more rapid than a corresponding Moran model. The transition between the exponential scaling of effective co-existence time to the rapid stochastic fixation in the Moran limit is governed by the niche overlap parameter, which for example can be derived in terms of the dynamics and interactions of the species and their secreted growth and death factors.

While I find that the fixation time is exponential in the system size unless the two species occupy exactly the same niche, the numerical factor in the exponential is highly sensitive to the value of the niche overlap, and smoothly decays to zero in the complete niche overlap case. These results can be understood by noticing that the escape from a deterministically stable co-existence fixed point can be likened to Kramers' escape from a pseudo-potential well [20, 48, 64, 122], where the mean transition time grows exponentially with well depth [122]. Approximating the steady state probability with a Gaussian shows that this well depth is proportional to $K(1 - a)$ and disappears when $a = 1$. With complete niche overlap the system develops a "soft" marginally stable direction along the Moran line

that enables algebraically fast escape towards fixation [34, 48]. Similar to the exponential term, the exponent of the algebraic prefactor is also a function of the niche overlap, and smoothly varies from -1 in the independent regime of non-overlapping niches to $+1$ in the Moran limit.

Niche overlap between two species, the similarity in how they interact with their shared environment, is of critical importance in determining whether they will co-exist. For typically large biological populations, effective co-existence occurs when escape time grows exponentially with the carrying capacity, which is the case for even slightly mismatched niches. Any niche mismatch leads to species which tend to exist for long times near their respective carrying capacities; in effect, niche models are apt even for large niche overlap. Only when niche overlap is complete will fixation be relatively rapid, algebraic in K. This has important implications for understanding the long term population diversity in many systems, such as human microbiota in health and disease [37, 88, 123], industrial microbiota used in fermented products [155], and evolutionary phylogeny inference algorithms [21, 134]. My results show that the generalized Lotka-Volterra model serves well as an extension to neutral models for problems such as maintenance of drug resistance plasmids in bacteria [59], strain survival in cancer progression [11], or the generation of coalescent or phylogenetic trees [87, 134, 136]. The theoretical results can also be tested and extended based on experiments in more controlled environments, such as the gut microbiome of a *c. elegans* [150], or in microfluidic devices [73].

Chapter 3

Ch3-AsymmetricLogistic

3.1 pre-intro

Half of this research has been submitted, in conjunction with the previous chapter, to be published in Journal of the Royal Society: Interfaces.

Note also that I talk about foreign invading immigrants in this chapter. This is not meant to be related to human immigrants into a country (which I view favourably).

3.2 Introduction

The previous chapter regarded an ecosystem with two competing species, and asked questions about the mean time until one of the species goes extinct and the other fixates in the system. In this chapter I aim to look at the reverse problem; starting with a system with one species, what is the probability and timescale that a second one will enter, given some overlap between the niches of the extant and immigrating species. First I would like to motivate the problem and discuss where a new species entering a system might come from.

Many of the historical giants I included in the introductory chapter have considered the problem of the arrival of a new strain or species in one way or another. Kimura and Crow analyzed the evolution of the probability of a single allele in a population, one that arose via genetic mutation or immigration , to find its fixation probability [43, 85?]. MacArthur and Wilson considered islands receiving an influx of immigrants from a neighbouring mainland to find the total number of distinct species on an island [101?]. Hubbell builds off of MacArthur and Wilson to predict the abundance distribution of species in a system balanced between influx of new species and extinction of extant ones [71]. New species can arise from speciation or from a larger reservoir. Bearing these historical precedents in mind, I do not distinguish where a new strain or species might enter from in my modelling below; mutation, speciation, and immigration are all viable. What is important to my research is that a distinctive second species is attempting to invade an already occupied system. Whether the invader is under selection [84] or the system is neutral [71?], the literature regards cases where the system is constrained to the Moran line, to constant population size. What has *not* been done is to look at an invasion attempt into an established niche when the invader has partial niche overlap with the established species.

By using the two species Lotka-Volterra model from the previous chapter I can study invasion in the

neutral case where the system is allowed to fluctuate off the Moran line, or even when the two species should happily coexist in the deterministic limit, *i.e.* with partial niche overlap and a single stable coexistence fixed point. I do this by continuing to use the truncated transition matrix inverse to solve the backward master equation for arbitrary accuracy. Furthermore, the literature typically argues that invasion attempts are sufficiently rare that when an invader arrives it will either successfully invade or die off before another member of the same strain invades. Indeed, this is one of my assumptions for the Lotka-Volterra model. But this need not be the case, depending on the immigration rate (see, for example, [61]); I also analyze the Moran model with an immigration term, which allows for repeated concurrent invaders of the same species. To this end immigration is more appropriate than mutation for the introduction of invaders, since it is unlikely to have the same mutation recurring independently, unless there is a common mutation pathway or if we categorize all equivalent mutants into one category of invader. In this chapter I shall investigate how the success probability and mean times scale with niche overlap, carrying capacity, and immigration rate, and in so doing I shall uncover critical combinations of these parameters as they affect the scaling of the mean times and the shape of the steady state distribution. Having increased niche overlap leads to lesser chance of invasion and greater times. In the Moran model with immigration, the steady state distribution changes from unimodal to bimodal around when the inverse immigration rate of a strain is equal to the expected population of that strain in the system.

There are a few steps needed to get to these conclusions. Within the generalized Lotka-Volterra model there is some ambiguity in the definition of a successful invasion, which I will discuss before providing a definition. Similar to the invasion success probability, I shall find the mean times conditioned on the success or failure of an attempt. In the previous chapter because of the initial conditions each species was equally likely to go extinct first. In this case, it is possible (and indeed true) that an invasion attempt that is ultimately successful will take a long time, whereas one that is unsuccessful fails quickly. These are the conditional mean times, and their scaling with carrying capacity will be analyzed, since exponential scaling implies that the event effectively does not happen. In order to extend these results to the circumstance of repeated concurrent invaders in the Moran limit I analyze the Moran model with an immigration term. I find the steady state probability distribution analytically to allow for an investigation of the critical parameter combinations that change the concavity of the curve. Along with the probability distribution I find the mean time to fixation, both unconditioned and conditioned on whether the species first fixates or goes extinct in the system. The application is in neutral theories like that of Hubbell [71]; I find the qualitatively different regimes of the probability distribution, which can be extended to abundance distributions. Neutral theories of the maintenance of biodiversity argue that no species truly establishes itself, and biodiversity is maintained by transient species in the system. Calculation of the steady state number of species requires the time that these transient species exist in the system. My results hold both for a species in an ecosystem (hence its relevance to conservation biology, where biodiversity is a marker of ecosystem robustness) and a gene in a population (hence its use in calculating heterozygosity, which confers resilience to environmental changes). There are also more practical applications, like the susceptibility of a microbiome ecosystem like the gut or lungs to invasion, say from salmonella or pneumococcus.

3.3   Defining Invasion in the Two-Dimensional Lotka-Volterra Model

As before, I employ the symmetric generalized LV model with niche overlap a and carrying capacity K. I study the case where the system starts with $K - 1$ individuals of the established species and 1 invader. This initial condition corresponds to a birth of a mutant. An initial condition of K established organisms and the 1 invader gives similar results. The species' dynamics are described by the birth and death rates defined by Equations (2.10) from the previous chapter, which I reproduce here:

$$b_i/x_i = r_i$$
$$d_i/x_i = r_i \frac{x_i + a_{ij}x_j}{K_i}.$$

An invasion is unsuccessful if the invading species dies out before establishing itself in the system. There are many ways to define what it means for a species to be established, and I will outline one such definition below. Deterministically the system would grow to asymptotically approach the co-existence fixed point; deterministically, all invasion attempts are successful, and stochasticity is required for non-trivial invasion probabilities. In a stochastic system, the populations could very easily fluctuate *near* the fixed point without touching that exact point. This would overestimate the time to establishment, or even misrepresent a successful invasion as unsuccessful if the system gets near the fixed point without reaching it but then fixates. Indeed, there is even a chance the established species dies out before the system reaches the co-existence fixed point, which would be counted as an unsuccessful invasion. For these reasons a successful invasion should not be defined as the system arriving at the co-existence point. The same arguments hold for a defined region near the fixed point (for instance, within three birth or death events, or within a circle of radius ε): the region might by chance be avoided for a time even after the invader is more populous than the original species, which could even go extinct before the invader. Inspired by the observation that in the symmetric case, the co-existence fixed point has the same population of each species, I consider the invasion successful if the invader grows to be half of the total population without dying out first. So long as the invader population matches that of the established species, regardless what random fluctuations may have made that population to be, the invasion is a success. I denote the probability of a invader success as $\mathcal{P}$.

Along with the probability of a successful invasion attempt, I am interested in the timescales involved. As such, I will consider conditional mean times, conditioned on either success or failure of the invasion attempt. The mean time to a successful invasion is written as τ_s, and the mean time of a failed invasion attempt as τ_f. More generally, invasion probability and the successful and failed times starting from an arbitrary state s^0 are denoted as $\mathcal{P}^{s^0}$, $\tau_s^{s^0}$ and $\tau_f^{s^0}$, respectively.

Similar to Equation (2.22) in a previous chapter, the invasion probability can be obtained from [77, 119]

$$\mathcal{P}^{s^0} = -\sum_s \hat{M}_{s,s^0}^{-1} \alpha_s \tag{3.1}$$

and the times from

$$\Phi^{s^0} = -\sum_s \hat{M}_{s,s^0}^{-1} \mathcal{P}^s, \tag{3.2}$$

where α_s is the transition rate from a state s directly to extinction or invasion of the invader and

$\Phi^{s^0} = \tau^{s^0} \mathcal{P}^{s^0}$ is a product of the invasion or extinction time and probability. Similar equations describe τ_f [77, 119].
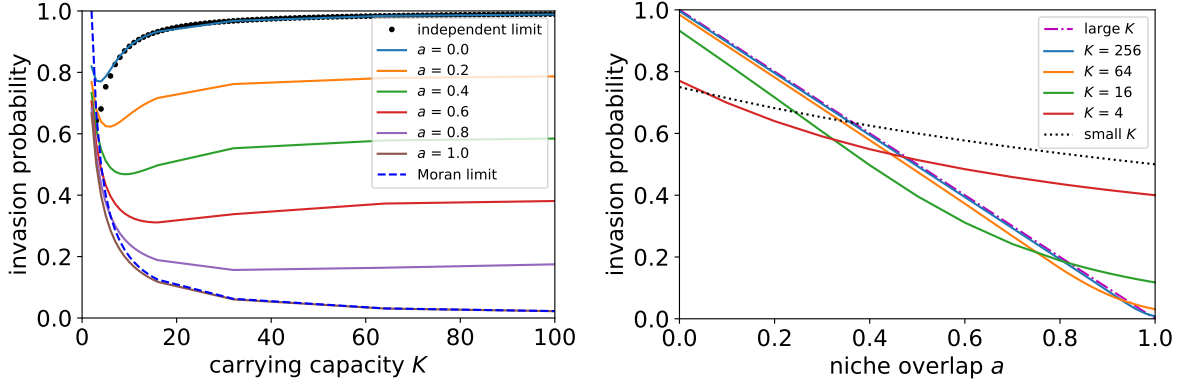
3.4   Results and some intuition behind them

Figure 3.1: *Probability of a successful invasion. Left:* Solid lines show the numerical results, from $a = 0$ at the top to $a = 1$ at the bottom. The black dotted line is the expected analytical solution in the independent limit. The blue dashed line is the prediction of the Moran model in the complete niche overlap case. *Right:* The solid blue line shows the results for small carrying capacity ($K = 4$), and matches well with the black dotted line $\frac{b_{mut}}{b_{mut}+d_{mut}}$. Successive lines are at larger system size, and approach the dash-dot purple line of $1 - d_{mut}/b_{mut} \approx 1 - a$.

Figure 3.1 shows the calculated invasion probabilities as a function of the carrying capacity K and of the niche overlap a between the invader and the established species. In the complete niche overlap limit, $a = 1$, the dependence of the invasion probability on the carrying capacity K closely follows the results of the classical Moran model, $\mathcal{P}^{s^0} = 2/K$ [111], shown in the blue dotted line in the left panel, and tends to zero as K increases. In the other limit, $a = 0$, the problem is well approximated by the one-species stochastic logistic model starting with one individual and evolving to either 0 or K individuals; the exact result in this limit is shown in black dotted line, referred to as the independent limit [119]. In the independent limit, $a = 0$, the invasion probability asymptotically approaches 1 for large K, reflecting the fact that the system is deterministically drawn towards the deterministic stable fixed point with equal numbers of both species. As K gets large, fluctuations are minimal and the system becomes more deterministic. Interestingly, the invasion probability is a non-monotonic function of K and exhibits a minimum at an intermediate/low carrying capacity, which might be relevant for some biological systems, such as in early cancer development [11] or plasmid exchange in bacteria [59].

For the intermediate values of the niche overlap, $0 < a < 1$, the invasion probability is a monotonically decreasing function of a, as shown in the right panel of Figure 3.1. For large K, the outcome of the invasion is typically determined after only a few steps: since the system is drawn deterministically to the mixed fixed point, the invasion is almost certain once the invader has reproduced several times. At early times, the invader birth and death rates (2.10) are roughly constant, and the invasion failure can be approximated by the extinction probability of a birth-death process with constant death d_{mut} and birth b_{mut} rates. The invasion probability is then $\mathcal{P} = 1 - d_{mut}/b_{mut} \approx 1 - a$. This heuristic estimate is in excellent agreement with the numerical predictions, shown in the right panel of Figure 3.1 as a purple

dashed and the blue lines respectively. Similarly, for small K either invasion or extinction typically occurs after only a small number of steps. The invasion probability in this limit is dominated by the probability that the lone mutant reproduces before it dies, namely $\frac{b_{mut}}{b_{mut}+d_{mut}} = \frac{K}{K(1+a)+1-a}$, as shown in black dotted line in the right panel of Figure 3.1.
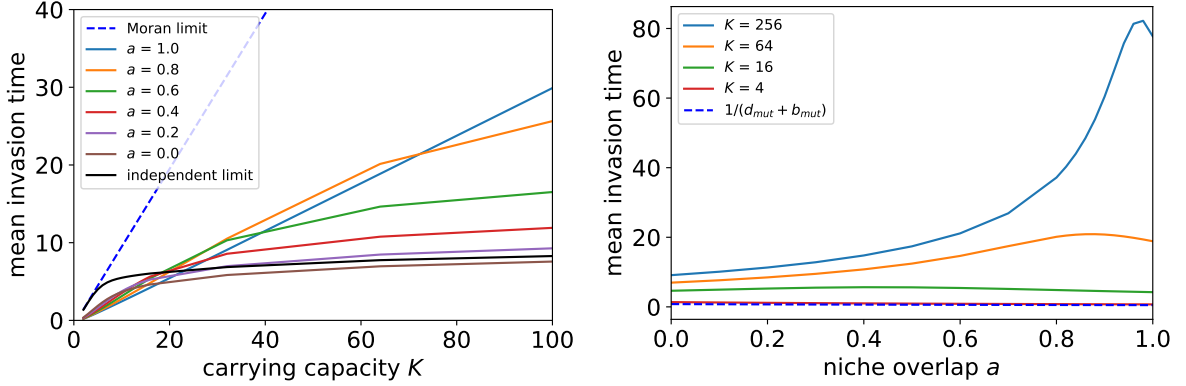

Figure 3.2: *Mean time of a successful invasion. Left:* Solid lines are the numerical results, from $a = 0$ at the bottom to $a = 1$ at top. The blue dashed line shows for comparison the predictions of the Moran model in the complete niche overlap limit, $a = 1$; see text. The black line correspond to the solution of an independent stochastic logistic species, $a = 0$. *Right:* The solid red line shows the results for small carrying capacity ($K = 4$), and successive lines are at larger system size, up to $K = 256$. The dashed blue line is $1/(b_{mut} + d_{mut})$ and matches with small K.

