

Eco-evolutionary dynamics of finite populations from first principles

A Thesis

submitted to
Indian Institute of Science Education and Research Pune
in partial fulfillment of the requirements for the
BS-MS Dual Degree Programme

by

Ananda Shikhara Bhat



Indian Institute of Science Education and Research Pune
Dr. Homi Bhabha Road,
Pashan, Pune 411008, INDIA.

May, 2023

Supervisor: Vishwesha Guttal
Co-supervisor: Rohini Balakrishnan
© Ananda Shikhara Bhat 2023

All rights reserved

Certificate

This is to certify that this dissertation entitled **Eco-evolutionary dynamics of finite populations from first principles** towards the partial fulfilment of the BS-MS dual degree programme at the Indian Institute of Science Education and Research, Pune represents work carried out by Ananda Shikhara Bhat at the Indian Institute of Science Education and Research, Pune, under the supervision of Vishwesha Guttal, Centre for Ecological Sciences, Indian Institute of Science with Rohini Balakrishnan, Centre for Ecological Sciences, Indian Institute of Science as a co-supervisor, during the academic year 2022-2023.

Vishwesha Guttal

Rohini Balakrishnan

Committee:

Prof. Vishwesha Guttal

Prof. Rohini Balakrishnan

Prof. Sutirth Dey

Declaration

I hereby declare that the matter embodied in the report entitled **Eco-evolutionary dynamics of finite populations from first principles** are the results of the work carried out by me at the Department of Biology, Indian Institute of Science Education and Research, Pune, and the Centre for Ecological Sciences, Indian Institute of Science, under the supervision of Vishwesha Guttal, Centre for Ecological Sciences, Indian Institute of Science, with Rohini Balakrishnan, Centre for Ecological Sciences, Indian Institute of Science as a co-supervisor, and the same has not been submitted elsewhere for any other degree

Ananda Shikhara Bhat

IISER Roll no.: 20181024

Date: 30th March 2023



This thesis is dedicated to obeying the inscrutable exhortations of one's soul.

Acknowledgements

First and foremost, I am grateful to my advisors, Prof. Vishwesha Guttal and Prof. Rohini Balakrishnan, for their masterful mentoring. The enormous academic freedom they gave me to pursue essentially whichever ideas seemed promising at the time, coupled with their readiness to provide detailed guidance whenever I asked for it, made working with them an incredibly enriching experience. Their constant focus on the bigger picture was vital to helping me see the forest when I was stuck looking at the trees. Both Vishu and Rohini are brilliant scientists and fantastic mentors, and their mentorship and guidance have deeply impacted my ways of thinking about science and mentoring. I am also grateful to all the members of the [Theoretical Ecology and Evolution lab](#) and the [Animal Communication and Bioacoustics Lab](#) for providing a supportive and productive working environment.

Prof. Sutirth Dey provided constructive comments during evaluation that greatly improved this thesis. Sutirth's brilliant course on evolution at IISER and various fascinating conversations with him and other lab members during my time at the [Population Biology Lab](#) have been instrumental in shaping my thoughts on evolution. I would like to thank Prof. Deepak Dhar for reviewing this thesis and providing insightful comments that greatly improved its content. I am also extremely grateful to Dr. Anand Krishnan for constantly supporting me and always being there, both as a mentor and a friend, when I needed advice.

I am grateful to many friends who made the IISER journey enjoyable. Countless long days and nights with Adithyan, Abhishek, Arjun, Chebi, Gaurav, Harshit, Manas, Milie, and Shruthi, have greatly helped me grow as a person, both intellectually and emotionally. Arjun has always been a great sounding board for my thoughts, no matter how inane, and has been very patient and supportive in more ways than I could list here. Milie is the occasional fitness friend that everyone needs, and occasionally texting her encouraged me to hit the gym at least semi-regularly even during hectic parts of the thesis project when schedules were uncertain. I am indebted to our D&D group in general and Manas in particular for introducing me to D&D and Critical Role, both of which were very helpful in getting through the pandemic and the thesis year. Shruthi is a constant source of cat related mental health boosts and has more generally taught me a lot about a lot during my time at IISER. I entered this project as someone who knew some mathematics and some biology but quickly realized I would benefit from reading physics papers (typically written in physics language) —

Abhishek always enthusiastically responded to my annoying physics-related questions and patiently explained physics language to me throughout the project year, and for that I am grateful. Insightful conversations about evolution with Gaurav were helpful in gaining fresh perspectives on various topics. Closer to home, Akshay and Prerana have been my closest friends since high school and continue to provide unerring and unconditional love and support. I don't know where I would be without them today.

I am also grateful to various non-human companions, canine, feline, and more exotic. The dogs and cats around my home are more helpful to my mental health than I could possibly articulate. Even though my thesis work may seem dreadfully abstract, running after animals in the Western Ghats and other wild places of India is how I got interested in science in the first place, and I certainly wouldn't be here today if the animals weren't so fascinating :)

Finally, I am eternally indebted to my parents, Raja and Vidya. They have unfailingly supported me at every step of my life, and have nurtured and encouraged an inquisitive wonder of the world that no doubt played a vital role in helping me choose this career path. Not many parents would raise their children on Durrell and Feynman or take time out of their day to help their child find the right plant or jar for keeping an insect in the house. Upon finding out that their child has an early fascination with praying mantids and finding that there is no accessible information about the topic, not many parents would encourage their child to try and write their own source on the subject by keeping these insects at home and documenting their behavior. Upon finding out that their child was interested in snakes, not many parents would encourage this interest and actively accompany them to animal rescue centers, snake parks, and herpetology summer camps. Not many parents would be okay with their child spending weeks in the rainforest with no cellular network catching snakes and looking for frogs in the dark with a flashlight. Certainly, not many Indian parents would let their child pursue this love of animals professionally and allow them to virtually ignore the more 'mainstream' entrance exams for engineering and medical colleges in favor of fully committing to the pipe dream of getting accepted by an IISER or IISc. Not many parents would provide an actively supporting environment at home for an MS project, without questioning their child's future plans or PhD application process at any stage, and just trusting that the child will figure things out. Not many parents would provide fantastic support during difficult emotional situations or not blink an eye when their child decides to grow their hair out or get a tattoo. Thank you for always believing in me and letting me make my own choices, both small and large. I love you.

Funding & Financial Support

I am a recipient (Registration no. SX-1711025) of the Kishore Vaigyanik Protsahan Yojana (KVPY) fellowship from the Department of Science and Technology, Government of India, and gratefully acknowledge the same for providing financial support.