Figure 3.2 shows the dependence of the mean time to successful invasion, τ_s, on K and a. Increasing K can have potentially contradictory effects on the invasion time, as it increases the number of births before a successful invasion on the one hand, while increasing the steepness of the potential landscape and therefore the bias towards invasion on the other. Nevertheless, the invasion time is a monotonically increasing function of K for all values of a. In the complete niche overlap limit $a = 1$ the invasion time scales linearly with the carrying capacity K, as expected from the predictions of the Moran model, $\tau_s = \Delta t K^2 (K - 1) \ln\left(\frac{K}{K-1}\right)$ with $\Delta t \simeq 1/K$, as explained above. The quantitative discrepancy arises from the breakdown of the $\Delta t \simeq 1/K$ approximation off of the Moran line. For all values $0 \leq a < 1$ the invasion time scales sub-linearly with the carrying capacity, indicating that successful invasions occur relatively quickly, even when close to complete niche overlap, where the invading mutant strongly competes against the stable species. In the $a = 0$ limit of non-interacting species, the invading mutant follows the dynamics of a single logistic system with the carrying capacity K, resulting in the invasion time that grows approximately logarithmically with the system size, as shown in the left panels of Figure 3.2 as a black line. This result is well-known in the literature, and is stated without reference for instance by Lande [92]. It is easy to see: by writing $\tau_s = \int dt = \int_{x_o}^{x_f} dx \frac{1}{\dot{x}}$ for initial state $x_0 = 1$ and final state $x_f = (1 - \epsilon)K$ with small ϵ and large K we get

$$\tau_s = \frac{1}{r} \int_{x_o}^{x_f} dx \frac{K}{x(K-x)} = \frac{1}{r} \int_{x_o}^{x_f} dx \left(\frac{1}{x} - \frac{1}{K-x}\right) = \frac{1}{r} \ln\left[\frac{x}{K-x}\right] \Big|_{x_o}^{x_f} = \frac{1}{r} \ln\left[\frac{x_f(K-x_o)}{x_o(K-x_f)}\right]$$
$$\approx \frac{1}{r} \ln\left[\frac{(1-\epsilon)K}{\epsilon}\right] \approx \frac{1}{r} \left(\ln\left[K\right] - \ln\left[\epsilon\right]\right)$$

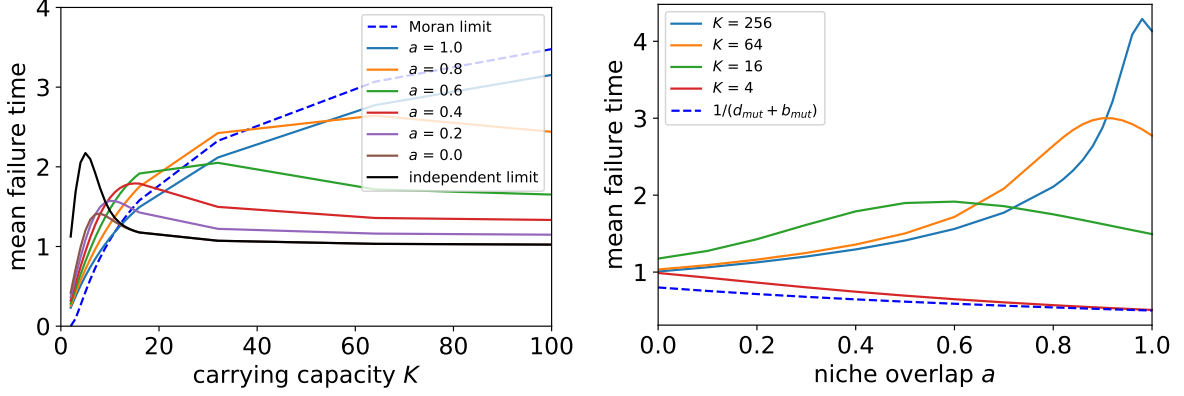and so expect the invasion time to grow logarithmically with carrying capacity.

Figure 3.3: *Mean time of a failed invasion. Left:* Solid lines are the numerical results, from $a = 0$ mostly being fastest to $a = 1$ being slowest, for large K. The blue dashed line is the analytical approximation of the Moran model result, and black is a 1D stochastic logistic system, which overestimates the time at small K but then converges to the same limiting value. *Right:* The solid red line shows the results for small carrying capacity ($K = 4$), and successive lines are at larger system size, up to $K = 256$. The dashed blue line is $1/(b_{mut} + d_{mut})$ and matches with small K.

Unlike the mean times conditioned on success, the failed invasion time, shown in Figure 3.3, is non-monotonic in K. The analytical approximations of the Moran model and the of two independent 1D stochastic logistic systems recover the qualitative dependence of the failed invasion time on K at high and low niche overlap, respectively. All failed invasion times are fast, with the greatest scaling being that of the Moran limit. For $a < 1$ these failed invasion attempts appear to approach a constant timescale at large K.

The dependence of the time of an attempted invasion (both for successful and failed ones) on the niche overlap a is different for small and large K, as shown in the right panels of Figures 3.2 and 3.3. For small K both τ_s and τ_f are monotonically decreasing functions of a, with the Moran limit having the shortest conditional times. In this regime, the extinction or fixation already occurs after just a few steps, and its timescale is determined by the slowest steps, namely the mutant birth and death. Thus $\tau \approx \frac{1}{b_{mut} + d_{mut}} = \frac{K}{K + 1 + a(K-1)}$, as shown in the figures as the dashed blue line. By contrast, at large K, the invasion time is a non-monotonic function of the niche overlap, increasing at small a and decreasing at large a. This behavior stems from the conflicting effect of the increase in niche overlap: on the one hand, increasing a brings the fixed point closer to the initial condition of one invader, suggesting a shorter timescale; on the other hand, it also makes the two species more similar, increasing the competition that hinders the invasion.

3.5 Discussion of one attempted invasion

Unlike the fixation times of the previous chapter, invasions into the system do not show exponential scaling in any limit. Indeed, all scaling with K is sublinear except in the complete niche overlap limit for successful invasion times. The timescale of a successful invasion varies between linear and logarithmic in the system size. The mean time of an unsuccessful invasion is even faster than logarithmic, and for large K it becomes independent of K. Curiously, these failed invasion attempts are unimodal, at intermediate carrying capacity and niche overlap values. As for the probabilities, the likelihood of a

failed invasion attempt grows linearly with niche overlap, for sufficiently large K. For complete niche overlap the invasion probability goes asymptotically to zero, but it is low even for partially mismatched niches.

High niche overlap makes invasion difficult due to strong competition between the species. In this regime, the times of the failed invasions become important because they set the timescales for transient species diversity. If the influx of invaders is slower than the mean time of their failed invasion attempts, most of the time the system will contain only one settled species, with rare "blips" corresponding to the appearance and quick extinction of the invader. On the other hand, if individual invaders arrive faster than the typical times of extinction of the previous invasion attempt, the new system will exhibit transient co-existence between the settled species and multiple invader strains, determined by the balance of the mean failure time and the rate of invasion [32, 47, 71]. Full discussion of diversity in this regime is beyond of the scope of the present work. The weaker dependence of the invasion times on the population size and the niche overlap, as compared to the escape times of a stably co-existing system to fixation, imply that the transient co-existence is expected to be much less sensitive to the niche overlap and the population size than the steady state co-existence. Curiously, both niche overlap and the population size can have contradictory effects on the invasion times (as discussed in section III) resulting in a non-monotonic dependence of the times of both successful and failed invasions on these parameters.

For species with low niche overlap, the probability of invasion is likely, and for large K decreases monotonically as $1 - a$ with the increase in niche overlap, independent of the population size K. The mean time of successful invasion is relatively fast in all regimes, and scales linearly or sublinearly with the system size K and is typically increasing with the niche overlap a.

The fixation times of two co-existing species, discussed in the previous chapter, determine the timescales over which the stability of the mixed populations can be destroyed by stochastic fluctuations. Similarly, the times of successful and failed invasions set the timescales of the expected transient co-existence in the case of an influx of invaders, arising from mutation, speciation, or immigration. Our results provide a timescale to which the rate of immigration or mutation can be compared. If the influx of invaders is slower than the mean time of their failed invasion attempts, each attempt is independent and has the invasion probability we have calculated. In the extreme case of this, that is, if the time between invaders is even longer than the fixation times calculated in the previous chapter, then serial monocultures are expected. If the rate in is greater than the mean failure time, the system will diversify. The balance between mutation or immigration coming into the system and these invaders failing to establish themselves determines how diverse a system will be. With different strains of invaders arising faster than the time it takes to suppress the previous invasion attempt, the new strains interact with one another in ways beyond the scope of this thesis, leading to greater biodiversity. We can get an idea of what it would be like, having a new immigrant come in before the previous invasion attempt is over, by considering a Moran model with immigration. This would correspond to the complete niche overlap limit, such that the population size is roughly constrained to the Moran line. In the following sections I solve steady state and dynamical properties of the Moran model with immigration. Compared to the Moran limit of the Lotka-Volterra system it is easier to treat, has well established literature results against which to compare, and offers analytic solutions.

3.6   Moran model in more detail

The previous results in this chapter have related to a single organism attempting to invade a system wherein another species is already established. The number of invader progeny fluctuates and ultimately it either dies out or occupies half of the total population. But recall the island model of MacArthur and Levins, in which a mainland system, which is large, is considered to be static, while a much smaller island system's dynamics are regarded, occasionally including immigration from the mainland. If the immigration rate is large then the invaders will receive reinforcements from the mainland in their attempt to establish themselves on the island system of interest. The easiest way to model this is a Moran model with immigration. In this section I will review the Moran model and some quantities that can be calculated from it. In the following section I calculate these quantities for a Moran model with immigration, and link said model to the results of a single invader analyzed above.

As a reminder, the Moran model [111] is a classic urn model used in population dynamics in a variety of ways. Its most prominent uses are in coalescent theory [21, 51] and neutral theory [18, 71?], describing how the relative proportion of genes in a gene pool might change over time. In fact it can describe any system where individuals of different species/strains undergo strong but unselective competition in some closed or finite ecosystem [36]: applications include cancer progression [11], evolutionary game theory [8, 67?], competition between species [38? ?], population dynamics with evolution [147], and linguistics [21].

In the classic Moran model, each iteration or time step involves a birth and a death event. Each organism is equally likely to be chosen (for either birth or death), hence a species is chosen according to its frequency, $f = n/N$, where N is the total population and n is the number of organisms of that species. We focus on one species of population n, which will be referred to as the focal species. Note that $N - n$ represents the remainder of the population, and need not all be the same species, so long as they are not the focal species. The focal species increases in the population if one of its members gives birth (with probability f) while a member of a different species dies (with probability $1 - f$); that is, in time step Δt the probability of focal species increase is $b(n) = f(1 - f)$. Similarly, decrease in the focal species comes from a birth from outside the focal group and a death from within, such that the probability of decrease is $d(n) = (1 - f)f$. By commutativity of multiplication, increase and decrease of the focal species are equally likely, with

$$b(n) = d(n) = n(N - n)/N^2. \tag{3.3}$$

Each time step, the chance that nothing happens is $1 - (b(n) + d(n)) = f^2 + (1 - f)^2$.

Note that, unlike in previous chapters where I used b and d as rates, here these are not rates, rather they are the probability of an increase or decrease of the focal species in one time step. I use the same notation not to be confusing but to hint at an approximation I employ in the following sections. Taking Δt to be infinitesimal, $b(n)\Delta t$ and $d(n)\Delta t$ serve as probabilities of birth and death of the focal species during this small time interval. This creates a continuous time analogue to the Moran model, with b and d serving as rates. The timescale is now in units of Δt, which is only relevant if one were to compare with other models, which I do not. With this approximation I can employ the formulae explored in chapter 1 for quantities like quasi-stationary probability distribution and mean time to extinction.

For reference, I include the mean and variance as a function of time [111?], so that I may later compare with the immigration case. If the system starts with n_0 individuals of the focal species, then

on average there should be n_0 individuals in the next time step as well. Therefore the mean population as a function of time is $\langle n \rangle(t) = n_0$. Since the extremes of $n = 0$ and $n = N$ are absorbing, The ultimate fate of the system is in one of these two states, despite the mean being constant. The variance starts at zero for this delta function initial condition. After k time steps the variance is

$$V_k = n_0(N - n_0)\big(1 - (1 - 2/N^2)^k\big).$$

For finite N the variance goes to $N^2 f_0(1 - f_0)$ at long times. This is easy to intuit: there is probability f_0 that the system ended in $n = N$, and probability $(1 - f_0)$ of ending at $n = 0$, since at long times the system has fixated at one end or the other. Notice that as $N \to \infty$ the variance, a measure of the fluctuations, goes to zero, and the system becomes deterministic, as any change of $\pm 1/N$ in the frequency of the focal species becomes meaningless.

The mean and variance characterize the distribution of outcomes that could occur when running an ensemble of identical trials of the same system. This is the the ensemble average denoted by $\langle \cdot \rangle$. Any individual trajectory, any individual realization, will take its own course, independent of any others, and after fluctuations will ultimately end up with either the focal species dying (extinction) or all others dying (fixation). Both of these cases are absorbing states, so once the system reaches either it will never change. Since a species is equally likely to increase or decrease each time step, the model is akin to an unbiased random walk [57], and therefore the probability of extinction occurring before fixation is just

$$E(n) = 1 - n/N = 1 - f. \tag{3.4}$$

The first passage time, however, does not match a random walk, as there is a probability of no change in a time step, and this probability varies with f.

The unconditioned first passage time can be found using the techniques outlined in chapter 1. As a reminder, I define the unconditioned first passage time $\tau(n)$ as the time the system takes, starting from n organisms of the focal species, to reach either fixation *or* extinction. It can be calculated by regarding how the mean from one starting position n relates to the mean of its neighbours.

$$\tau(n) = \Delta t + d(n)\tau(n - 1) + (1 - b(n) - d(n))\,\tau(n) + b(n)\tau(n + 1) \tag{3.5}$$

Substituting in the values of the increase and decrease rates and rearranging this gives

$$\tau(n + 1) - 2\tau(n) + \tau(n - 1) = -\frac{\Delta t}{b(n)} = -\Delta t \frac{N^2}{n(N - n)}.$$

Similar to the Fokker-Planck approximation, I approximate the LHS of the above with a double derivative (ie. $1 \ll N$) to get $\frac{\partial^2 \tau}{\partial n^2} = -\Delta t\, N \left(\frac{1}{n} + \frac{1}{N - n} \right)$. Double integrate and use the bounds $\tau(0) = 0 = \tau(N)$ gives

$$\tau(n) = -\Delta t\, N^2 \left(\frac{n}{N} \ln\left(\frac{n}{N} \right) + \frac{N - n}{N} \ln\left(\frac{N - n}{N} \right) \right). \tag{3.6}$$

Note that it was not necessary to use the large N approximation, there is an exact solution [?]:

$$\tau(n) = \Delta t\, N \left(\sum_{j=1}^{n} \frac{N - n}{N - j} + \sum_{j=n+1}^{N} \frac{n}{j} \right). \tag{3.7}$$

3.7   Moran model with immigration

In section (3.5) I argued that qualitatively different steady states are expected depending on a comparison of the timescales of immigration/mutation/speciation and invasion attempts. If new species enter the system faster than they go extinct, the number of extant coexisting species, and hence the biodiversity, should increase to some steady state. Conversely, if extinction is much more rapid than speciation, a monoculture of one single species is expected in the system. Whether the monocultural system consists of the same species over multiple invasion attempts or whether it experiences sweeps, changing from a monoculture of one species to the next, depends on the probability of a successful invasion. Results can easily be simulated, but to get better insight into the role of the parameters on the results I look for analytic solutions, and as such I treat a simplified model, that of the Moran model with immigration.

The Moran model with immigration is akin to the Hubbell model [71]. His work reinvigorated the debate between niche and neutral mechanisms of biodiversity maintenance. Early numerical solution of the Hubbell model was done by Bell [17]. Ultimately, it is simply a Moran model with immigration, where the immigrant species is never from one of the extant species (from Hubbell's perspective, the newcomers arise via speciation rather than immigration or simple mutation). Hubbell aggregated his theory to describe species abundance curves, rather than my interests of the population distribution or lifetime of a single species. With regard to the stationary probability distribution of a single species there was some pioneering work done by Crow and Kimura [43, 86], who had to assume both continuous time (as do I) and continuous population densities (which I do not), arriving at some numerical results. The techniques available in the field have evolved, and I highlight the work of McKane and Solé [109], which follows techniques similar to Hubbell but calculates the single species distribution. The difference between my work and that of McKane *et al.* is that I analyze the distribution in the context of differing timescales, so I calculate the conditions for monocultures versus biodiversity. I also look at the lifetime of a single species. Hubbell did this a little in his book [71] and later [**?**], and it has since been regarded in more detail by others [83**?**]. There remains a gap in that no one, so far as I know, has considered the conditional extinction time where the conditions are the focal species going extinct or else fixating in the system. My motivation is an impressive bit work from Gore and others [150], measuring the gut microbiome of bacteria-consuming *C. elegans* grown in a 50-50 environment of two strains of fluorescently-labeled but otherwise identical *E. coli*. After an initial colonization period, each nematode has a stable number of bacteria in their gut, presumably from a balance of immigration, birth, and death/emigration. The researchers find the population distribution depending on the comparison of two experimental timescales, those of establishment and fixation time conditioned on a successful invasion. In this section I calculate the stationary probability distribution of a single species, analyzing the critical parameter choices that change its qualitative form. Later, I find the probability of (temporary) extinction versus fixation and the first passage times conditioned on these two possibilities.