The support and the resources provided by the PARAM Brahma Facility under the National Supercomputing Mission, Government of India at the Indian Institute of Science Education and Research (IISER), Pune are gratefully acknowledged; However, it is worth explicitly noting that while I did briefly run some scripts on the PARAM-Brahma facility that were related to my thesis project, none of the scripts or simulations run on the PARAM-Brahma supercomputer ultimately contributed to this thesis — The main text is completely analytical, and Appendix D only uses very simple simulations that were run on my personal laptop.

Contributions

Contributor Name	Contributor Role
Ananda Shikhara Bhat, Vishwesha Guttal	Conceptualization
Ananda Shikhara Bhat	Methodology
Ananda Shikhara Bhat	Software
Ananda Shikhara Bhat, Vishwesha Guttal	Validation
Ananda Shikhara Bhat	Formal analysis
Ananda Shikhara Bhat	Investigation
Vishwesha Guttal	Resources
Ananda Shikhara Bhat	Data Curation
Ananda Shikhara Bhat	Writing - Original Draft
Ananda Shikhara Bhat, Vishwesha Guttal	Writing - Review and Editing
Ananda Shikhara Bhat	Visualization
Rohini Balakrishnan, Vishwesha Guttal	Supervision
Ananda Shikhara Bhat, Vishwesha Guttal	Project Administration
—	Funding Acquisition

This contributor syntax is based on the Journal of Cell Science CRediT Taxonomy¹

¹<https://journals.biologists.com/jcs/pages/author-contributions>

Abstract

Several recent theoretical studies have shown that noise can have strong impacts on evolutionary dynamics in the limit of small population sizes. In this thesis, I analytically describe the evolutionary dynamics of finite fluctuating populations from first principles to capture the fundamental phenomena underlying such noise-induced effects. Starting from a density-dependent ‘birth-death process’ describing a population of individuals with discrete traits, I derive stochastic differential equations (SDEs) for how the relative population sizes and trait frequencies change over time. These SDEs generically reveal a directional evolutionary force, ‘noise-induced selection’, that is particular to finite, fluctuating populations and is present even when all types have the same fitness. The strength of noise-induced selection depends directly on the difference in turnover rates between types and inversely on the total population size. Noise-induced selection can reverse the direction of evolution predicted by infinite-population frameworks. This general derivation of evolutionary dynamics helps unify and organize several previous studies – typically performed for specific evolutionary and ecological contexts – under a single set of equations. My SDEs also recover well-known results such as the replicator-mutator equation, the Price equation, and Fisher’s fundamental theorem in the infinite population limit, illustrating consistency with known formal descriptions of evolution. Finally, I extend the birth-death formalism to one-dimensional quantitative traits through a ‘stochastic field theory’ that yields equations such as Kimura’s continuum-of-alleles and Lande’s gradient dynamics in the infinite population limit and provides an alternative approach to modelling the evolution of quantitative traits that is more accessible than current measure-theoretic approaches.

Contents

Front Matter	i
Abstract	xvi
Contents	xx
I Motivation & Outline	1
1 Introduction	3
1.1 A (very, very) brief sketch of high-level modelling frameworks in population biology	4
1.2 A very brief outline of the rest of this thesis	9
1.3 A more expanded outline of the rest of this thesis	11
II Theory	15
2 Mathematical background and an expository example	17

2.1	Mathematical Background	18
2.1.1	Birth-death processes	18
2.1.2	SDEs and the Fokker-Planck equation	19
2.1.3	Density-dependence and the intuition for system-size expansions via ecology	23
2.1.4	The intuition for the weak noise approximation in ecology	24
2.2	Warm up: One-dimensional processes for population size	25
2.2.1	Description of the process and the Master Equation	25
2.2.2	The system-size expansion	28
2.2.3	Stochastic fluctuations and the weak noise approximation	29
3	Population dynamics from stochastic first principles	33
3.1	Description of the process and the Master Equation	34
3.2	The system-size expansion	36
3.3	Functional forms of the birth and death rates	37
3.4	Statistical measures for population-level quantities	38
3.5	Stochastic Trait Frequency Dynamics	40
3.6	The infinite population limit	41
3.7	Stochastic fluctuations and the weak noise approximation	43
4	Stochastic field equations for the evolution of quantitative traits	49
4.1	Description of the process and the Master Equation	51
4.2	The functional system-size expansion	55
4.3	The infinite population limit	57
4.4	Stochastic trait frequency dynamics in the infinite-dimensional case	64

4.5 Stochastic fluctuations and the weak noise approximation	64
III Major Takeaways & Discussion	67
5 A unified view of population dynamics	69
5.1 Fundamental equations for evolution in finite populations	70
5.1.1 The fundamental equation for changes in type frequencies	70
5.1.2 The fundamental equation for the mean value of a type-level quantity	71
5.1.3 The fundamental equation for the variance of a type-level quantity .	75
5.2 A stochastic field theory for quantitative traits	78
6 Discussion and Outlook	83
6.1 Noise-induced selection is a generic outcome of evolution in finite, fluctuating populations	87
6.2 Noise-induced selection has concrete implications for finite populations . .	89
6.3 Connections with other theoretical frameworks	92
6.4 Outlook	95
Appendices	99
A From Itô to Fokker-Planck	99
B Stochastic trait frequency dynamics using Itô's formula	103
C A Price-like equation for the variance of a type-level quantity	109
D Some Examples	115

D.1	An example in one dimension: The stochastic logistic equation	115
D.2	An example for discrete traits: Lotka-Volterra and matrix games in finite populations	119
D.3	An example of systematic deviations from neutrality despite equal fitness due to noise-induced selection	123
D.4	Interlude: Detecting modes in quantitative trait distributions through Fourier analysis	127
D.5	An example for quantitative traits: The quantitative logistic equation	131
E	A more elegant representation of the stochastic integrals of Chapter 5	135
F	Behavior of our fundamental equation for trait frequencies at boundaries	139
References		141

Part I

Motivation & Outline

Chapter 1

Introduction

The theory of evolution by natural selection is an ecological theory—founded on ecological observation by perhaps the greatest of all ecologists. It has been adopted by and brought up by the science of genetics, and ecologists, being modest people, are apt to forget their distinguished parenthood

John Harper (1967)

Idealization and generalization are part and parcel of science, be it theorists making unrealistic assumptions on paper to model specific phenomena or experimentalists creating artificially controlled conditions in the laboratory to test specific hypotheses (Zuk and Travisano, 2018). Indeed, some philosophers of science argue that “the epistemic goal of science is not truth, but understanding” (Potochnik, 2018), an idea generally echoed by practicing scientists (Levins, 1966; Servedio et al., 2014; Zuk and Travisano, 2018; Grainger et al., 2022). In other words, since the world is complicated and humans are limited, general understanding inevitably comes at the cost of other desirable qualities such as the ability to make precise quantitative predictions. This is especially true for complex phenomena such as those that are the domain of ecology and evolution, where we are often not even aware of all

the factors that are at play or how they interact. We thus benefit from formulating simple ‘general’ models that provide simple qualitative predictions and help us think about the phenomena we study in a cohesive, unified framework that we can understand well (Potochnik, 2018; Luque and Baravalle, 2021).

A general approach need not (and often will not) be perfect or all-encompassing. As Robert MacArthur once remarked, “general events are only seen by ecologists with rather blurred vision. The very sharp-sighted always find discrepancies and are able to see that there is no generality, only a spectrum of special cases” (Kingsland, 1985). MacArthur was speaking primarily about biological generalities and special cases, but in a related vein, if our language of choice for expressing our general events is mathematics, making any non-trivial observations in complex fields such as ecology and evolution often requires *mathematical ‘coarse-graining’ approximations and idealizations* that ~~may exclude or underplay~~ ignores or ‘lumps together’ several ‘low-level’ model-specific details in favor of a more general description at a ‘higher’ level achieved in some limit that only contains a small number of ‘model-independent’ quantities, often derived from first principles. Formulating and studying such general frameworks can be greatly beneficial as an aid to thinking, sometimes precisely *because* ‘blurry eyed’ thinking that begins from a small set of fundamental first principles and only looks for general broad-brush regularities can be much more insightful than accounting for every little detail or special case. The success of such an approach is perhaps best illustrated by the success of statistical mechanics in physics - statistical mechanics was essentially born from the idea that various useful statements about systems with many moving parts can be made without the need for knowing the excruciating details of every single moving part, and indeed, starting from first principles, this sort of explicitly ‘blurry-eyed’ thinking that only looked at approximate properties was shown to be able to recover the phenomenological laws of thermodynamics as *statistical* laws.

1.1 A (very, very) brief sketch of high-level modelling frameworks in population biology

In biology, arguably the greatest example of a general organizing framework is the idea of evolution by natural selection as synthesized by Darwin from myriad detailed observations of particular systems. Theoretical population genetics has also had a long-standing tradition in building general organizing frameworks that ‘abstract away’ some biological specificities in

favor of a small number of ‘fundamental’ notions like selection and mutation which act on a small number of ‘fundamental’ quantities like fitness. The description of evolution in these general terms was first laid out in formal mathematical terms during the Modern Synthesis by authors such as Wright, Fisher, and Haldane, an extremely successful venture that unified two major schools of thought — Mendelian genetics and Darwinian evolution — that were, at the time, considered to be incompatible (Provine, 2001). It is currently thought that this unification would have been unlikely or would have taken much longer if the architects of the Modern Synthesis had stuck to verbal arguments instead of working with formal models in explicitly mathematical terms (Thompson, 2014).

The success of the Modern Synthesis illustrates the value of formulating abstract mathematical models that only provide a ‘high-level’ description of the fundamental processes required to capture the essence of biological evolution (Provine, 2001; Thompson, 2014). However, the evolutionary play that architects of the Modern Synthesis studied famously unfolds in the ecological theatre (Hutchinson, 1965). Thus, quantities like fitness are not truly fundamental but instead emerge as the net result of various ecological interactions, tradeoffs, and constraints (Metz et al., 1992), a fact that can have important consequences for evolution (Coulson et al., 2006; Schoener, 2011; Kokko et al., 2017). Trying to understand such ‘eco-evolutionary feedbacks’ or ‘eco-evolutionary dynamics’ has sprouted a rich body of literature under the broad heading of ‘evolutionary ecology’ that has greatly enriched our understanding of biological populations (Coulson et al., 2006; Metcalf and Pavard, 2007a; Schoener, 2011; Brown, 2016; Kokko et al., 2017; Lion, 2018; Govaert et al., 2019; Svensson, 2019; Hendry, 2019). Several major theoretical frameworks in the slightly more general setting of eco-evolutionary dynamics — such as evolutionary game theory and adaptive dynamics — as well as the standard equations of population genetics and quantitative genetics, can still be recovered (in a very general sense) as special cases of a slightly reformulated version of something called the ‘Price equation’ (Page and Nowak, 2002; Lion, 2018).

Formulated in the mid to late 20th Century, the Price equation is the most general standard mathematical framework we have for evolutionary population biology (Frank, 2012; Frank, 2017; Queller, 2017; Luque, 2017; Lion, 2018; Lehtonen, 2018; Lehtonen, 2020; Luque and Baravalle, 2021). Indeed, much like statistical mechanics in physics, the Price equation is derived from a small number of very general first principles and is able to recover several standard equations as special cases. The Price equation partitions changes in population composition into multiple terms, each of which lends itself to a straightforward interpretation

in terms of ‘high-level’ evolutionary forces such as selection and mutation, thus providing a useful conceptual framework for thinking about how populations change over time (Frank, 2012). However, the greater complexity of ecology and evolution relative to physics has meant that the generality of the Price equation comes with a much bigger cost in predictive power. The Price equation in its most general formulation is dynamically insufficient¹ (Van Veelen, 2005; Frank, 2012; Simon, 2014; Queller, 2017). However, this need not be the case if we are willing to compromise slightly - several authors have put forth more predictive versions of the Price equation by moving to a continuous time differential equation framework in which the Price equation is dynamically sufficient but manifests in a slightly less general form (Page and Nowak, 2002; Lion, 2018; Day et al., 2020).

These general formulations are still often very difficult to coax concrete quantitative predictions from, but they do lend themselves to simple biological interpretation, and in the dynamically sufficient formulations, often provide *qualitative* predictions. These qualitative predictions, as well as the decomposition of terms in the original Price equation, are useful primarily for their generality — the Price equation gives us a clear idea of which evolutionary forces operate in which systems and when in an almost entirely ‘model-independent’ language (Okasha, 2006; Frank, 2012; Queller, 2017; Luque, 2017). It also leads to a small number of simple yet insightful ‘fundamental theorems’ of population biology (Queller, 2017; Lion, 2018; Lehtonen, 2018) that serve a similar function, and unifies several various seemingly disjoint formal structures of evolution under a single theoretical banner (Lehtonen, 2020; Luque and Baravalle, 2021).