The Moran model in this section is as before: there is a focal species of n organisms, with the remaining $N - n$ organisms being of a different strain (or strains). Again I define a fractional abundance $f = n/N$ of the species on which I focus. The Moran population is treated as a rapidly evolving population, with immigrants coming from a more static metapopulation of larger size and diversity. As with the Moran population, the metapopulation contains the focal species and other species, with m, M and g analogous to n, N and f. That is, an immigrant into the Moran population is a member of the focal species with probability g, and of another species with probability $1 - g$. The metapopulation contains $m = g M$ members of the focal species out of M total organisms. In principle g should be a

random number drawn from the probability distribution associated with an evolving metapopulation, but for $M \ll N$ one can treat it as fixed. In practice, I am assuming that the metapopulation changes much slower than the Moran population. In the context of the Gore experiment, the system of interest is the nematode gut, and the metapopulation is the environment in which the nematode lives (and eats). The consumption of one bacterium will influence the gut microbiome while having a negligible effect on the external environment. In a more general setting, the system of interest is a small island receiving immigrants from a larger mainland; the arrival of one immigrant on the island can be impactful even when the loss of that same emigrant is negligible to the mainland.

Each step of the Moran model involves one birth and one death. I leave the death unchanged, killing the focal species with probability f. Immigration is incorporated by having a fraction ν of the birth events be replaced by immigration events. The regular Moran model has the focal species increasing in population with probability $f(1-f)$; this is now modified to occur only a fraction $(1-\nu)$ of the time, and there is also a contribution $\nu g(1-f)$ that increases the focal population when an immigrant enters (fraction ν) of the focal species (fraction g) when a death of a non-focal species occurs (fraction $1-f$). As before, I take the time interval Δt of each step to be infinitesimal, such that b and d are rates, which are:

transition	function	value
$n \to n+1$	$b(n)$	$f(1-f)(1-\nu) + \nu g(1-f)$
$n \to n-1$	$d(n)$	$f(1-f)(1-\nu) + \nu(1-g)f$
$n \to n$	$1 - b(n) - d(n)$	$\left(f^2 + (1-f)^2\right)(1-\nu) + \nu\left(gf + (1-g)(1-f)\right)$

Note that the rates of increase and decrease of the focal species are no longer the same as each other (as they are in the classical Moran model); there is a bias in the system, toward g. Notice that setting the immigration rate ν to zero recovers the classic Moran model.

If a new mutant or immigrant species is unlikely to enter again (ie. if $g \simeq 0$) then the model corresponds to the Moran model with selection [?], which I will not explicitly treat, though it is included in the general treatment below. Also included here results similar to those of the Moran limit of section (3.5) above, with a single immigrant entering the community and then either successfully invading or going (locally) extinct. Since there is immigration from the static metacommunity, the system will never truly fixate, as there will always be immigrants of the 'extinct' species to be reintroduced to the population. Rather, the system will settle on a stationary distribution of P_n, the probability of having n organisms of the focal species. The process has the master equation $\frac{d\,P_n(t)}{dt} = P_{n-1}(t)b(n-1) + P_{n+1}(t)d(n+1) - \left(b(n) + d(n)\right)P_n(t)$, the difference equation of which can be solved in steady state to give

$$\widetilde{P}_n = \frac{q_n}{\sum_{i=0}^{\infty} q_i} \tag{3.8}$$

where

$$q_0 = \frac{1}{b(0)} = \frac{1}{\nu g}$$

$$q_1 = \frac{1}{d(1)} = \frac{N^2}{(N-1)(1-\nu) + \nu N(1-g)}$$

$$q_i = \frac{b(i-1)\cdots b(1)}{d(i)d(i-1)\cdots d(1)}, \qquad \text{for i}_{\dot{\iota}}1$$

$$= \frac{1}{d(i)}\prod_{j=1}^{i-1}\frac{b(j)}{d(j)}$$

recalling that $\frac{b(i)}{d(i)} = \frac{i(N-i)(1-\nu)+\nu Ng(N-i)}{i(N-i)(1-\nu)+\nu N(1-g)i}$. The unnormalized steady-state probability q_n can be written compactly as

$$q_n = \frac{N^2\Gamma(N)\Gamma\left(n + \frac{gN\nu}{1-\nu}\right)\Gamma\left(N - n + 1 + \frac{(1-g)N\nu}{1-\nu}\right)}{\left(n(N-n)(1-\nu) + (1-g)nN\nu\right)\Gamma(n)\Gamma(N-n+1)\Gamma\left(1 + \frac{gN\nu}{1-\nu}\right)\Gamma\left(N + \frac{(1-g)N\nu}{1-\nu}\right)}$$

and the sum of these is the normalization

$$\sum q_i = \frac{1}{g\nu}\frac{\Gamma[1 - \frac{N(1-g\nu)}{1-\nu}]\Gamma[N + 1 - \frac{N}{1-\nu}]}{\Gamma[N + 1 - \frac{N(1-g\nu)}{1-\nu}]\Gamma[1 - \frac{N}{1-\nu}]}$$

which follows formally from the definition of the hypergeometric function $_2F_1$. See figure 3.4 for a visualization of the steady-state probability distribution for different immigration rates.
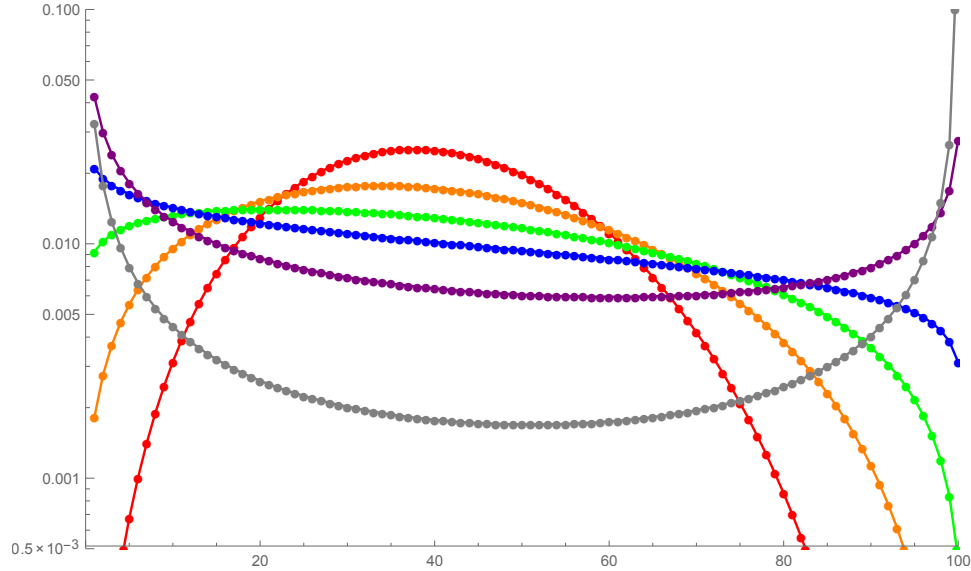
Figure 3.4: PDF of stationary Moran process with immigration. Metapopulation focal fraction $g = 0.4$, local system size $N = 100$, immigration rate ν is given by the colour; red is $10/N$, orange is $5/N$, green is $3/N$, blue is $2/N$, purple is $1/N$, and grey is $0.2/N$. Notice that the curvature of the distribution inverts around $\nu = 2/N$. For high immigration rate the distribution should be centered near the metapopulation fraction gN whereas for low immigration the system spends most of its time fixated.

While the population probability distribution is difficult to calculate before steady state, the mean and variance of the distribution are more tractable at all times. If the mean μ at some time step k has $\mu_k = n_k$ individuals, then after one time step there should be $\mu_{k+1} = n_k - d(n_k) + b(n_k) = n_k + \nu(g - f_k)$ individuals. That is, $\mu_{k+1} - \mu_k = \nu(g - \mu_k/N)$. This is solved by

$$\mu_k = \langle n \rangle (k) = gN \left(1 - (1 - n_0)(1 - \nu/N)^k \right). \tag{3.9}$$

At long times the mean fraction f approaches g, the fraction of the focal species in the metapopulation. To get the an approximation of the variance, I consider the continuous time analogue to the model by taking Δt to be infinitesimal, as described previously. First, the above difference equation of the mean is written as a differential equation $\partial_t \mu(t) = \langle b(n) - d(n) \rangle = \nu(g - \mu(t)/N)$, which has solution $\mu(t) = gN + (\mu_0 - gN)e^{-\nu t/N}$, and the timescale is set by ν/N. The dynamical equation for the second moment is

$$\partial_t \langle n^2 \rangle = 2\langle nb(n) - nd(n) \rangle + \langle b(n) + d(n) \rangle$$
$$= 2\nu \left(g\mu - \langle n^2 \rangle/N \right) + 2(1 - \nu)\left(N\mu - \langle n^2 \rangle \right)/N^2 + \nu(\mu + gN - 2\mu g)/N$$

which is an inhomogeneous linear differential equation. Recalling that $\sigma^2(t) = \partial_t \langle n^2 \rangle(t) - \mu^2(t)$ I solve the above equation and write the variance as

$$\sigma^2(t) = \sigma^2(\infty) + A \exp\{-\frac{\nu}{N}t\} - B \exp\{-2\frac{\nu}{N}t\} + C \exp\{-\frac{2}{N}\left(\nu + \frac{(1-\nu)}{N}\right)t\}$$

where $A = \left(1 + g\nu - g(1 - \nu)/N \right)N^2 \frac{\mu_0 - gN}{N\nu + 2(1-\nu)}$, $B = (gN - \mu_0)^2$, and C is an integration constant; $C = \sigma^2(0) - \sigma^2(\infty) + (gN - \mu_0)^2 + (gN - \mu_0)(2 - \nu)(1 - 2g)/\left(N\nu + 2(1 - \nu)\right)$ if the initial variance is $\sigma^2(0)$. $\sigma^2(\infty) = g(1 - g)N^2 \frac{1}{1+\nu(N-1)}$ is the long time, steady state variance of the system.

Notice that for $g = 0, 1$ the long term variance $\sigma^2(\infty)$ goes to zero. This contrasts with the results of the Moran model without immigration, which has a nonzero variance due to there being a nonzero chance of ending up with the focal species fixated or extinct, with fixation ultimate probability equal to initial fractional abundance. Having a supply of immigrants destabilizes one of these absorbing states; for instance for $g = 0$ the ultimate fate is none of the focal species. This is true even if the initial population fraction was almost entirely of the focal species. If immigration is rare the system may temporarily fixate with the focal species, but with the repeated invasion attempts eventually a non-focal species will fixate, after which the system cannot recover the focal species. Ultimately there is only one fate, hence no variance.

For $g \neq 0, 1$, I would like to consider the case when the time $1/\nu$ between immigration events is truly smaller than the classic Moran model, which has a time scaline proportional to N. In this case we recover similar results to the no immigration case of the Moran model. Instead of $f_0(1 - f_0)N^2$ from the Moran model we get $\sigma^2(\infty) \approx g(1 - g)N^2$, with the metapopulation focal species abundance g acting analogously to the initial abundance f_0. This is easy to intuit. Because the immigration events are rare, each time an immigrant arrives it does so into a system that has already fixated into a monoculture, either of the focal species or without the focal species. A fraction g of the time the immigrant is of the focal species; this is akin to having multiple independent iterations of a classic Moran model, hence the appearance of g as the initial abundance analogue. In the other extreme, immigration happens much more rapidly than the fixation time of the classic Moran model. When $N\nu \gg 1$ the system is

still evolving when a new immigrant is introduced, which acts to keep the probability distribution near g and away from fixation. In this limit the long term variance tends to $\sigma^2(\infty) \approx g(1-g)N/\nu$. For a fixed system size N, increasing the immigration rate decreases the variance, as the system is drawn more toward the metapopulation abundance and away from the extremes of focal species fixation or extinction.

The variance limits, and indeed figure 3.4, suggest that there are at least two parameter space regimes of the Moran model with immigration. At low immigration rate the system undergoes a series of monocultures punctuated by the occasional immigrant. It spends most of its time resting in the fixated state, rarely seeing a new immigrant which quickly either dies out or takes over in a new fixation. When immigration is frequent the system follows the metapopulation, and deviations away from the metapopulation abundance are suppressed. Probability of having n focal organisms gathers near the mean value gN. These regimes will be investigated further in the following paragraphs.

Like the mean, another way to characterize the distribution is the extremum, which for large immigration rate corresponds to the mode of the system. The extremum occurs at the n for which $\partial_n \widetilde{P}_n = 0$. For ease of analysis it is notable that $\partial_n \widetilde{P}_n = \partial_n (q_n / \sum_i q_i) = \partial_n q_n = q_n \partial_n \ln(q_n)$ and therefore I can instead calculate the value which gives $\partial_n \ln(q_n) = 0$. First,

$$\ln(q_n) = 2\ln[N] - \ln\left[n(N-n)(1-\nu) + (1-g)nN\nu\right] + \ln[(N-n)!] + \ln\left[\left(n-1+\frac{\nu gN}{1-\nu}\right)!\right] \quad (3.10)$$

$$+ \ln\left[\left(N-n+\frac{\nu(1-g)N}{1-\nu}\right)!\right] - \ln[(N-n)!] - \ln[(n-1)!] - \ln\left[\left(\frac{\nu gN}{1-\nu}\right)!\right] - \ln\left[\left(N-1+\frac{\nu(1-g)N}{1-\nu}\right)!\right].$$

I have employ the Stirling approximation $\ln[x!] = x\ln[x] - x + O(1/x)$. Setting $\partial_n \ln[q_n] = 0$ and collecting all the non-logarithmic terms to the right-hand side gives

$$\ln\left[\frac{(N-n)(n-1+\nu gN/(1-\nu))}{(n-1)(N-n+\nu(1-g)N/(1-\nu))}\right] = \frac{-2n + N(1-\nu-g\nu)/(1-\nu)}{n\left(-n+N(1-\nu-g\nu)/(1-\nu)\right)}$$

$$= \ln\left[\frac{(1-f)(f-\gamma+\epsilon g)}{(f-\gamma)(1-f+\epsilon(1-g))}\right] = \gamma\frac{1-2f-\epsilon g}{f\left(1-f-\epsilon g\right)} \quad (3.11)$$

where $\gamma = 1/N$ and $\epsilon = \nu/(1-\nu)$, and recalling that $f = n/N$. The parameters γ and ϵ are typically small, so I perform an expansion in them. For this expansion the lowest order in these parameters is $O(\gamma)$, followed by $O(\epsilon\gamma)$. The left-hand side has an infinite series in ϵ starting at $O(\epsilon^1)$, before picking up $O(\epsilon\gamma)$ terms. Keeping only the $O(\epsilon^1)$ terms from the left and $O(\gamma^1)$ terms from the right gives

$$f^* = \frac{1 - g\epsilon/\gamma}{2 - \epsilon/\gamma}. \quad (3.12)$$

This analysis agrees with the observation that there are multiple regimes in parameter space. When immigration is large, $\epsilon/\gamma \approx N\nu \gg 1$, and the maximum or mode of the distribution matches with the mean. The bulk of the probability is centred near gN. But in the opposite limit, when the probability is concentrated at zero and one, the minimal value is half way between these two.

The question remains, how does the distribution switch between these two qualitatively different regimes as ν changes. First, note that there is in fact an intermediate regime, as shown by the blue line $N\nu = 2$ in figure 3.4. The probability need not only be concentrated near both extremes or near gN: for moderate values of immigration there is the possibility that the curvature near one edge of

the domain is positive while it is negative near the other. To this end, I calculate whether the ratio of $\widetilde{P}_0/\widetilde{P}_1$ is greater than one for g (assuming $g < 0.5$) and for the symmetric case $g \leftrightarrow 1 - g$ (rather than also considering $\widetilde{P}_N/\widetilde{P}_{N-1}$ as a function of g). There are three regimes, with two critical parameter combinations dividing them. At the lower division,

$$\frac{\widetilde{P}_0}{\widetilde{P}_1} - 1 = \frac{q_0}{q_1} - 1 = \frac{N - \nu N^2 g - \nu N g - 1 + \nu}{\nu N^2 g} \approx \frac{N - \nu N^2 g}{\nu N^2 g} < 0 \tag{3.13}$$

which implies the probability distribution is concave down when $N\nu > 1/g$. By symmetry the other bound is at $1/(1 - g)$, below which the distribution is concave down. It turns out these same bounds can be found by requiring $0 < f^* \approx \frac{1 - gN\nu}{2 - N\nu} < 1$, since only when the extremum f^* is inside the domain can the distribution have a consistent curvature; when the extremum is outside the domain then the distribution is monotonic and therefore in the intermediate regime. The regimes are shown in figure 3.5.
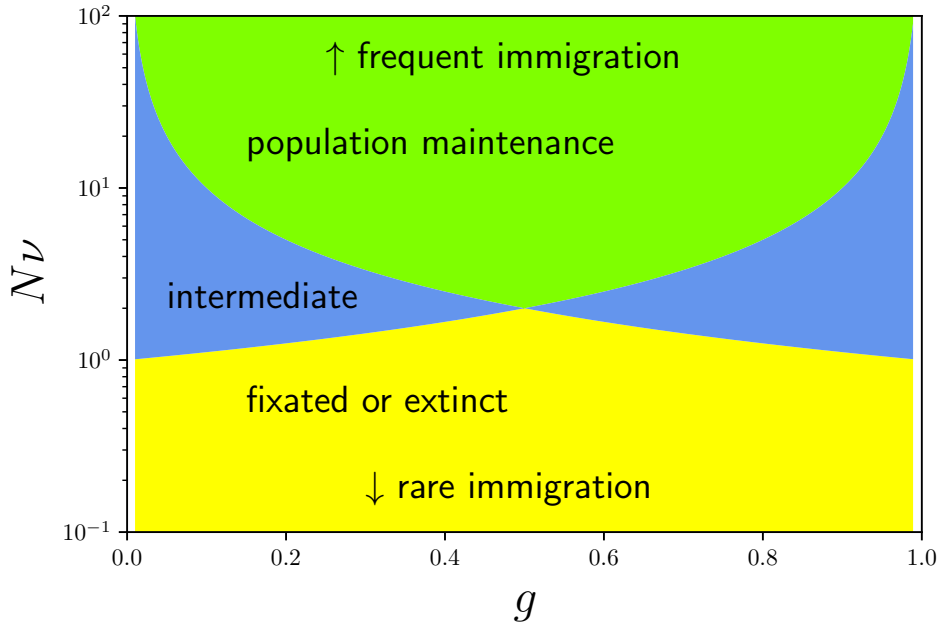
Figure 3.5: The qualitatively different regimes of the system based on the system size N, the immigration rate ν, and the focal species metapopulation abundance g. When immigration is frequent (green region) the focal species is maintained in the population by new immigrants. When immigration is rare (yellow region) the steady state of the system is either an absence or monoculture of the focal species. There is an intermediate regime (blue region) for which the focal species is present but not fixated.

To recapitulate, one should expect the Moran model with immigration to have populations close to one or other of fixation and extinction of the focal species most of the time when the immigration rate is low, specifically $N\nu < \min\left(1/g, 1/(1 - g)\right)$. In the case of frequent immigration, with $N\nu > \max\left(1/g, 1/(1-g)\right)$, the system spends most of its time near the average value gN, with a fraction of the focal species equal to the faction in the metacommunity from which the system receives its immigrants. Qualitatively, these regimes correspond to the system spending most of its time as a monoculture or as having multiple species present, respectively. And there is a third, intermediate regime for $N\nu$ between $1/g$ and $1/(1-g)$ for which the system is often fixated to one extreme but not the other (of $f = 0, 1$), with occasional fluctuations bringing the system away from this extreme. If the metapopulation is equally

likely as not to provide an immigrant of the focal species then there are only the two qualitative regimes of low and high immigration rate.

I had previously written that $N\nu \gg 1$ was the condition of frequent immigration. One also needs to make the comparison between $N\nu$ and $1/g$ to predict, for the focal species, whether it is expected to be locally extinct most of the time (for $N\nu < 1/g$) or maintained at the fractional abundance g (for $N\nu > 1/g$). The parameter regime of the focal species has no direct impact on the rest of the species. The metapopulation is expected to contain many species, thus when any one of them is the focal species it is likely that the associated g is small. For each species i that $N\nu > 1/g_i$ we expect it to exist in the system, and so the number of species with g_i's greater than $1/(N\nu)$ gives a estimate of the expected number of species extant in the system when immigration is frequent. To this extent, the distribution of g_i's in the metapopulation affects the biodiversity of the local system.

Figure 3.4 gives the probability distribution of the species of interest at steady state, but does not allow us to infer anything about the timescales or dynamics of the system. We can guess that if immigration is common the system will fluctuate about its mean, and if immigrants are rare the system will be in a fixated state punctuated by occasional invasion attempts. Given that the focal species in at an intermediate abundance, I want to find the probability of locally fixating before going extinct, and the timescales of these conditions, recognizing that both local fixation and extinction are temporary states, since there is always another immigrant on the way. To make the mathematics more tractable, we must regard at a slightly modified problem, with transition rates changed such that $b(0) = d(N) = 0$. This allows us to find the mean time the system first reaches focal species fixation or extinction. The difference between these results and the results earlier in this chapter is that there is still immigration into the system as it is evolving, which alters the dynamics depending on the rate ν and the immigrant focal fraction g.

The technique I employ follows that laid out in the introduction/appendix/chapter 1. As a brief reminder, define E_i as the probability that the focal species goes extinct in this modified system with absorbing states at $n = 0$ and $n = N$, ie. the system goes to the former before the latter, given that it starts at $n = i$. Then $E_i = \frac{b(i)}{b(i)+d(i)}E_{i+1} + \frac{d(i)}{b(i)+d(i)}E_{i-1}$. Further define $S_i = \frac{d(i)\cdots d(1)}{b(i)\cdots b(1)}$. Then

$$E_i = \frac{\sum_{j=i}^{N-1} S_j}{1 + \sum_{j=1}^{N-1} S_j}. \tag{3.14}$$

See figure 3.6 for a graphical representation of the results. As with the stationary distribution, the extinction probabilities can be written explicitly in terms of N, ν, and g, but graphical interpretation is easier than understanding such a complicated expression.

Unsurprisingly, when immigrants of the focal species are uncommon ($g < 0.5$) the extinction probability is generally increased compared to the Moran model without immigration. The exception is for some n/N values less than g, when the focal fraction in the metapopulation is notably greater than in the local system; in this case the immigration acts to stabilize the population, lessening the probability of extinction before fixation. See the inset of figure 3.6. Unlike with the steady state results, the different trends for the extremes of $N\nu$ compared to $1/g$ and $1/(1-g)$ are less pronounced. Certainly for large immigration rate and population size, for $g < 0.5$ the temporary extinction is almost certain, as is fixation for $g > 0.5$. But all parameter combinations (with $g \neq 0, 1$) result in a leveling of the probability as compared to the Moran model (in black).

Similar to the extinction probabilities, we can write the unconditioned mean first passage time to
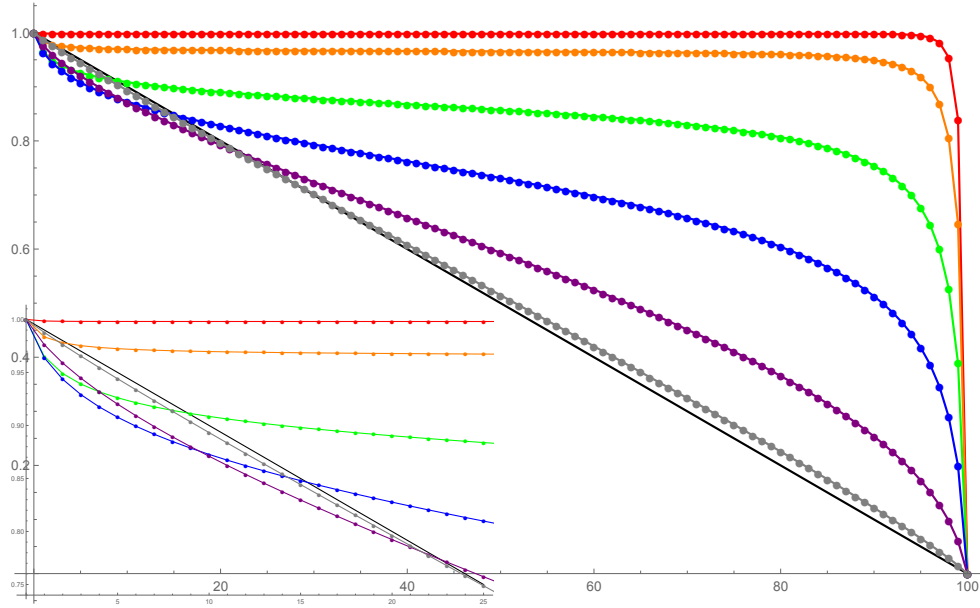
Figure 3.6: Probability of first going extinct rather than fixating, given starting population of the focal species. Metapopulation focal fraction $g = 0.4$, local system size $N = 100$, immigration rate ν is given by the colour; red is $10/N$, orange is $5/N$, green is $3/N$, blue is $2/N$, purple is $1/N$, and grey is $0.2/N$ (same as in figure 3.4). The black line is the regular Moran result without immigration. The inset shows that immigration acts to lower the probability of extinction as compared to the Moran model for some f less than g, even though $g < 0.5$ and more often than not the immigrant is not from the focal species.

either fixation or extinction:

$$\tau[i] = \sum_{k=1}^{N-1} q_k + \sum_{j=1}^{i-1} S_j \sum_{k=j+1}^{N-1} q_k. \tag{3.15}$$

At $n = 0$ the focal species has temporarily gone extinct and at $n = N$ it has fixated; for both of these cases we get $\tau[n] = 0$ since the system has already attained one of these extremes. Again there is a closed form, which I include in the appendix, but it is a sum of hyperbolic functions and so I rely on the graphical representation for interpretation. It is approximated numerically and displayed graphically in figure 3.7. Immigration acts to stabilize the system, drawing it towards g and hence away from the extremes, at which momentary fixation or extinction occurs. Thus a higher immigration rate is expected to increase the mean time until fixation when compared to the regular Moran model. What is more, immigration skews this unconditioned first passage time away from g. Consider the parameters chosen for figure 3.7 with $g = 0.1$. At small n the focal species is more likely to go extinct before it fixates, thus the largest contribution to the unconditioned time is from the mean time conditioned on extinction. Immigration may help delay the inevitable, but the effect is not great, as most immigrants do not increase the focal species population. At large n, however, fixation is the main contributor to the unconditioned time. Most of the immigrants act in opposition to fixation of the focal species, greatly increasing the time to either fixation or extinction. I once again remind the reader that since immigration is unending into the system, both fixation and extinction of the focal species are temporary.

In interpreting the unconditioned mean time I made reference to the times conditioned on local focal fixation or extinction. With the clock stopping when the focal population first reaches 0 or N individuals, I calculate the conditional times, respectively to extinction and to fixation. The extinction probability
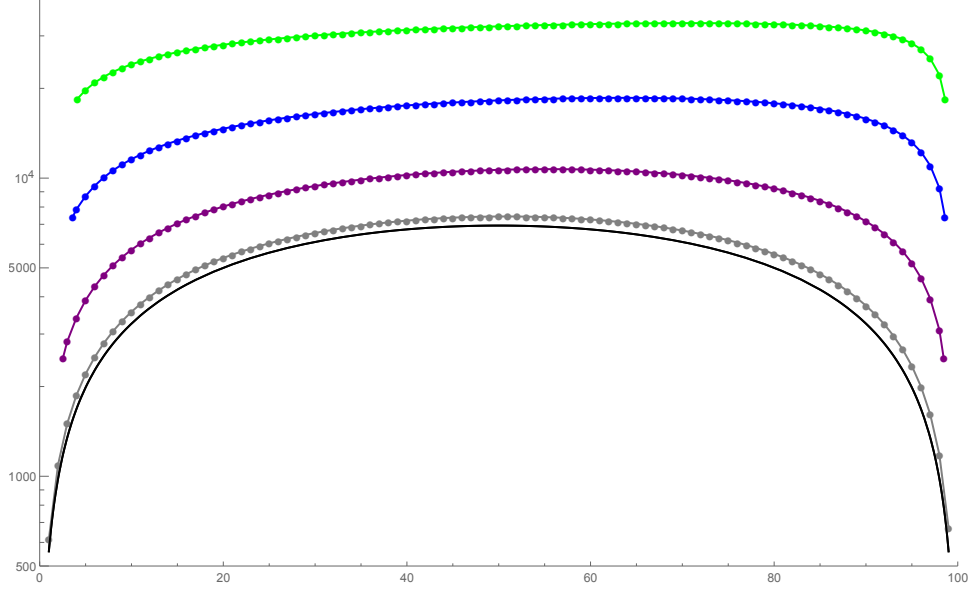
Figure 3.7: Mean time to first reaching either fixation or extinction, from a given starting population of the focal species. Focal immigration fraction is $g = 0.4$, system size is $N = 100$, and immigration rate ν is coloured as before. The black line shows regular Moran results without immigration. Immigration acts to increase the first passage time, and the effects are greatest away from gN.

is given by equation 3.14. Equation 3.3 gives the general equation for solving the conditional time, but it can be written more clearly, following the notation of the fixation probability and unconditioned time, as

$$\phi_n = \phi_1 + \sum_{j=1}^{n-1} \left(\phi_1 - \sum_{i=1}^{j} q_i E_i \right) S_j. \tag{3.16}$$

Here $\phi_i \equiv E_i \theta_i$ (not a dot product, just multiplication of elements), where θ_i is the conditional extinction time [77]. The probability of first fixating before being locally extinct when starting with zero members of the focal species is zero ($E_0 = 0$), so too is $\phi_0 = 0$. The other boundary condition is that the mean first passage time conditioned on fixation is zero if the system starts fixated ($\theta_N = 0$), thus $\phi_N = 0$. These boundary conditions allow me to rearrange the previous equation to get

$$\phi_1 = \frac{\sum_{j=1}^{N-1} \sum_{i=1}^{j} q_i E_i}{1 + \sum_{j=1}^{N-1} S_j}. \tag{3.17}$$

Equation 3.7 substituted into equation 3.7 allows us to solve for ϕ_n, and therefore the conditional time θ_n. One arrives at the graph in figure 3.8. The conditional times mostly follow the unconditioned time, except near the rare events that do not much contribute to the average.

3.8   Discussion

As a reminder, in the research above, even though not all species are equal to each other, their interactions have been symmetric. That is, no species has been given an explicit fitness advantage. The complete neutrality of Hubbell comes when the species not only interact with each other symmetrically but also
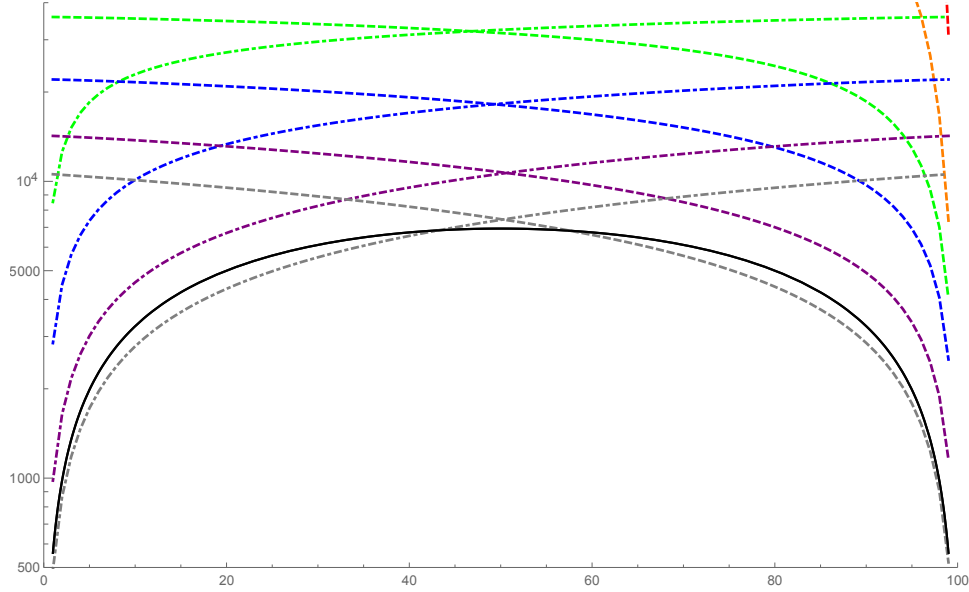
Figure 3.8: Mean time to fixation or extinction, conditioned on that event happening, given starting population/fraction. $g = 0.7$, $N = 100$, ν varies from 0.3 (highest) to 0.0001 (lowest). Grey is regular Moran results without immigration.

interact with other species as strongly as they interact with themselves.