One of the general guiding principles of much of this mathematization has been the assumption that incorporating the reality of finite population sizes into models leads to no major qualitative differences in behavior, only ‘adding noise’ or ‘blurring out’ the predictions of simpler infinite population models (Page and Nowak, 2002). Consequently, several major theoretical frameworks in the field, such as adaptive dynamics, are explicitly formulated

¹Since this is not very standard nomenclature outside theoretical biology and related fields: ‘Dynamically insufficient’ means that the equation relates quantities through time in a manner that cannot be presented in the form of a difference equation $x_{t+1} = F(x_t)$ or a differential equation $\dot{x} = F(x)$. Thus, we cannot predict a dynamic ‘trajectory’ given an initial condition x_0 . The term ‘insufficient’ is to indicate that the equation is only true ‘in retrospect’, requiring complete information about the system at all times. It is thus ‘insufficient’ for prediction. In our case, in its most general setting, the Price equation is usually formulated in a way that partitions a *given* amount of phenotypic change between two populations (usually, but not necessarily, the same population at two different times) into change due to selection, transmission bias, etc., rather than *predicting* a trajectory for how much phenotypic change will occur at various future times based on the *current* phenotypic distribution.

in deterministic terms at the infinite population size limit. However, this assumption is largely unjustified, and since populations in the real world are finite and stochastic, checking whether stochastic models differ from their deterministic analogs is vital to furthering our understanding of the fundamentals of population biology (Hastings, 2004; Coulson et al., 2004; Shoemaker et al., 2020). Today, we increasingly recognize that incorporating the finite and stochastic nature of the real world routinely has much stronger consequences than simply ‘adding noise’ to deterministic expectations (Boettiger, 2018), with important consequences for both ecological (Schreiber et al., 2022) and evolutionary (DeLong and Cressler, 2023) theory. In ecology and evolution, stochastic models need not exhibit phenomena predicted by their deterministic analogues (Proulx and Day, 2005; Johansson and Ripa, 2006; Claessen et al., 2007; Wakano and Iwasa, 2013; Débarre and Otto, 2016; B. Johnson et al., 2021). In addition, they exhibit novel phenomena not predicted by the deterministic approximations (Rogers et al., 2012a; Rogers et al., 2012b; Rogers and McKane, 2015; Veller et al., 2017; DeLong and Cressler, 2023), ~~for example, even completely ‘reversing’ the predictions of deterministic models (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Constable et al., 2016; McLeod and Day, 2019).~~

A striking example of such novel phenomena is given by the complete ‘reversal’ of the predictions of deterministic models that are seen in many finite-population evolutionary models (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Constable et al., 2016; McLeod and Day, 2019). These latter models work with two competing types of individuals (in various model-specific contexts, ex: cheaters and cooperators in evolutionary games), and find in their specific models that while deterministic models in the infinite-population limit predict one particular type being superior due to natural selection, the stochastic finite-population models evolve in the exact opposite direction, with the type that is classically ‘disfavored’ in the infinite-population limit having a higher fixation probability. This phenomenon of selection for the classically disfavored type that leads to the ‘reversal’ of the prediction of deterministic natural selection has been dubbed ‘noise-induced selection’ (Constable et al., 2016; McLeod and Day, 2019; Week et al., 2021), and has been seen in several particular models in fields as diverse as epidemiology (Kogan et al., 2014; Humplik et al., 2014; Parsons et al., 2018; Day et al., 2020), cell-cycle dynamics (Wodarz et al., 2017), life-history evolution (Gillespie, 1974; Veller et al., 2017; Kuosmanen et al., 2022), and social evolution (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Chotibut and Nelson, 2015; Constable et al., 2016; McLeod and Day, 2019; Wang et al., 2023). However, despite these suggestions that stochastic finite-population effects can have profound consequences in a wide variety of

biological systems, as yet, no general model-independent description of noise-induced selection analogous to the Price equation as a description for classical selection and transmission bias exists in the literature.

Studies of neutral or near-neutral dynamics in population and quantitative genetics usually do take stochasticity seriously, explicitly modeling finite populations that follow stochastic dynamics. Unfortunately, the classic or standard stochastic models in both population genetics (Fisher, 1930; S. Wright, 1931; Moran, 1958; Kimura, 1964) and quantitative genetics (Crow and Kimura, 1970; Lande, 1976) typically assume a fixed total population size, thus restricting their validity in a world where population sizes routinely fluctuate (Lambert, 2010). This is unfortunate, since empirical studies indicate that populations with fluctuating population size can exhibit evolutionary dynamics that are starkly different from the predictions of constant population models (Sanchez and Gore, 2013; Chavhan et al., 2019; Chavhan et al., 2021). Theoretical studies concerned with evolution in populations of non-constant size usually impose deterministic and typically phenomenological rules for how the total population size must vary (Kimura and Ohta, 1974; Ewens, 1967; Otto and Whitlock, 1997; Engen et al., 2009; Waxman, 2011). These rules further usually do not depend on population composition. Such models are thus somewhat artificial since demography and population size are forced to be independent quantities even though this is obviously not the case in natural populations, where population size is a ‘bulk’ property whose value emerges from an intricate interplay of the individual-level demographic processes of birth and death (Metcalf and Pavard, 2007a; Lambert, 2010; Geritz and Kisdi, 2012; Doebeli et al., 2017). Notably, the most general framework we have, the Price equation, is typically formulated in a deterministic setting (Page and Nowak, 2002; Frank, 2012; Queller, 2017; Lion, 2018; Day et al., 2020) that ignores stochasticity (but see Rice (2008) and Rice (2020) for a discrete time formulation that is stochastic but, as far as I can tell, is dynamically insufficient, just like the original formulation of the Price equation). Since real-life populations are stochastic, finite, and of non-constant population size, this is somewhat of a problem, since we know that such deterministic approximations that may not capture important dynamics of the real systems of interest.

Incorporating stochasticity into deterministic systems is a tricky business, and, if done in a phenomenological manner by adding noise to a ‘deterministic skeleton’ (Coulson et al., 2004) in an ad-hoc fashion, can very easily result in models that are ill-behaved ~~not self-consistent~~. For example, such models may predict negative population sizes, or lead to biologically

unreasonable predictions such as ‘extinction time paradoxes’ where, for instance, a strong Allee effect can actually appear to increase time to extinction even when a population begins below the Allee threshold, a clearly unreasonable prediction that vanishes if noise is derived properly from first principles (Black and McKane, 2012; Strang et al., 2019). Further, this procedure of adding noise in an ad-hoc manner provides no insight into the mechanistic factors actually responsible for the stochasticity in the first place. Stochastic individual-based models, in which (probabilistic) rules are specified at the level of the individual and population level dynamics are systematically derived from first principles, are self-consistent, much more natural, and can fundamentally differ from the predictions made by simply adding noise terms to a deterministic model (Black and McKane, 2012; Strang et al., 2019). Formulating the fundamental formal structures of evolutionary biology in terms of the mechanistic demographic processes of birth and death at the individual level is also greatly desirable for biological reasons (Metcalf and Pavard, 2007a; Geritz and Kisdi, 2012) because ‘all paths to fitness lead through demography’ (Metcalf and Pavard, 2007b). In other words, since demographic processes such as birth and death rates explicitly account for the ecology of the system, they can more accurately reflect the complex interplay between ecological and evolutionary processes and provide a more fundamental mechanistic description of the relevant evolutionary forces and population dynamics (Doebeli et al., 2017). In this thesis, I present a formulation of population dynamics constructed from mechanistic first principles grounded in individual-level birth and death. The mathematical formalism itself is very general and applies equally to the ‘high level’ forces of population genetics and the ‘high level’ forces of community ecology as postulated by Vellend (2016), though I will mostly stick to the population genetics interpretation in my discussions.

1.2 A very brief outline of the rest of this thesis

This section provides a bullet-point chapter-wise outline for convenience. A more detailed outline of the thesis, with explanations of the technical content covered, originality, etc., is provided in the next section.

This thesis develops a mathematical formalism for describing finite fluctuating populations from first principles and is structured as follows. The rest of this thesis is divided into two parts. Part II provides the complete formalism in all its gory mathematical detail, but in a (hopefully) accessible pedagogical style. Part III then presents some major results from part II and discusses their implications and connections with previous studies. Appendix D

presents some concrete examples of models for clarity regarding the major ideas.

- Chapter 2 provides the necessary mathematical background and provides a toy example studying the size of a population of identical individuals that illustrates the major ideas used. If I am required to shoehorn this thesis into a ‘intro-methods-results-discussion’ format, then chapter 2 can be thought of as providing a mathematical ‘introduction’ and illustration of the ‘methods’ that will be used (in chapter 3) and generalized (in chapter 4) in a biological context to get ‘results’.
- Chapter 3 develops a formalism for describing the evolution of finite fluctuating populations of individuals that come in arbitrarily many ‘types’ that vary in arbitrarily many *discrete* characters. This yields equations that generalize the classic Price equation and replicator-mutator equation to finite fluctuating populations.
- Chapter 4 extends the ideas developed in chapter 3 to populations that vary in a single one-dimensional *quantitative* character and derives a so-called ‘stochastic field theory’ that describes evolution in such populations. This also results in some mathematical equations that may be of independent interest to physicists and applied mathematicians.
- Chapter 5 provides a technical summary of the major results by presenting three important stochastic differential equations. These are equations for type frequencies, the population mean value of an arbitrary type-level quantity, and the population variance of an arbitrary type-level quantity, and respectively generalize the replicator-mutator equation, the Price equation, and Lion’s (2018) variance equation to finite, stochastically fluctuating populations. These equations predict a directional evolutionary force called ‘noise-induced selection’ that is only seen in finite, fluctuating populations. Some implications for social evolution and community ecology are discussed. I also briefly discuss the field equation formalism I develop in chapter 4. Readers who are comfortable with equations but don’t want to go through the details and intermediate steps of the formalism itself can skip to this chapter.
- Chapter 6 provides a quick summary of the major results and discusses biological implications, connections with previous studies, and opportunities for future work. This chapter has no equations (!). Thus, readers who do not like explicitly seeing mathematics in their biology may skip all other chapters and directly read chapter 6 if they are interested only in the final results and takeaways.

1.3 A more expanded outline of the rest of this thesis

Chapter 2 provides the basic mathematical background required and illustrates the major ideas that we will use. To facilitate readership by a broad audience, I only assume passing familiarity with calculus (derivatives, integrals, Taylor expansions) and probability. Familiarity with stochastic calculus is helpful for some sections but is not required, and I present a brief introduction to the relevant notions from both Markov theory and stochastic calculus in section 2.1. In section 2.2, I present a toy example of tracking population size of a population of identical individuals in section. I introduce a description of the system via a ‘master equation’, and then conduct a ‘system-size expansion’ to obtain a Fokker-Planck equation for the system, thus illustrating all the major tools required. For completeness, I also conduct a weak noise approximation to arrive at a so-called ‘linear’ Fokker-Planck equation that can be solved exactly to arrive at a closed-form solution.

Chapter 3 deals with the evolution of discrete traits. In this case, the system is finite-dimensional, since we can completely specify the state of the system by simply listing out the number of individuals of each type in a vector. I introduce a general multivariate process to describe the evolution of discretely varying traits, and use the system size expansion to arrive at a continuous description of change in trait frequencies as an SDE under mild assumptions on the functional forms of the birth and death rates. Unlike many classic stochastic formulations in evolutionary theory (Fisher, 1930; S. Wright, 1931; Moran, 1958; Crow and Kimura, 1970; Lande, 1976; Kimura and Ohta, 1974), I do not assume a fixed (effective) population size and instead allow the total population size to be a natural emergent property from the demographic processes of birth and death. I show that the deterministic limit of this process is the well-known replicator-mutator equation (or equivalently, the dynamic version of the Price equation), thus establishing the microscopic basis of well-known equations from stochastic first principles. I also illustrate some general predictions that can be made using the weak noise approximation for the sake of completeness.

While the mathematics of chapter 3 is standard and well-understood, it has, to the best of my knowledge, not been applied before in the generality and context we use here. Several specific models of specific systems do use these mathematical techniques, but these papers are often written assuming familiarity with notions in physics and/or mathematics and thus may not be very accessible to theoretical ecologists who do not have formal training in these subjects (but see Czuppon and Traulsen (2021) for a recent pedagogical review on the

general approach applied to Wright-Fisher and Moran processes, where total population size is constant). As such, chapters 2 and 3 together also serve as a tutorial and technical introduction to some theoretical ideas: For ecologists, the chapter introduces ‘system size expansions’ and illustrates their use in a general setting, and can be seen as a tutorial on modelling finite populations analytically with minimal assumptions; For population geneticists, the chapter illustrates how system-size approximations (‘diffusion approximations’ in the population genetics literature) can be carried out without assuming a constant (effective) population size and how this generalization has important consequences for the evolutionary forces at play; For physicists and applied mathematicians, the chapter presents a study of the consequences of applying the system-size expansion to the kind of density-dependent birth-death processes that are widely applicable in ecology and evolution - unlike many physical systems, though demographic processes like birth and death over ecological timescales are usually formulated in terms of population numbers or densities, predictions in evolution are typically in terms of frequencies of types, and this fact has subtle consequences that can be overlooked if one only works with densities.