Based on the results above, I can make some comments about the switching behaviour of the Moran population with immigration. Suppose the focal species population starts in state $n = 0$, before any mutations have arisen or immigrants have entered. At a rate νg there will be an attempted invasion by the focal species. Once the invader arrives the dynamics and its ultimate fate depend on how much its niche overlaps with the species currently present in the system. It will be most excluded by those with high population and those with large niche overlap. In the research above I have considered the case of only one extant species upon the arrival of an invader. For species with low niche overlap, the probability of invasion is likely, and for large K decreases monotonically as $1 - a$ with the increase in niche overlap, independent of the population size K. The invader is least likely to be successful in the Moran limit when niche overlap is complete. For invaders that are mutants of the extant wildtype species, this $a = 1$ is the niche overlap they are most likely to experience, and so more similar a mutant to the wildtype, the less likely it is to reach half the population size, which is how I have defined a successful invasion.

Whether or not a mutant invasion is successful, the timescale is longest when niche overlap is high. The times of successful and failed invasions into a stable population set the timescales of the expected transient coexistence in the case of an influx of invaders, arising from mutation, speciation, or immigration. The mean time of successful invasion is relatively fast in all regimes, and scales linearly or sublinearly with the system size K. By contrast, high niche overlap makes invasion difficult due to strong competition between the species. In this regime, the times of the failed invasions become particularly salient because they set the timescales for transient species diversity. We must compare the rate of invasion attempts νg to the time of an invasion attempt. If the influx of invaders is slower than the mean time of their failed invasion attempts, most of the time the system will contain only one settled species, with rare "blips" corresponding to the appearance and quick extinction of the invader [32, 47, 71]. On the other hand, if individual invaders arrive faster than the typical times of extinction of the previous

invasion attempt, they will buoy the population in the system, maintaining its presence. I deal with both of these cases, high or low immigration rate, using the Moran model with immigration when the niche overlap is $a = 1$. For incomplete niche overlap, once a species successfully invades it will persist for long times, based on the results of chapter 2.

Within the Moran model with immigration I have explicitly considered the cases of high and low immigration rate. When immigration is sufficiently high, such that $N\nu > \max\left(1/g, 1/(1-g)\right)$, the focal species is maintained at steady state most often at a fractional abundance equal to that in the metapopulation from which the immigrants arrive. For low immigration rate, specifically $N\nu < \min\left(1/g, 1/(1-g)\right)$, the focal species spends the bulk of its time either locally extinct or else fixated in the system. One way to characterize biodiversity is by the number of different species that reside in a system [71, 105? ?]. An estimate of the expected number of species in a system, at least when immigration is frequent, is given by the number of g_i's greater than $1/(N\nu)$, where g_i is the fractional abundance of species i in the static large metapopulation that provides immigrants to the system.

For incomplete niche overlap, the number of species in a system, as well as their abundances, depends on how their K_i's are distributed. This can be connected back to theories of the apportionment of resources common to niche theories of biodiversity []. Unfortunately calculating the abundance distribution as a function of immigration rate, ecosystem carrying capacity, and niche overlap is outside of the scope of this thesis. Hubbell's species abundance distribution is well known, and is similar to that of Fisher's log series distribution when diversity is high [? ?]. Given that Hubbell's theory is equivalent to the Moran model with immigration, I can make a qualitative argument as to how this will be disrupted with incomplete niche overlap. Those species in disparate niches will exist at their local carrying capacity, and will not be suppressed by their neighbours, thus there should be more species at higher abundance. Only those species that have a naturally low carrying capacity and those that have high niche overlap with others in the system will be found at low abundance or in a transient state. Based on my results, an observed species distribution that is greater than Hubbell's at high abundance but lesser at low abundance is a signature of an ecosystem influenced by niches.

Chapter 4

Ch4-ClosingRemarks

4.1   Experimental tests

It is possible experimentally to corroborate some of the claims made in the first chapter. As previously discussed, measuring the average dynamics alone is insufficient. For example, in a bacterial species the birth rate can be inferred by the uptake and usage of radioisotope-doped nucleotides in nucleic acid synthesis [89]. The death rate is easily inferred from the birth rate and the average dynamics, or it too can be measured using radioisotopes [?]. With these two rates and a couple of 96 well plates it should be a routine procedure to probe the quasi-steady state population probability distribution. A patient experimentalist can also verify the dependence of the MTE on δ and q at low carrying capacity.

The extinction times from the last few chapters are long, and not just those that scale exponentially with the carrying capacity. Even the relatively fast results of the Moran model, which scale linearly with K, will be longer than is experimentally viable for typical biological populations and timescales. The fastest reproducing model organism is the *E. coli* bacterium, which reproduces every twenty minutes in ideal conditions [?]. A carrying capacity as low as 10^3 would show fixation in the Moran limit in a week, but the complete extinction of the population (as described in the first chapter) would take longer than life has existed on the Earth. For example, the famous long running experiments by Lenski [?] would take something like 10^{10^8} years for one of their vials to have all the bacteria die out due to demographic fluctuations. A typical bacterial density is $10^6 - 10^8$ per millilitre. Larger organisms tend to have lesser densities but also slower bithrates.

With lab space graciously provided by Josh Milstein at the University of Toronto, and inspired by similar designs [?], some undergraduate students and I attempted to build a microfluidic device that would constrain a bacterial population to a population on the order of tens or a few hundred. Figure 4.1 shows an image of the design, filled with bacteria... The challenge was to design that allowed for the population to receive a constant small influx of nutrients in a confined space (hence a small carrying capacity) without flushing them out of the system too quickly. The flow rate could not be too slow, however, as it was the mechanism by which the bacteria would be well mixed. Otherwise the bacteria would be spacially arranged and death (in this case, being flowed out of the system) would not be random but depend on proximity to the main channel. It later turned out that this was indeed the case, which was problematic both in that it would not allow for competition of species (instead simply favouring the one farthest from the mouth of the chamber) and it would not allow for extinction of a single species
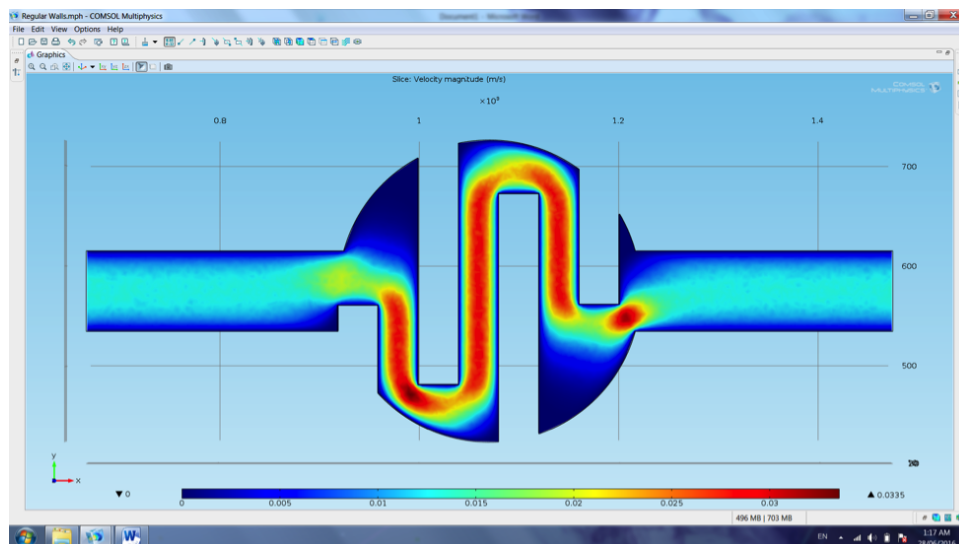
Figure 4.1: words words and more words

(as the death rate dropped to negligible magnitude for bacteria in far back positions). Nevertheless the bacteria grew and were maintained. Further investigation of this system, both in experimental refinement and theoretical modelling, is warranted.

Other researchers have set up systems similar to the microfluidic one I worked on, though typically with different research questions and with larger population sizes.

One remarkable manifestation of the theories I have analyzed in this thesis was done by the Gore lab, growing bacteria in the guts of nematodes [?]. These *C. elegans* worms are grown in an environment filled with red- and green-tagged *E. coli* bacteria that are otherwise identical. The bacteria invade the nematode guts by infrequently surviving the eating process and then slowly reproducing to colonize the system. As the bacteria have the same reproductive rate and chance of entering the gut unscathed this is an experimental realization of the Moran model with immigration from chapter 3 with $g = 0.5$. The researchers use a more simplified model to get steady state probabilities simmilar to those of the data and the Moran model with immigration. The data are too sparse to distinguish between their simple model and the Moran model with immigration. Both theories appear to fit nicely.

4.2 Applications of the theory

The experimental realizations outlined above show very controlled situations where my research could be applicable. The obvious application of the theories investigated in this thesis is to biological systems of few competing species or strains within constant environments. This is somewhat artificial, but there are some systems for which it is relevant. Microfluidic work, in addition to its testing capabilities, could model systems with flow, like industrial food processing or the digestive tract. The *C. elegans* gut microbiome work of Gore and others [? ?] suggests that the theories in this thesis could be applicable to microbiota more generally, be they in the gut or lungs or other areas [?]. My results rely on a well-mixed assumption, so environments like the gut or the ocean surface are good candidates, but situations like plaque growth or trees(desert scrubs?) are not. Species need not rely on the environment to mix them well; animals that are mobile or trees with far-travelling seeds would also be candidates.

Chapters 2 and 3 briefly touch on the idea of small population sizes. For instance, the second chapter's figure ?? suggests that the exponential term is less relevant than the algebraic term when it comes to the fixation time of two competing species as compared to a similarly sized Moran-like system. Indeed, exponential dependence is only dominant when system sizes are large. This large population size is relevant in many (if not most) biological contexts, from *e. coli* and their typical $10^6 - 10^8$ bacteria/mL density [?] to hare in Canada's arctic which are more spread out but number in the ten thousands [?]. But not all biology is overflowing with individuals. One example is that of plasmids in a cell [?]. Plasmids are small loops of DNA that typically code for a few/handful genes, often one of which confers some antibiotic resistance. [?] Their reproduction can be thought of as asexual, since only one plasmid copy is required as a template to make a new copy. Plasmid copy numbers, the average number of copies of a plasmid per cell, tend to be low, ranging from 10^1 to 10^3. The copy number can be thought of as the carrying capacity of that plasmid in the cell, and is maintained primarily by a negative regulatory circuit, whereby a protein expressed by the plasmid acts to inhibit the replication of new plasmids. There is also a stochastic effect when the cell divides and the plasmids are distributed between the two daughters, but given the differing time scales of plasmid (DNA) replication and cell division it is not a necessity that cellular division be the sole mechanism of local extinction; demographic fluctuations like those described in this thesis may also be relevant. Seeing as the bacterial chromosome is typically thousands or a million times longer than that of a plasmid, and as a bacterium can only divide when it has copied its chromosome, I expect the relative time scales to be similarly disparate. A low copy number plasmid might have a carrying capacity of $K = 20$, which would suggest a mean extinction time of 20 million replications, which is a long time for a bacterium, despite the differing time scales of DNA replication and cell division, though for a quickly replicating, low copy number plasmid in a slowly dividing bacterium this could be of relevance. What's more, different types of plasmids can have the same or similar maintenance mechanisms, such that they compete, as mediated by their shared inhibitor proteins. Then the relevant comparison would not be chapter 1's extinction of a single species but the competition of chapters 2 and 3. Each plasmid type would have its carrying capacity given by its copy number if it were alone in the bacteria. The niche in this case is defined by the replication inhibitor, and niche overlap relates how much the inhibitor of each plasmid type stymies the replication of the other; if there is a shared inhibitor, the plasmids are in the Moran limit, and we expect fixation to be rapid, much faster than the cell division time scale.

Similar to plasmids, the number of mitochondria in a cell is small and tends to be controlled within a cell. In [brewer's] yeast there are typically 34 ± 2 mitochondria per cell [?]. Of interest to researchers is how the integrity of mitochondrial DNA is maintained [?]. Sometimes a mitochondrion will have large deletions in its DNA [?]. There are conflicting selection forces in this system: the mutant mitochondria [presumably] reproduce faster than the wildtype but hinder the viability of the cell, hence its reproduction. But when one mutant arises in a population, what is the chance it will fixate, and will fixation occur before the cell reproduces? The analyses of chapter 3 are relevant to such a problem.

Coalescent theory is a model that predicts the time in the past when two variants of a gene most recently shared a common ancestor [?]. In its simplest form it treats all mutants (at least, all that survived to the present day) as equally viable, interacting with other variants as strongly as they do with themselves [?]. In this way it is like a Moran model, or a symmetric Lotka-Volterra model with complete niche overlap. The time it predicts since the last common ancestor is approximately exponentially distributed, with a mean proportional to the system size [?]. An inclusion of incomplete

niche overlap, such as has been investigated in chapter 2, would only act to increase the estimated times. Niche overlap less than one would be most appropriate for very disimilar variants that now serve different but equally vital purposes in the organism. The longer timescale of incomplete niche overlap of course would not change the historical record. Rather, coalescent theory is used to make inferences about historical population genetic parameters, like mutation rate and population size. For those genes to which incomplete niche overlap applies, the longer timescales are counteracted by a smaller population size or greater mutation rate. Conversely, if we know the historic population size and mutation rate, we could infer the niche overlap between disparate gene variants. Phylogenetic reconstruction is related to coalescent theory, and would be similarly affected by incomplete niche overlap [?].

As discussed in the previous chapter, the Hubbell model also relies on assumptions similar to those of the Moran model, to the point that it is effectively a Moran model with immigration [71?]. It is used to predict the number of species that on average can be found in a system, and the distribution of abundances of the species therein. As with coalescent theory, the underlying assumption of Hubbell's neutral theory is that species occupy the same niche. This is effectively what he means by the term "neutral", a claim which has been controversial [?]. Apologists defend the theory by explaining that, in an abstracted sense, species in the same trophic level are effectively competing with each other as much as they compete with themselves [?]. The alternative to neutral theory is traditionally taken to be niche theory, the idea that each species in a system occupies its own niche, and any newcomers must either die out as a transient species or invade a niche, suppressing and evicting its previous occupant. I have shown, in chapter 3, that there are other alternatives. My research investigated the situation where two species have overlapping, but not necessarily identical, niches, such that they do not exclude each other entirely. Neither is selected for, and I have focussed on symmetric interactions where each species affects the other to the same degree, has the same carrying capacity, and turnover rate. Thus I situate myself among those theories which accommodate both niche and neutral theories [?].

4.3   Limitations and caveats

Of course, some of the assumptions underlying the theories outlined in this thesis might fail. My results should be applied with caution. That said, they can at least serve as a minimal working model, from which deviations can be noted and explained. Comparing data against my results would be illuminating in that the way in which the theories fail is indicative of what is a relevant in affecting the system beyond the base model.

For the bulk of this thesis I have assumed the species interactions are neutral; either truly neutral, as in the models of Moran and Hubbell with complete niche overlap, or at least symmetric, in that no species has an explicit fitness advantage. Selection due to differing fitnesses tends to lead to greater fixation probabilities and faster fixation times. It is hard to believe that two species that occupy even partially overlapping niches would not have a fitness difference between them such that one of them is favoured [?] This dissonance is especially potent if two species evolved in different systems and are optimized for different conditions before one of them immigrates to the environment of the other. I have largely been silent on where new strains might arise from. On a molecular level, mutation is most likely to come from single nucleotide polymorphisms, which has three common outcomes: most often, the mutation is deleterious and selected against; often the mutation is synonymous and therefore neutral; occasionally the mutation is beneficial. Synonymous mutations should obey the Moran limit of my results, but to

have symmetric partial niche overlap the mutant must be worse use a resource less effectively than the wildtype strain, while simultaneously not decreasing its basal growth rate. Intuitively this should require at least two mutations; one to decrease the usage efficacy and a second to improve a different resource usage to compensate. The idea is not so far-fetched, however, since most enzymes have some activity on a variety of substrates (typically optimized for one compound), it is not inconceivable that a mutation should decrease the catalytic activity of an enzyme on one substrate while simultaneously improving it for another.