Chapter 4 introduces a function-valued process to model the evolution of quantitative traits such as body size, which can take on uncountably many values. This function-valued process can then also be analyzed via an analog of the system-size approximation to arrive at a ‘functional’ Fokker-Planck equation in which derivatives are replaced by functional derivatives. I show that classic equations from quantitative genetics such as Kimura’s continuum-of-alleles model (Kimura, 1965) and Lande’s gradient dynamics (Lande, 1982) can be derived as the infinite population limit of this stochastic process. I also conduct a weak noise approximation to arrive at a linear functional Fokker-Planck equation that can be analyzed for specific systems as required. Unlike the systems studied in Chapter 3, formalizing the study of the kind of processes we study in Chapter 4 is an active area of mathematical research (Carmona and Rozovskii, 1999; Da Prato and Zabczyk, 2014; Prévôt and Röckner, 2007; Liu and Röckner, 2015; Bogachev et al., 2015; Balan, 2018) and the mathematics itself is far from settled.

Chapter 4 generalizes the work of Tim Rogers and colleagues (Rogers et al., 2012a; Rogers et al., 2012b; Rogers and McKane, 2015) to a wide class of eco-evolutionary systems, and to the best of my knowledge, has never been presented in full generality before. Mathematically, chapter 4 presents heuristic, accessible alternatives to the rigorous tools of martingale theory and measure-valued branching processes that are usually employed to describe the evolution

of quantitative traits (Champagnat et al., 2006; Etheridge, 2011; Week et al., 2021) by generalizing the idea of a system size expansion of density-dependent (finite-dimensional) birth-death processes to the infinite-dimensional case using the notion of functional differentiation. Biologically, Chapter 4 provides ‘stochastic field equations’ that describe the dynamics of one-dimensional quantitative traits in finite populations and illustrates that these equations are consistent with well-known formalisms in quantitative genetics at the infinite population limit.

Part III summarizes the major results of the formalism developed in Part II and presents some simple equations that can be argued to be ‘fundamental equations’ of population biology in the sense of Queller (2017), and together form a ‘unifying perspective’ in the sense of Lion (2018). These equations reduce to well-known results such as the Price equation, the replicator-mutator equation from evolutionary game theory, and Fisher’s fundamental theorem from population genetics in the infinite population limit. For finite populations, these same equations predict a new evolutionary force, ‘noise-induced selection’, that has still not found its way into the standard formal canon of evolutionary biology and whose significance is only recently being recognized (Constable et al., 2016; McLeod and Day, 2019; Mazzolini and Grilli, 2022; Kuosmanen et al., 2022). Implications of noise-induced selection are also discussed in part III. Readers who are okay with mathematical equations but do not want any intermediate derivations (or just trust my math) can skip to Chapter 5 for the major equations that emerge as being important, though I strongly encourage working through the entire formalism properly if possible. Readers who are averse to or do not care for equations can safely skip to Chapter 6 directly for the major takeaways of this thesis.

Part II

Theory

Chapter 2

Mathematical background and an expository example

Like most mathematicians, he takes the hopeful biologist to the edge of a pond, points out that a good swim will help his work, and then pushes him in and leaves him to drown.

Charles Elton (1935), speaking about Lotka

Theorists are often accused of presenting somewhat intuitive ideas in a highly inaccessible formalism that discourages those unfamiliar with the required mathematics. Indeed, many models that use the kind of stochastic processes I use in this thesis assume familiarity with stochastic calculus or Markov theory, or at the very least a willingness to ‘fill in the blanks’ between major results of the calculations. In an attempt to make the ideas I use more accessible to a broader audience, I will use this expository chapter to present a pedagogical summary of the basic mathematical tools required, and present a toy model tracking the population size of a population of identical individuals as an example of the major ideas used.

2.1 Mathematical Background

Here, I provide a brief, informal introduction to some basic notions in stochastic processes. I will make no attempt at rigor and will actively avoid jargon like ‘martingale’ and ‘filtration’. Readers looking for a more comprehensive introduction can refer to standard mathematics texts such as Øksendal (1998), Ethier and Kurtz (1986), or Karatzas and Shreve (1998) for a rigorous treatment of the mathematical foundations, or physics-style texts such as Gardiner (2009) or Van Kampen (1981) for useful tools and techniques to study real systems.

2.1.1 Birth-death processes

Mathematically, a birth-death process is a stochastic process unfolding in continuous time such that

- The process is ‘Markov’, meaning that the future is statistically independent of the past given the present. In more mathematical terms, if the value of the stochastic process at time t is given by X_t , $\mathbb{P}(\cdot|E)$ denotes probability conditioned on E , and $u < s \leq t$ are any three times, then

$$\mathbb{P}(X_t|X_s, X_u) = \mathbb{P}(X_t|X_s)$$

This equation is simply saying that if we have the information about the state of a process at time s , then we do not gain any more predictive power about the process at a future time t if we have additional knowledge about the process from some past time $u < s$. A series of tosses of a fair coin is a simple example of a Markov process, since knowing whether a coin landed on heads during a previous toss does not change your predictive power about whether the coin will land on heads the next time you toss it.

- Transitions are in units of one individual. In one dimension, the phrase ‘birth-death process’ is usually reserved for processes that take values in the non-negative integers $\{0, 1, 2, 3, 4, \dots\}$ such that the only direct transitions are from n to $n \pm 1$. Biologically, this is saying that we observe the population on a fine enough timescale that the probability of two or more births/deaths occurring at the exact same time is very low and we can disallow it entirely in our models. The conditions for higher dimensional birth-death processes look similar.

Since these processes unfold in continuous time, they are characterized not by transition

probabilities but by transition *rates*, which can be thought of as the probability of transition ‘per unit time’. The quantity of interest is usually the probability of being in a particular state at a given point in time. The entire birth-death process can be described in terms of such a quantity, through a so-called ‘Master equation’. The master equation is a partial differential equation (PDE) for the probability of being in a given state at a given time, However, in all but the simplest cases, we can’t actually solve this PDE, because it is simply too hard. The primary source of difficulty is non-linearity in the transition rates and the fact that transitions occur in discrete, discontinuous ‘jumps’. It is much easier to describe and analyze systems that change ‘continuously’.

2.1.2 SDEs and the Fokker-Planck equation

Stochastic systems which change continuously (in the state space) can be described in terms of a ‘stochastic differential equation’ (SDE), which here is interchangeable with the phrase ‘Itô process’. An SDE for a stochastic process $\{X_t\}_{t \geq 0}$ is an equation of the form

$$X_t = \int_0^t F(s, X_s) ds + \int_0^t G(s, X_s) dW_s \quad (2.1)$$

where $F(t, x)$ and $G(t, x)$ are ‘nice’ functions. In the math, physics, and related literature, F and G are often called the ‘drift’ and ‘diffusion’ terms of the process respectively. However, I will not use this terminology here to avoid potential confusion with genetic/ecological drift (which actually manifests in the ‘diffusion’ term G , whereas directional effects like selection manifest in F , the term called ‘drift’ in the math/physics terminology. This can obviously be very confusing).