One conceit of this thesis is that most of the parameters, like turnover rate r, carrying capacity K, and niche overlap a, are phenomenological. Often these parameters can be connected to real physical quantities, as I demonstrated in section XXXX with the example of two bacterial strains producing antibacterial toxins. Niche overlap, for example, was an amalgam of the basal birth and death rates, the toxin production and degradation rates, and the death rate increases due to the presence of toxins. For the most part, deriving the phenomenological paramters from physical and biological quantities is outside the scope of this thesis. Phenomenological parameters see wide use in the literature; nevertheless an astute reader should be wary whenever they are presented.

At times I have suggested that those organisms not from the focal species could be all from some second species or from a variety of non-focal species. For incomplete niche overlap this means that those non-focal species all occupy the same niche as each other and have the same overlap with the focal species, a situation which is unlikely. Lumping all non-focal species together is more justified under the assumptions of the Moran and Hubbell models, that in some way each species affects the others as strongly as itself. Neutral theory apologists have made such arguments [**?**]. My results are more readily applied to those systems wherein only two species might coexist.

One of my major assumptions is that species are well-mixed, such that each organism has a chance of interacting with every other organism in the system. I have also allowed for self-interactions, but this is of secondary import. The well-mixed assumption is more valid for some systems than others. It seems to work well with mobile organisms like animals or tree seeds [**?**], or mobile bacteria in a fluid. There are situations for which it is clearly not applicable, where spacial arrangement matters. Then there are situations for which it is unclear. One of the motivating questions for this thesis is the paradox of the plankton, how come there are so many species of plankton that seem to coexist in a seemingly small niche with little variety of resources. One resolution of this paradox is that spacial effect act to stabilize the competing populations, and that otherwise they would collapse to only a few species coexisting [**?**].

Another major caveat of my research is that I have regarded demographic fluctuations while ignoring environmental noise. In reality both sources of stochasticity should contribute, to varying degrees. Typically, environmental fluctuations are of larger magnitude and therefore lead to larger variances and faster first passage times. Thus for even a moderately noisy environment the demographic fluctuations on which I base my results will be superceded and my predictions invalidated.

There are many other simplifications I have employed to make my research questions tractable. For instance, I have assumed that reproduction is (or can at least be effectively treated as being) asexual. As such I can make no comment on how demographic stochasticity and interspecies competition should affect heterozygosity. I have also not considered any life stages or structured populations. Genotypically an organism has died when it can no longer reproduce, yet it still can compete for resources. I treat birth events as being exponentially distributed, and yet even bacteria show a refractory time after reproducing before they can again produce offspring [**?**]. There are many ways that the research contained in this

thesis could be made more realistic. They come at the cost of also being more complicated. My results come from very simple models, which can nevertheless capture qualitative behaviours in extinction rates and coexistance phase diagrams.

4.4   Conclusions and retrospective

This thesis treated extinction from a one dimensional logistic system in chapter 1, extinction from a coupled logistic system in chapter 2, and invasion into a coupled logistic system in chapter 3. The coupled logistic system of chapters 2 and 3 acts like a neutral model in the Moran limit, and the transition to this limit was characterized. Chapter 3 also looks at dynamics and steady state distributions of a single species in a neutral model. The results, in short, are: higher commensurate birth and death rates leads to faster extinction; the WKB approximation is usually fine to use, as often is the Fokker-Planck equation; two species will effectively coexist unless they have exactly the same niche; similarly, greater niche overlap leads to longer invasion times, and less likelihood of success of an invasion attempt; and in a neutral model with repeated immigration a focal species will be most likely to have a moderate population size if $K\nu > \max\big(1/g, 1/(1-g)\big)$, where K is the system size, ν is the immigration rate, and g is the fractional abundance of the focal species in the nearby reservoir population. More detailed discussions and conclusions specifically related to these topics can be found in their respective chapters; below I shall summarize them and extend them beyond the topics of their chapters and to the broader questions discussed in this thesis.

How long will a single species exist with only intraspecies interactions? I have shown that the simplest model of a deterministically stable species, namely the logistic model, does indeed give a mean extinction time whose scale is dominated by an exponential dependence on the system size. This was already known in the literature. What was *not* known was the particulars of how the parameters affect the extinction time. I have demonstrated that, after the carrying capacity, the most impactful parameter on the quasi-steady state distribution and extinction time is the magnitude of the birth and death rates together (rather than their difference). Similarly, acting to simultaneously increase the effect of interspecies interactions on both birth and death while holding their difference constant broadens the distribution and decreases the time. More significantly, this chapter 1 research serves as a warning to those researchers who start their modelling with the deterministic equation and only add noise later. The details of the birth and death rates individually are just as important for stochastic quantities like extinction as are the deterministic parameters. Furthermore, I used this exemplar system to investigate which mathematical techniques are effective to model stochastic extinction. If a researcher wants to model noise in their system but is only concerned with near equilibrium dynamics, the Fokker-Planck and WKB approximations appear acceptable. It is only when investigating far-from-equilibrium quantities like the rare fluctuations that lead to extinction in a deterministically stable system do subtleties start to arise. Specifically, while these two approximations do still capture the exponential dependence on system size, they incorrectly calculate the algebraic prefactor. These approximations remain appealing for getting an idea of the qualitative far-from-equilibrium behaviour but if the goal is to be precise I recommend using the approximation employed throughout most of this thesis, namely introducing a cutoff to the transition matrix and then inverting it to solve the master equation directly.

It has been long known in the literature that extinction from the stochastic logistic model is dominated by an exponential scaling with carrying capacity [15, 80, 120, 122], and so too would be the fixation time

of two independent logistic systems. More recently, it was noted that when the Lotka-Volterra model has equal intra and interspecies interactions it obeys fixation dynamics similar to the Moran model [34, 38, 97, 159]. So then how long will a species exist with intraspecies interactions and some lesser but non-zero interspecies interactions? That is, how long will two species coexist given some niche overlap? It turns out the MTE transitions from exponential scaling with carrying capacity to the relatively fast extinction of algebraic scaling as found in the Moran limit by having the exponential prefactor decrease continuously to zero. Only when there is complete niche overlap does the exponential dependence disappear. For large carrying capacity, this means that two species will effectively coexist if there is even a slight mismatch in their niches. Hubbell's model, being neutral, has equal intra and interspecies interactions. . There have been many pages written either decrying the unintuitive supposition that two disparate species might be equivalent, and many others defending the theory, arguing there must be a way in which species are effectively neutral. My research shows that there is no room for compromise between neutral and niche theories of two species coexisting. Unless the species are truly neutral they will coexist, and their respective carrying capacities should regulate their abundance, rather than the random fluctuations of a neutral model. It remains possible that collections of species are neutral with each other but together exist in a different niche from other such collections.

In terms of coexistence and extinction, neutral models remain just as viable as they were before I did my research. How about regarding the entrance of a new species? For an ecosystem with a species already established, the probability of failure of an invasion attempt of a new species in the large carrying capacity limit is directly proportional to the niche overlap between the established and invading species. The timescale of a successful invasion attempt is never exponential in the system size, being linear at most, in the Moran limit of complete niche overlap, and following the deterministic trend of logarithmic scaling when the invader has independent resource needs from the established species. Unsuccessful invasion attempts are even faster to resolve. The implication is that one could test the Hubbell model by measuring invasion success probability or timescale. If invasion (as defined in chapter 3) is rare then it suggests that a neutral model has some utility in describing the system. If, however, invasion attempts, whether successful or not, resolve themselves rapidly then the system is variegated by different niches, at least in part.

The Hubbell model assumes that each invader is from a new species, but was inspired by the island model of MacArthur and Wilson, which supposes that a small island gets repeated invasions from a larger repository of species on the mainland. This was the motivation for considering the Moran model with immigration in chapter 3, which allows me to answer questions about what the probability distribution of a species will be in a neutral system, and how this distribution depends on the immigration rate. I found that the probability distribution has three characteristic qualitative shapes, with the probability either concentrated at the extremes of extinction and fixation, somewhere near the middle, or mostly near the middle but with sizable density at one of the extremes. The three shapes occur in different regimes of parameter space, bounded by how $K\nu$ compares with $1/g$ and $1/(1-g)$. Each mainland species will have its own abundance g, and the distribution of g's, specifically the number of species with g's greater than $1/(N\nu)$, suggests the number of species we expect to see with population around Kg in the island system. Those with lesser g's will likely have low or zero population in the system.

In the end, I cannot conclude whether neutral or niche theories are more appropriate for explaining the maintenance of biodiversity. Neutral theory is more parsimonious; above I have suggested some bounds or checks to what it should predict, namely that invasions should be rare and slow, and that the

island abundance distribution should follow the mainland abundance distribution. Niche theories are hard to disprove, as they suffer from an overabundance of parameters. Competitive exclusion within a niche and the observation that the ocean surface has very few resources compared to its abundance of species were the initial motivators for my research: the paradox of the plankton. What I can say is that coexisting species need not be in entirely disparate niches; they can effectively coexist with even large, albeit incomplete, niche overlaps.

4.5 Next steps for the research

I will now indulge in some speculation. In this section I present some straightforward extensions of my research. Also included are some more extensive next steps. As with all my research, there is the obvious next step of applying my results to specific real biological systems, finding a way to estimate the phenomenological parameters based on measurable evidence, and then making preditions.

In terms of the approximations covered in this thesis, not much more could be done, unless I omitted a major one. Biophysics ias a field is dnamic and as such I would not be surprised if a new technique gains populatirty in the next few years. Likely the technique already exists and has not made its way to our discipline. It could be in far from equilibrium condensed matter or high energy physics, it might be evolutionary game theory, or already commonly employed in linguistics, economics, graph theory, or the more mathematical side of stochastic processes. Martingales are soluble, so perhaps the next approach will be to map everything to a martingale or a convolution of martingales. Artificial intelligence is also trendy at the time of writing, and the many layered neural nets that are becoming available could easily be turned to stochastic processes, with each neuron representing one population state of the system. Current routinely used nets have XXXX neurons, which corresponds to a sqrt(XXXX) carrying capacity.

The content of the first chapter has a couple extensions I would like to try. The "hidden parameters" were those that did not show up in the deterministic analogue of the system which nevertheless affect the stochastic dynamics. Suppose we write the system as

$$b(n) = r\,n + f(n)$$
$$d(n) = r\,n^2/K + f(n)$$

for arbitrary $f(n)$. I should like to see how f affects the stochastic measurables like MTE. Writing f as a polynomial, I predict the higher orders will have a lesser and lesser effect, and that there will be some conditions placed on $f^{(k)}(0)/k!$, the kth coeffient. Furthermore, I investigated the deterministic equation $\dot{x} = r\,x(1 - x/K)$. I suspect any concave down system will behave similarly [?]. It is unclear what the effects of other forms would be. For instance, Kamenev and others studied inclusion of the Allee effect [?], which to my mind would give extinction time scaline with an effective carrying capacity equal to the difference between the stable and unstable fixed points.

In chapter 2 I regarded how the scaling of the extinction time changes as niche overlap increases from independence to $a = 1$, at which point the MTE is that of the Moran model. It is only in this one limit that coexistence is not long-lived (with the large population caveat examined in section 2.7); with the size of phasespace, this one point seems unlikely, being of measure zero. A next step would be to account for mutation and evolution in this model. If a mutant strain arises from a single population, it is likely to have a very similar niche to the wildtype, hence a large niche overlap. And from there, what would

the forces of evolution dictate? I would continue to constrain the system to the case wherein neither species has an intrinsic fitness advantage. The niche overlap could be allowed to evolve, but it would be more sensible to instead define a niche space (perhaps 1D, perhaps multidimensional) in which the niche of each species is defined (likely with a Gaussian distribution [?]). The desire to increase fitness might have a stabilizing on the system, encouraging differing niches. Or it might keep the niche overlap close to unity, which would go far to explaining why the Hubbell model has had such success. In any case, it would be insightful to see whether fitness considerations act to make the already-unlikely Moran limit entirely untenable, or whether this limit is actually a natural consequence of evolution.

A natural extension of my work would be to include selection, explicit fitness advantages, to the system. I have not done so already because selection has been treated many times in many ways [?] and actually tends to act to simplify the system. The higher fitness species rapidly fixates. Nevertheless there are some advantages to how I treat stochastic systems that would aid the analysis of systems with selection. In the independent limit any fitness advantage should be irrelevant, so there will at least be a transition between coexistence in the independent limit and rapid fixation otherwise. What is more, most models with selection, including those of Kimura [?] and Moran [?] [and others???] must make the assumption that the effect of selection is weak, typically much less than $1/K$. Inverting the transition matrix, as I do in chapter 2, is arbitrarily precise and so does not suffer from any requirement of such an assumption. It can treat the range of selection, from weak to strong. Not only does this give access to regimes not normally considered, it provides a way to verify the small selection results that rely on approximation. In the literature there is a debate about selective sweeps [?] that requires more careful treatment, with fewer assumptions; the method employed in chapter 2 of this thesis would be ideal to explore the differences between hard and soft selective sweeps. Selection can also be incorporated into the invasion dynamics of chapter 3.

One extension applicable to all my work, that I did not account for, is the inclusion of environmental noise. This has a fundamentally different action on the system. Whereas demographic stochasticity comes from the discrete nature of population sizes and therefore can be mapped onto a transition matrix or graph (as in figure 1.1),environmental noise is continuous, especially for the phenomenological parameters I use, which cannot be fixed to any one cause, let alone a discrete one. Such an advancement in my research would involve the abandonment of the transition matrix, and likely would require the use of Fokker-Planck, which I am dubious about, based on my findings in chapter 1.

In chapter 3 I considered the probability and timescales of a single individual attempting to invade a system with an established species, and observed the effect of niche overlap. I then looked at the steady state population distribution and conditional exit times of the focal species in the case of complete niche overlap, via the Moran model. This Moran model was equivalent to the system employed by Hubbell [71], albeit with different questions being asked. However, I could have also asked the questions of Hubbell in the case of incomplete niche overlap. Namely, if species do not occupy the same niche, what should we expect the species abundance curves to look like? With lessened pressure of competing for exactly the same niche I showed that an invader can more easily invade a system, and in lesser time, regardless of its ultimate success or failure. This suggests that a Hubbell model with only partial niche overlap would certainly have more species than the classic Hubbell model with $a = 1$. It is unclear how the abundance curves would change. The Hubbell model has its total population size K, but if species occupy different niches they could in principle have different carrying capacities. Therefore the abundance curve would be influenced by the distribution K is drawn from. In effect this change of niche overlap being partial

and carrying capacities drawn from a distribution would move the model of Hubbell closer to those models of niche apportionment [**?**]. Insofar as no species would have an explicit fitness advantage this new model would be neutral, but with its differing Ks it would also be a niche model, hence a hybrid of the two. There is a body of work for hybrid models [**?**] against which one could compare my proposed incomplete niche overlap Hubbell model.

Just as the Hubbell model is used to make sense of current extant species, coalescent theory uses a Moran or Kimura model to understand paleontological data and other data of the past [**?**]. It seeks to find the time to coalescence, which means the time (in the past) when a common ancestor existed for two variants of a species (in the present). The classic model assumes the variants behave the same - in my language, they occupy the same niche, and have niche overlap of one. The theory has been modified to include explicit fitness advantage [**?**] but the unbiased case of incomplete niche overlap has not been tried. Seeing as decreasing the niche overlap from unity acts to increase the timescales of the system, as I have shown in this thesis, using my results to extend coalescent theory when appropriate would similarly suggest that variants diverged longer ago than the classic theory predicts.

The techniques applied in this thesis, specifically the use of a truncated transition matrix being inverted to solve for the first passage times exactly, can easily be applied to other low species number systems. The calculation for each point in parameter space is not lengthy; the main constraint is RAM, rather than time, and so a larger memory computer could deal with problems with larger carrying capacities or more species. For example, in chapter 2 I considered a two dimensional generalized Lotka-Volterra system to explore competition between two species. The Lotka-Volterra system has been generalized to any dimension, any number of species. A judicious choice of parameters (*e.g.* $a_{12} < 0$, $a_{21} > 0$) will recreate the predator-prey dynamics from which the generalized Lotka-Volterra system gets its name. Assaf *et al.* [**?**] analyzed the predator-prey system using a clever rotating reference frame and the WKB approximation. It would be nice to get the arbitrarily correct data to compare to the approximate results of their research. Furthermore, the deterministic system is marginally stable, with trajectories orbiting a zero-eigenvalue fixed point. As such, I expect an algebraic time to extinction, as well as a stronger dependence on initial conditions. However, the period depends on the value of the Lyapunov energy associated with the orbit. Thus as it "diffuses" tangentially to the orbit its characteristic time scale will change, which may complicate the analysis. Nevertheless, the data that can be generated by inverting a truncated transition matrix to solve the extinction time to arbitrary accuracy would be correct, regardless of the analysis.

From the two dimensional Lotka-Volterra system one arrives at the predator-prey system by choosing the parameters correctly. One can also move to the third dimension, in order to account for a third species in the system. This allows for the investigation of many systems of interest, with much more diversity. The simplest extension in this regard would be to have three species all with overlapping niches. I could observe how a species whose niche is situated between those of two others (such that they each overlap with the first species but not with each other) would go extinct more readily as the overlap of the encroaching species is increased.

A fun model of three interacting species has ones that each are excluded by one other and exclude the other, *e.g.* $a_{12}, a_{23}, a_{31} > 0$ and $a_{21}, a_{32}, a_{13} < 0$. For this reason it is called a rock-paper-scissors system, after the children's game. In any pairwise competition there is a winner, but when all three are combined none has a clear advantage. The model is not just fun, it is also relevant to real-world systems, for instance of lizards [**?**] or bacteria [**?**]. Again I highlight some work from the Gore lab, where

they identified the different bacteria found in a soil sample and performed the combinatorial pairwise competitions. One possible resolution of the unexpectedly large biodiversity one typically observes is that a rock-paper-scissors-like dynamic interaction cyclic dynamics that allow for a greater variety of species than could otherwise coexist.

Another three species model that has enjoyed widespread usage is the SIR model of disease outbreak [?]. It counts the number of susceptible (S), infected (I), and recovered (R) individuals with regards to the disease. There are many variations of this model, but the simplest has a stable spiral node deterministic fixed point. Application of the WKB approximation showed a route to extinction that spiraled in the opposite direction [?]. In chapter 1 I showed that WKB does not recover the correct prefactor for the extinction time, and in chapter 2 I showed that the idea of a route to extinction is questionable. Thus the SIR model is ripe for further analysis.

Armstrong and McGehee showed that, despite there not being a fixed point at which all three species coexisted, there are conditions that allow for three species to indefinitely avoid extinction in a deterministic system with only one resource [9, 143]. This contradicts the intuition that the number of resources constrains the number of species that can exist in a system. I propose an investigation as to whether there is a qualitative difference in the extinction time of such a system as compared to one with the same number of species but a sufficient number of resources to allow for a coexistence fixed point. With stochastic analysis both such systems should eventually exhibit extinction, but the way the mean time scales with the population size scale of the systems could differ. If so, this would serve as a way to distinguish between systems with a fixed point and those with an extended attractor.

For the two dimensional Lotka-Volterra stochastic system I created figure 2.2 identifying qualitatively different regions in parameter space. Especially for a greater number of species, it might be instructive to look at parameters drawn from distributions. Complete niche overlap is of measure zero if niche overlap is a continuous parameter, so it is unlikely to be observed. Distributed parameters may also have an effect on the higher moments of the extinction time beyond the mean, something I have not considered but which is a conceptually-straightforward extension of my research. The extinction time distribution in figure 2.5 looks roughly exponential, in which case the moments can be estimated. Extending to three or more dimensions also allows for the possibility of observing a chaotic system. I would first have to determine what parameter values lead to deterministic chaos, and based on how the parameters are distributed I could estimate how likely this chaos would be observed. But it is unclear whether chaotic and not chaotic systems can be distinguished with the inclusion of stochastic fluctuations. This is a broad research question.

The final next step I will outline is the coupling of the ecology of this thesis with evolution. If species were to change their parameters in response to their environment and the other species, how would the dynamics change? Perhaps each species would evolve to minimize niche overlap while maximizing r and K. If the niches were tied to underlying resources, the species might be constrained in how their niche overlap could evolve, such that they would reach some equilibrium. Or if the evolution itself were stochastic, it might be that whichever happens to find a beneficial mutation first would end up dominating the system. Coupling ecology with evolution is a large field of research, and little has been done with stochasticity, especially demographic stochasticity.

Bibliography

[1] P.A. Abrams. Density-Independent Mortality and Interspecific Competition: A test of Pianka's niche overlap hypothesis. *Am. Nat.*, 111(979):539–552, 1977.

[2] Peter Abrams. Some Comments on Measuring Niche Overlap. *Ecology*, 61(1):44–49, 1980.

[3] Peter Abrams. The Theory of Limiting Similarity. *Annu. Rev. Ecol. Syst.*, 14(1):359–376, 1983.

[4] Peter B. Adler, Stephen P. Ellner, and Jonathan M. Levine. Coexistence of perennial plants: An embarrassment of niches. *Ecol. Lett.*, 13(8):1019–1029, 2010.

[5] Edward J Allen, Linda J S Allen, and Henri Schurz. *A comparison of persistence-time estimation for discrete and continuous stochastic population models that include demographic and environmental variability.*, volume 196. jul 2005.

[6] Linda J S Allen. *An Introduction to Stochastic Processes with Applications to Biology.* Pearson Education, Inc, New Jersey, 2003.

[7] Linda J S Allen and Edward J. Allen. A comparison of three different stochastic population models with regard to persistence time. *Theor. Popul. Biol.*, 64(4):439–449, 2003.

[8] Tibor Antal and Istvan Scheuring. Fixation of strategies for an evolutionary game in finite populations. *Bull. Math. Biol.*, 68(8):1923–1944, 2006.

[9] Robert A Armstrong and Richard McGehee. Coexistence of species competing for shared resources. *Theor. Popul. Biol.*, 9:317–328, 1976.

[10] Robert A. Armstrong and Richard McGehee. Competitive Exclusion. *Am. Nat.*, 115(2):151–170, 1980.

[11] Peter Ashcroft, Franziska Michor, and Tobias Galla. Stochastic tunneling and metastable states during the somatic evolution of cancer. *Genetics*, 199(4):1213–1228, 2015.

[12] Michael Assaf, Alex Kamenev, and Baruch Meerson. Population extinction risk in the aftermath of a catastrophic event. *Phys. Rev. E*, 79(1):011127, jan 2009.

[13] Michael Assaf and Baruch Meerson. Spectral theory of metastability and extinction in birth-death systems. *Phys. Rev. Lett.*, 97(20):1–4, 2006.

[14] Michael Assaf and Baruch Meerson. WKB theory of large deviations in stochastic populations. *J. Phys. A. Math. Gen.*, page 63, 2016.

[15] Michael Assaf, Baruch Meerson, and Pavel V Sasorov. Large fluctuations in stochastic population dynamics: momentum-space calculations. *J. Stat. Mech. Theory Exp.*, 2010(07):P07018, jul 2010.

[16] MattheW A Badali, Jeremy Rotheschild, and Anton Zilman. Intraspecies interactions in birth or in death rates. *(in Prep.*, 2018.

[17] Bell and G. Neutral Macroecology. *Ecology*, 293(5539):2413–2418, 2001.

[18] Graham Bell. The Distribution of Abundance in Neutral Communities. *Am. Nat.*, 155(5):606–617, 2000.

[19] A. Belle, A. Tanay, L. Bitincka, R. Shamir, and E. K. O'Shea. Quantification of protein half-lives in the budding yeast proteome. *Proc. Natl. Acad. Sci.*, 103(35):13004–13009, 2006.

[20] W Bez and P Talkner. A new variational method to calculate escape rates in bistable systems. *Phys. Lett. A*, 82(7):313–316, 1981.

[21] Richard A Blythe and Alan J Mckane. Stochastic Models of Evolution in Genetics , Ecology and Linguistics. *J. Stat. Mech. Theory Exp.*, 2007(07):P07018, 2007.

[22] Immanuel M Bomze. Lotka-Volterra Equation and Replicator Dynamics: A Two-Dimensional Classification. *Biol. Cybern.*, 211:201–211, 1983.

[23] T Börgers and R Sarin. Learning through reinforcement and replicator dynamics. *J. Econ. Theory*, 14:1–14, 1997.

[24] W. Brian Arthur, Yu M. Ermoliev, and Yu M. Kaniovski. Path-dependent processes and the emergence of macro-structure. *Eur. J. Oper. Res.*, 30(3):294–303, 1987.

[25] Vanni Bucci, Carey D Nadell, and João B Xavier. The evolution of bacteriocin production in bacterial biofilms. *Am. Nat.*, 178(6):E162–73, dec 2011.

[26] Yang Cao, Daniel T Gillespie, and Linda R Petzold. Efficient step size selection for the tau-leaping simulation method. *J. Chem. Phys.*, 124(4):044109, jan 2006.

[27] Youfang Cao, Anna Terebus, and Jie Liang. State Space Truncation with Quantified Errors for Accurate Solutions to Discrete Chemical Master Equation. *Bull. Math. Biol.*, 78(4):617–661, 2016.

[28] José A. Capitán, Sara Cuenda, and David Alonso. Stochastic competitive exclusion leads to a cascade of species extinctions. *J. Theor. Biol.*, 419(1980):137–151, 2017.

[29] Ian T. Carroll and Roger M. Nisbet. Departures from neutrality induced by niche and relative fitness differences. *Theor. Ecol.*, 8(4):449–465, 2015.

[30] Ted J Case and RG Casten. Global stability and multiple domains of attraction in ecological systems. *Am. Nat.*, 113(5):705–714, 1979.

[31] P. Chesson. MacArthur ' s Resource Model. *Theor. Popul. Biol.*, 37:26–38, 1990.

[32] Peter Chesson. Mechanisms of Maintenance of Species Diversity. *Annu. Rev. Ecol. Syst*, 31(May):343–66, 2000.

[33] Peter Chesson and Jessica J Kuang. The interaction between predation and competition. *Nature*, 456(7219):235–238, 2008.

[34] Thiparat Chotibut and David R. Nelson. Evolutionary dynamics with fluctuating population sizes and strong mutualism. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 92(2), 2015.

[35] Thiparat Chotibut, David R. Nelson, and Sauro Succi. Striated populations in disordered environments with advection. *Phys. A Stat. Mech. its Appl.*, 465:500–514, 2017.

[36] Jens Claussen and Arne Traulsen. Non-Gaussian fluctuations arising from finite populations: Exact results for the evolutionary Moran process. *Phys. Rev. E*, 71(2):025101, feb 2005.

[37] Bryan Coburn, Pauline W Wang, Julio Diaz Caballero, Shawn T Clark, Vijaya Brahma, Sylva Donaldson, Yu Zhang, Anu Surendra, Yunchen Gong, D Elizabeth Tullis, Yvonne C W Yau, Valerie J Waters, David M Hwang, and David S Guttman. Lung microbiota across age and disease stage in cystic fibrosis. *Sci. Rep.*, 5:10241, 2015.

[38] George W.A. Constable and Alan J. McKane. Models of genetic drift as limiting forms of the Lotka-Volterra competition model. *Phys. Rev. Lett.*, 114(3):1–5, 2015.

[39] Otto X Cordero and Manoshi S Datta. Microbial interactions and community assembly at microscales. *Curr. Opin. Microbiol.*, 31:227–234, 2016.

[40] J Theodore Cox, Mathieu Merle, and Edwin Perkins. Coexistence in a two-dimensional lotka-volterra model. *Electron. J. Probab.*, 15:1190–1266, 2010.

[41] J Cremer, T Reichenbach, and E Frey. Neutral evolution can maintain cooperation in social dilemmas. *Imprint*, 2009.

[42] Jonas Cremer, Tobias Reichenbach, and Erwin Frey. The edge of neutral evolution in social dilemmas. *New J. Phys.*, 11, 2009.

[43] James Crow and Motoo Kimura. Some Genetic Problems in Natural Populations. In *Proc. Third Berkeley Symp. Math. Stat. Probab. Vol. 4 Contrib. to Biol. Probl. Heal.*, number 580, pages 1–22, 1956.

[44] Peter Czuppon and Arne Traulsen. Fixation probabilities in populations under demographic fluctuations. pages 1–33, 2017.

[45] Michael M. Desai and Daniel S. Fisher. Beneficial mutation-selection balance and the effect of linkage on positive selection. *Genetics*, 176(3):1759–1798, 2007.

[46] Jonathan Desponds, Thierry Mora, and Walczak Aleksandra. Fluctuating fitness shapes the clone size distribution of immune repertoires. *Proc. Natl. Acad. Sci.*, 113(2):274–279, 2016.

[47] Paula C Dias. Sources and sinks in population biology. *Trends Ecol. Evol.*, 11(8):326–330, 1996.

[48] Alexander Dobrinevski and Erwin Frey. Extinction in neutrally stable stochastic Lotka-Volterra models. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 85(5):1–14, 2012.

[49] Charles R Doering, Khachik V Sargsyan, and Leonard M Sander. Extinction times for birth-death processes: exact results, continuum asymptotics, and the failure of the Fokker-Planck approximation. *Multiscale Model. Simul.*, 3(2):283–299, 2005.

[50] Vlad Elgart and Alex Kamenev. Rare event statistics in reaction-diffusion systems. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 70(4 1):1–12, 2004.

[51] Alison M Etheridge, Robert C Griffiths, and Jesse E Taylor. A coalescent dual process in a Moran model with genic selection, and the lambda coalescent limit. *Theor. Popul. Biol.*, 78(2):77–92, sep 2010.

[52] C. K. Fisher and P. Mehta. The transition between the niche and neutral regimes in ecology. *Proc. Natl. Acad. Sci.*, 111(36):13111–13116, 2014.

[53] Charles K. Fisher, Thierry Mora, and Aleksandra M. Walczak. Habitat Fluctuations Drive Species Covariation in the Human Microbiota. pages 1–18, 2015.

[54] R.A. Fisher. *The Genetical Theory of Natural Selection.* Clare, Oxford, 1930.

[55] Jonathan Friedman, Logan M. Higgins, and Jeff Gore. Community structure follows simple assembly rules in microbial microcosms. *Nat. Ecol. Evol.*, 1(5):1–7, 2017.

[56] Alan Gabel, Baruch Meerson, and S. Redner. Survival of the scarcer. *Phys. Rev. E*, 87(1):010101, jan 2013.

[57] C.W. Gardiner. *Handbook of Stochastic Methods.* Springer, New York, 3 edition, 2004.

[58] Daniel T Gillespie. Exact Stochastic Simulation of couple chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.

[59] Robert Gooding-Townsend, Steven Ten Holder, and Brian Ingalls. Displacement of Bacterial Plasmids by Engineered Unilateral Incompatibility. *IEEE Life Sci. Lett*, 1:19–21, 2015.

[60] Omer Gottesman and Baruch Meerson. Multiple extinction routes in stochastic population models. *Phys. Rev. E*, 85(2):021140, feb 2012.

[61] Sidhartha Goyal, Sanggu Kim, Irvin SY Chen, and Tom Chou. Mechanisms of blood homeostasis: lineage tracking and a neutral model of cell populations in rhesus macaques. *BMC Biol.*, 13(1):85, 2015.

[62] David Greenhalgh. An epidemic model with a density-dependent death rate. *Math. Med. Biol.*, 7(1):1–26, 1990.

[63] Charles M Grinstead and Laurie J Snell. *Introduction to Probability.* American Mathematical Society, Providence, RI, 2nd edition, 2003.

[64] P Hänggi, P Talkner, and M Borovec. Reaction rate theory - 50 years after Kramers. *Rev. Mod. Phys.*, 62(2):251–341, 1990.

[65] Garrett Hardin. The competitive exclusion principle. *Science (80-.).*, 131(1960):1292–1297, 1960.

[66] D. T. Haydon, N. C. Stenseth, M. S. Boyce, and P. E. Greenwood. Phase coupling and synchrony in the spatiotemporal dynamics of muskrat and mink populations across Canada. *Proc. Natl. Acad. Sci.*, 98(23):13149–13154, 2001.

[67] Christian Hilbe. Local replicator dynamics: a simple link between deterministic and stochastic models of evolutionary game theory. *Bull. Math. Biol.*, 73(9):2068–87, sep 2011.

[68] Josef Hofbauer and Karl Sigmund. Evolutionary game dynamics. *Bull. Am. Math. Soc.*, 40(04):479–520, jul 2003.

[69] Robert D Holt, James P Grover, and David Tilman. Simple Rules for Interspecific Dominance in Systems with Exploitative and Apparent Competition. *Am. Nat.*, 144(5):741–771, 1994.

[70] Bahram Houchmandzadeh, Eric Wieschaus, and Stanislas Leibler. Establishment of developmental precision and proportions in the early Drosophila embryo. *Nature*, 415(6873):798–802, 2002.

[71] S.P. Hubbell. *The Unified Theory of Biodiversity and Biogeography.* Princeton University Press, 2001.