W_t denotes the so-called ‘Wiener process’ or ‘standard Brownian motion’. Named after the botanist Robert Brown, who was looking at the random erratic motion of pollen grains in water, the (standard) Brownian motion $\{W_t\}_{t \geq 0}$ is a stochastic process that is supposed to model ‘random noise’ or ‘undirected diffusion’ of a particle in a medium. If one imagines W_t as recording the position of a small pollen grain at time t , then W_t can be formally thought of as a model with the following assumptions:

- The pollen grain starts at the origin, *i.e* $W_0 = 0$. This is a harmless assumption made for convenience and amounts to a choice of coordinate system.

- The pollen grain moves without discontinuous jumps across regions of space, *i.e.* the map $t \rightarrow W_t$ is continuous.
- The future movement of the pollen grain is independent of its past history. That is, given any three times $u < s < t$, the displacement $W_t - W_s$ is independent of the past position W_u .
- Between two observations, the pollen grain is equally likely to have moved in any direction, and the distance moved is normally distributed with variance corresponding to the time interval between the two observations (*i.e.* your uncertainty regarding its position is greater if it has been longer since you last saw it). More precisely, given two times $s < t$, the displacement $W_t - W_s$ follows a normal distribution with a mean of 0 and a variance of $t - s$.

It can then be shown that since the motion is equally likely to be in any direction, the expected position at any point of time is the same as the initial position, *i.e.* $\mathbb{E}[W_t | W_0] = W_0 = 0$.

The second integral in equation (2.1) is Itô's 'stochastic integral', and is to be interpreted in the following sense: Fix a time $T > 0$. For any $n \geq 2$, let $\Pi_n = \{t_1, t_2, \dots, t_n\}$ be a partition of the interval $[0, T]$. In other words, the points contained in Π_n divide $[0, T]$ into n slices of the form $[t_i, t_{i+1}]$ such that $0 = t_0 < t_1 < t_2 < \dots < t_n = T$. Then, the (Itô) stochastic integral of the function $G(t, x)$ over the time interval $[0, T]$ is given by

$$\int_0^T G(s, X_s) dW_s := \lim_{n \rightarrow \infty} \sum_{t_i \in \Pi_n} G(t_i, X_{t_i})(W_{t_{i+1}} - W_{t_i})$$

That is to say, it is obtained by dividing our time interval into slices of the form $[t_i, t_{i+1}]$, computing the 'area of the rectangle' formed with $W_{t_{i+1}} - W_{t_i}$ and $G(t_i, X_{t_i})$ as sides, and then taking the limit¹ of finer and finer slices of time. This should look similar to the classic Riemann integral, with the uniform width $t_{i+1} - t_i$ of the Riemann integral replaced by a random width $W_{t_{i+1}} - W_{t_i}$ corresponding to the (random) displacement of a Brownian particle during the uniform time interval $[t_i, t_{i+1}]$.

¹If you are familiar with some real analysis, it bears noting that this limit is in $L^2(\mathbb{P})$, whereas the corresponding limit in the usual Riemann-Stieltjes integral is evaluated pointwise. If you don't know what this sentence means, just ignore it for the purposes of this thesis :)

Equation (2.1) is often represented in the ‘differential’ form:

$$dX_t = F(t, X_t)dt + G(t, X_t)dW_t \quad (2.2)$$

The physics literature also often uses the ‘Langevin’ form:

$$\frac{dx}{dt} = F(t, x) + G(t, x)\eta(t) \quad (2.3)$$

where $\eta(t)$ is supposed to be ‘Gaussian white noise’, a ‘function’ that is defined indirectly such that the integral $\int_0^t G(s, x)\eta(s)ds$ behaves identically to $\int_0^t G(s, X_s)dW_s$. However, it is important to remember that these are both purely formal² expressions - equation (2.2) is meaningless on its own and is really just shorthand for equation (2.1), which is well-defined as explained above; Equation (2.3) is even worse, because the Brownian motion is known to be non-differentiable, and as such, $\eta(t)$ cannot really exist - both equations are thus to be interpreted as shorthand for equation (2.1). SDEs are convenient because they satisfy several ‘nice’ analytical properties. For example, using the fact that the Brownian motion has no expected change in value (*i.e.* $\mathbb{E}[W_t|W_0] = W_0 = 0$), it can be shown that the stochastic integral also has an expectation value of 0 for all t , *i.e.*:

$$\mathbb{E} \left[\int_0^t G(s, X_s)dW_s \middle| X_0 \right] = 0$$

~~Using this, and the fact that the future path of the Brownian motion itself is independent of its history, one can derive the following ‘multiplication table’ for manipulating products of formal expressions containing stochastic integrals with respect to the Brownian motion~~ One important but not immediately obvious consequence of ~~this table~~ the definition of stochastic integrals and the Brownian motion is that we can no longer rely on the normal rules of calculus when dealing with stochastic integrals. In regular calculus, if we had a continuous quantity $x(t)$ satisfying

$$\frac{dx}{dt} = f(x) + g(x)$$

for two ‘nice’ real functions f and g , then, given any real function h , we can calculate how

²Perhaps confusingly, theoretical people often use ‘formal’ to refer to notation or calculation that is devoid of semantic content, to contrast with things that have rigorous meaning. For example, a ‘formal calculation’ can often mean just manipulating the symbols without any rigorous justifications for whether the terms being manipulated exist. Most normal people would probably call this an ‘informal’ calculation :)

the quantity $h(x(t))$ changes over time using the chain rule of differentiation, which says that

$$\frac{dh}{dt} = \frac{dh}{dx} \frac{dx}{dt} = h'(x)f(x) + h'(x)g(x)$$

i.e.