[72] Stephen P Hubbell. Neutral Theory and the Evolution of Ecological Equivalence. *Ecology*, 87(6):1387–1398, 2006.

[73] Paul J Hung, Philip J Lee, Poorya Sabounchi, Nima Aghdam, Robert Lin, and Luke P Lee. A novel high aspect ratio microfluidic design to provide a stable and uniform microenvironment for cell growth in a high throughput mammalian cell culture array. *Lab Chip*, 5(1):44–8, 2005.

[74] Stuart H. Hurlbert. The Measurement of Niche Overlap and Some Relatives. 59(1):67–77, 1978.

[75] G.E. Hutchinson. The Paradox of the Plankton. 95(882):137–145, 1961.

[76] Yen T ing Lin, Hyejin Kim, and Charles R. Doering. Demographic stochasticity and evolution of dispersion I. Spatially homogeneous environments. *J. Math. Biol.*, 70(3):647–678, 2015.

[77] Srividya Iyer-Biswas and Anton Zilman. First passage processes in cellular biology. *Adv. Chem. Phys.*, 160(261), 2016.

[78] C O Jacob. Cytokines and anti-cytokines. *Curr. Opin. Immunol.*, 2(2):249–257, 1989.

[79] Alex Kamenev, Baruch Meerson, and Boris Shklovskii. How Colored Environmental Noise Affects Population Extinction. *Phys. Rev. Lett.*, 101(26):268103, dec 2008.

[80] Alex Kamenev, Baruch Meerson, and Boris Shklovskii. How Colored Environmental Noise Affects Population Extinction. *Phys. Rev. Lett.*, 101(26):268103, dec 2008.

[81] David A Kessler, Robert H Austin, and Herbert Levine. Resistance to chemotherapy: Patient variability and cellular heterogeneity. *Cancer Res.*, 74(17):4663–4670, 2014.

[82] David a. Kessler and Nadav M. Shnerb. Extinction Rates for Fluctuation-Induced Metastabilities: A Real-Space WKB Approach. *J. Stat. Phys.*, 127(5):861–886, 2007.

[83] David A. Kessler and Nadav M. Shnerb. Generalized model of island biodiversity. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 91(4), 2015.

[84] Motoo Kimura. Stochastic processes and distribution of gene frequencies under natural selection. *Cold Spring Harb. Symp. Quant. Biol.*, 20:33–53, 1955.

[85] Motoo Kimura. Diffusion Models in Population Genetics. *J. Appl. Probab.*, 1(2):177–232, 1964.

[86] Motoo Kimura. *The Neutral Theory of Molecular Evolution*. Cambridge University Press, 1983.

[87] J Kingman. The coalescent. *Stoch. Process. their Appl.*, 13(3):235–248, 1982.

[88] James M Kinross, Ara W Darzi, and Jeremy K Nicholson. Gut microbiome-host interactions in health and disease. *Genome Med.*, 3(3):14, 2011.

[89] David Kirchman, Hugh Ducklow, and Ralph Mitchell. Estimates of bacterial growth from changes in uptake rates and biomass . Estimates of Bacterial Growth from Changes in Uptake Rates and Biomass. *Appl. Environ. Microbiol.*, 44(6):1296–1307, 1982.

[90] Peter H. Klopfer and Robert MacArthur. On the Causes of Tropical Species Diversity : Niche Overlap. *Univ. Chicago Press Am. Soc. Nat.*, 95(883):223–226, 1961.

[91] T. Korem, D. Zeevi, J. Suez, A. Weinberger, T. Avnit-Sagi, M. Pompan-Lotan, E. Matot, G. Jona, A. Harmelin, N. Cohen, A. Sirota-Madi, C. A. Thaiss, M. Pevsner-Fischer, R. Sorek, R. Xavier, E. Elinav, and E. Segal. Growth dynamics of gut microbiota in health and disease inferred from single metagenomic samples. *Science (80-.).*, 349(6252):1101–1106, 2015.

[92] Russell Lande. Risks of Population Extinction from Demographic and Environmental Stochasticity and Random Catastrophes. *Am. Nat.*, 142(6):911–927, 1993.

[93] M A Leibold and Mark A McPeek. Coexistance of the Niche and Neutral Perspectives in Community Ecology. *Ecology*, 87(6):1399–1410, 2006.

[94] Matthew A Leibold. The Niche Concept Revisited : Mechanistic Models and Community Context. *Ecology*, 76(5):1371–1382, 1995.

[95] Richard E. Lenski, Michael R. Rose, Suzanne C. Simpson, and Scott C. Tadler. Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations. *Am. Nat.*, 138(6):1315, 1991.

[96] Simon A. Levin. Community Equilibria and Stability, and an Extension of the Competitive Exclusion Principle. *Am. Nat.*, 104(939):413–423, 1970.

[97] Yen T Lin, Hyejin Kim, and Charles R. Doering. Features of Fast Living: On the Weak Selection for Longevity in Degenerate Birth-Death Processes. *J. Stat. Phys.*, 148(4):646–662, 2012.

[98] R. H. MacArthur. On the relative abundance of bird species. *Proc. Natl. Acad. Sci. U. S. A.*, 43(3):293–295, 1957.

[99] R. H. MacArthur and E. O. Wilson. *The theory of island biogeography*. Princeton University Press, Princeton, New Jersey, 1 edition, 1967.

[100] Robert MacArthur. Species Packing and Competitive Equilibrium for Many Species. *Theor. Popul. Biol.*, 11, 1970.

[101] Robert MacArthur and Richard Levins. The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *Am. Nat.*, 101(921):377–385, 1967.

[102] Chaysavanh Manichanh, Jens Reeder, Prudence Gibert, Encarna Varela, Marta Llopis, Maria Antolin, Roderic Guigo, Rob Knight, and Francisco Guarner. Reshaping the gut microbiome with bacterial transplantation and antibiotic intake. *Genome Res.*, 20(10):1411–1419, 2010.

[103] Rachael A Maplestone, Martin J Stone, and Dudley H Williams. The evolutionary role of secondary metabolites–a review. *Gene*, 115(1-2):151–7, 1992.

[104] BJ Matkowsky, Z Schuss, C Knessl, C Tier, and M Mangel. Asymptotic solution of the Kramers-Moyal equation and first-passage times for Markov jump processes. *Phys. Rev. A*, 29(6), 1984.

[105] R May. Unanswered questions in ecology. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.*, 354(1392):1951–9, 1999.

[106] Margaret M. Mayfield and Jonathan M. Levine. Opposing effects of competitive exclusion on the phylogenetic structure of communities. *Ecol. Lett.*, 13(9):1085–1093, 2010.

[107] Richard McGehee and Robert A. Armstrong. Some mathematical problems concerning the ecological principle of competitive exclusion. *J. Differ. Equ.*, 23(1):30–52, 1977.

[108] Alan McKane, David Alonso, and Ricard V. Sole. Mean-field stochastic theory for species-rich assembled communities. *Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top.*, 62(6 B):8466–8484, 2000.

[109] Alan McKane, David Alonso, and Ricard V. Sole. Analytic solution of Hubbell's model of local community dynamics. *Evolution (N. Y).*, 65:10, 2004.

[110] Baruch Meerson and Pavel V. Sasorov. Noise-driven unlimited population growth. *Phys. Rev. E - Stat. Nonlinear, Soft Matter Phys.*, 78(6):2–5, 2008.

[111] Patrick Moran. *The Statistical Processes of Evolutionary Theory*. Clarendon Press, Oxford, 1962.

[112] Brian Munsky and Mustafa Khammash. The finite state projection algorithm for the solution of the chemical master equation. *J. Chem. Phys.*, 124(4), 2006.

[113] Carey D Nadell, Vanni Bucci, Knut Drescher, Simon a Levin, Bonnie L Bassler, and João B Xavier. Cutting through the complexity of cell collectives. *Proc. Biol. Sci.*, 280(1755):20122770, 2013.

[114] Carey D. Nadell, Joao B. Xavier, Simon A. Levin, and Kevin R. Foster. The evolution of quorum sensing in bacterial biofilms. *PLoS Biol.*, 6(1):0171–0179, 2008.

[115] S. Naeem, D. Raffaelli, B. Schmid, J. Bengtsson, J. P. Grime, A. Hector, M. Loreau, D. U. Hooper, D. Tilman, P. Inchausti, D. A. Wardle, and M. A. Huston. Biodiversity and Ecosystem Functioning: Current Knowledge and Future Challenges. *Science (80-.).*, 294(5543):804–808, 2001.

[116] I Nåsell. Extinction and quasi-stationarity in the Verhulst logistic model. *J. Theor. Biol.*, 211(1):11–27, jul 2001.

[117] Claudia Neuhauser and Stephen W . Pacala. An Explicitly Spatial Version of the Lotka-Volterra Model with Interspecific Competition. *Ann. Appl. Probab.*, 9(4):1226–1259, 1999.

[118] T J Newman, Jean-Baptiste Ferdy, and C Quince. Extinction times and moment closure in the stochastic logistic process. *Theor. Popul. Biol.*, 65(2):115–26, mar 2004.

[119] R.M. Nisbet and W.S.C. Gurney. *Modelling Fluctuating Populations*. John Wiley & Sons, Toronto, 1982.

[120] R H Norden. On the Distribution of the Time to Extinction in the Stochastic Logistic Population-Model. *Adv. Appl. Probab.*, 14(4):687–708, 1982.

[121] I. D. Ofiteru, M. Lunn, T. P. Curtis, G. F. Wells, C. S. Criddle, C. A. Francis, and W. T. Sloan. Combined niche and neutral effects in a microbial wastewater treatment community. *Proc. Natl. Acad. Sci.*, 107(35):15345–15350, 2010.

[122] Otso Ovaskainen and Baruch Meerson. Stochastic models of population extinction. *Trends Ecol. Evol.*, 25(11):643–52, 2010.

[123] Jr Palmer, K. Kazmerzak, M. C. Hansen, and P. E. Kolenbrander. Mutualism versus independence: Strategies of mixed-species oral biofilms in vitro using saliva as the sole nutrient source. *Infect. Immun.*, 69(9):5794–5804, 2001.

[124] Matthew Parker and Alex Kamenev. Extinction in the Lotka-Volterra model. *Phys. Rev. E*, 80(2):021129, 2009.

[125] Robin Pemantle. A survey of random processes with reinforcement. *Probab. Surv.*, 4:1–79, 2007.

[126] Elizabeth Pennisi. What determines species diversity? *Science (80-.).*, 309:90, 2005.

[127] Garry Peterson, Craig R. Allen, and C. S. Holling. Ecological Resilience, Biodiversity, and Scale. *Ecosystems*, 1(1):6–18, 1997.

[128] E. R. Pianka. Niche Overlap and Diffuse Competition. *Proc. Natl. Acad. Sci.*, 71(5):2141–2145, 1974.

[129] E.R. Pianka. The Structure of Lizard Communities. 1973.

[130] S Pimm, H L Jones, and Diamond J. On the risk of extinction. *Am. Nat.*, 132(6):757–785, 1988.

[131] Anna Posfai, Thibaud Taillefumier, and Ned S. Wingreen. Metabolic Trade-Offs Promote Diversity in a Model Ecosystem. *Phys. Rev. Lett.*, 118(2):028103, 2017.

[132] E O Powell. Growth rate and generation time of bacteria, with special reference to continuous culture. *J. Gen. Microbiol.*, 15(3):492–511, 1956.

[133] Daniel J Rankin, Leighton A Turner, Jack A Heinemann, and Sam P Brown. The coevolution of toxin and antitoxin genes drives the dynamics of bacterial addiction complexes and intragenomic conflict. *Proc. Biol. Sci.*, 279(1743):3706–15, 2012.

[134] Sean H. Rice. *Evolutionary theory: mathematical and conceptual foundations*. Sinauer Associates, Sunderland, Mass., 2004.

[135] Guus Roeselers, Erika K. Mittge, W. Zac Stephens, David M. Parichy, Colleen M. Cavanaugh, Karen Guillemin, and John F. Rawls. Evidence for a core gut microbiota in the zebrafish. *ISME J.*, 5(10):1595–1608, 2011.

[136] Jeffrey Rogers and Richard A Gibbs. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nat. Rev. Genet.*, 15(5):347–59, 2014.

[137] James Rosindell, Stephen P. Hubbell, and Rampal S. Etienne. The Unified Neutral Theory of Biodiversity and Biogeography at Age Ten. *Trends Ecol. Evol.*, 26(7):340–348, 2011.

[138] I. M. Rouzine, A. Rodrigo, and J. M. Coffin. Transition between Stochastic Evolution and Deterministic Evolution in the Presence of Selection: General Theory and Application to Virology. *Microbiol. Mol. Biol. Rev.*, 65(1):151–185, 2001.

[139] Shovonlal Roy and J. Chattopadhyay. Towards a resolution of 'the paradox of the plankton': A brief overview of the proposed mechanisms. *Ecol. Complex.*, 4(1-2):26–33, 2007.

[140] Thomas W. Schoener. Some comments on Connell's and my reviews of field experiments on interspecific competition. *Am. Nat.*, 125(5):730–740, 1985.

[141] P Schuster and K Sigmund. Replicator dynamics. *J. Theor. Biol.*, pages 533–538, 1983.

[142] A Shmida and S Ellner. Coexistence of plant-species with similar niches. *Vegetatio*, 58(1):29–55, 1984.

[143] S Smale. On the differential equations of species in competition. *J. Math. Biol.*, 3:5–7, 1976.

[144] Emily R. Stirk, Grant Lythe, Hugo A. van den Berg, and Carmen Molina-Paris. Stochastic competitive exclusion in the maintenance of the naive T cell repertoire. *J. Theor. Biol.*, 265(3):396–410, 2010.

[145] Casey M Theriot, Mark J Koenigsknecht, Paul E Carlson Jr, Gabrielle E Hatton, Adam M Nelson, Bo Li, Gary B Huffnagle, Jun Li, and Vincent B Young. Antibiotic-induced shifts in the mouse gut microbiome and metabolome increase susceptibility to Clostridium difficile infection. 2014.

[146] David Tilman. Biodiversity : Population Versus Ecosystem Stability. *Ecology*, 77(2):350–363, 1996.

[147] Arne Traulsen, Jens Christian Claussen, and Christoph Hauert. Coevolutionary dynamics in large, but finite populations. *Phys. Rev. E*, 74(1):011901, jul 2006.

[148] N.G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. North-Holland, Amsterdam, 1992.

[149] Laurence Van Melderen and Manuel Saavedra De Bast. Bacterial toxin-Antitoxin systems: More than selfish entities? *PLoS Genet.*, 5(3), 2009.

[150] Nicole M. Vega and Jeff Gore. Stochastic assembly produces heterogeneous communities in the Caenorhabditis elegans intestine. *PLoS Biol.*, 15(3):1–20, 2017.

[151] Vito Volterra. Fluctuations in the Abundance of a Species considered Mathematically. *Nature*, 118(2972):558–560, 1926.

[152] Marin Vulic and Roberto Kolter. Evolutionary Cheating in. *Genetics*, (March), 2001.

[153] Joshua A Weinstein, Ning Jiang, Richard A White, Daniel S Fisher, and Stephen R Quake. High-Throughput Sequencing of the Zebrafish Antibody Repertoire. *Science (80-.).*, 324(5928):807–810, 2009.

[154] Michael Wink. Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry*, 64(1):3–19, 2003.

[155] Benjamin E. Wolfe, Julie E. Button, Marcela Santarelli, and Rachel J. Dutton. Cheese rind communities provide tractable systems for in situ and in vitro studies of microbial diversity. *Cell*, 158(2):422–433, 2014.

[156] Sewall Wright. Evolution in Mendelian Populations. *Genetics*, 16(2):97, 1931.

[157] Thomas A. Wynn. Type 2 cytokines: mechanisms and therapeutic strategies. *Nat. Rev. Immunol.*, 15:271–282, 2015.

[158] H. Yan, L. Zhao, L. Hu, X. Wang, E. Wang, and J. Wang. Nonequilibrium landscape theory of neural networks. *Proc. Natl. Acad. Sci.*, 110(45):E4185–E4194, 2013.

[159] Glenn Young and Andrew Belmonte. Explicit probability of fixation formula for mutual competitors in a stochastic population model under competitive trade-offs. pages 1–24, 2018.

[160] Xiaoquan Yu and Xiang-Yi Li. Application of WKB and Fokker-Planck methods in analyzing demographic stochasticity. pages 1–19, 2017.

[161] J. X. Zhou, M. D. S. Aliyu, E. Aurell, and S. Huang. Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface*, 9(77):3539–3553, 2012.