$$dh = h'(x)f(x)dt + h'(x)g(x)dt$$

Naively, we may expect the same logic to still hold true for one-dimensional Itô processes of the form

$$dX_t = F(X_t)dt + G(X_t)dW_t$$

with gdt simply being replaced by GdW_t on the RHS. However, this does not work. The correct relation is instead given by *Itô's formula*³:

$$dh(X_t) = h'(X_t)F(X_t)dt + h'(X_t)G(X_t)dW_t + \frac{h''(X_t)}{2}G^2(X_t)dt$$

There is now an extra $h''(X_t)G^2(X_t)/2$ term that does not exist in the deterministic setting(!). Using Itô's formula and some simple algebra, one can then show that given any process X_t taking values in \mathbb{R} satisfying the SDE (2.2), the associated probability density $P(x, t)$ of finding the process in a state $x \in \mathbb{R}$ satisfies the PDE

$$\frac{\partial P}{\partial t}(x, t) = -\frac{\partial}{\partial x}\{F(t, x)P(x, t)\} + \frac{1}{2}\frac{\partial^2}{\partial x^2}\{(G(t, x))^2P(x, t)\} \quad (2.4)$$

I present a simple informal derivation in Appendix A for the sake of completeness. Equation (2.4) is called the ‘Fokker-Planck equation’ in the physics and applied mathematics literature (Gardiner, 2009) and is often called the ‘Kolmogorov forward equation’ in the population genetics (Ewens, 2004; Barton and Etheridge, 2019) and pure mathematics (Øksendal, 1998) literature. As I explain in Appendix A, the Fokker-Planck equation can be viewed as a ‘conservation law’ for probability that mathematically expresses the common-sense observation that the sum (or integral) of probabilities over the entire state space cannot change over time (since it must always equal 1). If the function G is independent of x , it comes out of the derivatives in equation (2.4), and the resultant Fokker-Planck equation is said to be ‘linear’ (and is much easier to solve). **It bears noting that the Fokker-Planck equation is always linear in the probability density $P(x, t)$, and thus the linearity here means that the equation is**

³Itô's formula also additionally requires $h \in C^2(\mathbb{R})$, meaning that h is continuous and the first and second derivatives of h exist and are also continuous

linear in the *noise* term in the corresponding SDE (Van Kampen, 1981; Gardiner, 2009).

This link between SDEs and Fokker-Planck equations goes both ways: One can show that every stochastic process with a probability density described by a Fokker-Planck equation of the form (2.4) corresponds to the solution of an SDE of the form (2.2), though the proof is much more technical and will not be discussed here. This two-way correspondence proves to be extremely useful, ~~as one approach often works for applications in which the other fails.~~ since the two formulations are complementary and one may be much easier than the other for solving a particular problem. This correspondence makes it greatly desirable to be able to describe our stochastic process of interest as either the solution to an Itô SDE of the form (2.2) or as the solution to a Fokker-Planck equation of the form (2.4). System-size expansions facilitate such a description for birth-death processes.

2.1.3 Density-dependence and the intuition for system-size expansions via ecology

The fundamental idea behind the system-size expansion relates to the nature of the jumps between successive states of a birth-death process. In most situations of interest to population dynamics, at an individual level, births and deaths of individuals are affected by ecological rules that depend on the local population density and not directly on the total population size. Despite this, the jumps themselves occur in terms of the addition (birth) or removal (death) of a *single individual* from the population. If there are many individuals, each individual contributes a negligible amount to the density, and thus, the discontinuous jumps due to individual-level births or deaths can look like a small, *continuous* change in population density. This is the essential idea behind the system-size expansion. The name derives from the formalization of this idea as a change of variable from the discrete values $\{0, 1, 2, \dots, n-1, n, n+1, \dots\}$ to the approximately continuous values $\{0, 1/K, 2/K, \dots, x-1/K, x, x+1/K, \dots\}$ through the introduction of a ‘system size parameter’ K . In physics and chemistry, K is usually the total volume of a container in which physical or chemical reactions take place and is thus a ‘hard’ limit on the number of discrete values allowed. In ecology, we will only impose a ‘soft’ limit by requiring that births and deaths must scale with K in a way that the population almost surely cannot grow indefinitely if K is finite, reflecting the empirical fact that the total amount of resources in the world is limited (of course, this should also be reflected in how birth and death rates vary as functions of population density). Thus, in our systems, K will manifest as some fundamental limit on resources, such as habitat size

or carrying capacity. When K is large, the fact that transitions occur in units of a small value $1/K$ can be exploited via a Taylor expansion of the transition rates in the Master equation, which then yields a Fokker-Planck equation upon neglecting higher order terms⁴. A similar approximation is well-known (ever since Fisher) in theoretical population genetics, where it goes by the name of the ‘diffusion approximation’ (Ewens, 2004; Barton and Etheridge, 2019) or ‘continuum limit’ (Czuppon and Traulsen, 2021), and has been heavily used by Kimura (Crow and Kimura, 1970) in his stochastic models. However, the population genetics version of the approximation usually either relies on total population size being fixed (Crow and Kimura, 1970; Lande, 1976; Ewens, 2004) or is conducted in an ad-hoc manner without specifying an explicit system size parameter.

2.1.4 The intuition for the weak noise approximation in ecology

If the parameter K is sufficiently large, then the Fokker-Planck equation obtained via the system-size expansion can be further simplified to obtain a linear Fokker-Planck equation. This is accomplished by viewing the stochastic dynamics as fluctuating about a deterministic trajectory⁵ (obtained by letting $K \rightarrow \infty$) and only works if K is large enough to be able to neglect all but the highest-order terms. This is usually an excellent approximation for populations in which the deterministic trajectory has already reached an attractor (stable fixed point, stable limit cycle, etc.). Since many deterministic eco-evolutionary models are expected to relax to such attractors, such an approximation is a useful first step in increasing the generality of existing models (which are usually studied only in the equilibrium regime) to incorporate the dynamics of finite populations. ~~Importantly, this approximation only works if we can discard all but highest order terms of K : Including higher order terms leads to equations that do not form Fokker Planck equations and do not even describe probability densities. As such,~~ This approximation is best suited to describe populations that are ‘medium sized’ - small enough that they cannot be assumed to be infinitely large, yet large enough that stochasticity is rather weak and the deterministic limit is somewhat predictive - a situation that occurs frequently in ecology and evolution.

⁴If this sounds handwavy to you, see chapter 11, section 3 in Ethier and Kurtz (1986) for a more rigorous treatment.

⁵This idea can be made much more rigorous via an analog of the central limit theorem for density-dependent Markov chains. See chapter 11, section 2 in Ethier and Kurtz (1986)

2.2 Warm up: One-dimensional processes for population size

The simplest birth-death processes are those in which the state at any time can be characterized by a single number. Populations of identical individuals are an obvious example of such a system. I will use this toy system as an illustration of the techniques that will be used for the actual problems we intend to tackle in the next sections. The mathematics below are adapted from sections 6.3 and 7.2 of Gardiner (2009) to use biological language and more intuitive notation/explanations.

2.2.1 Description of the process and the Master Equation

Consider a population of identical individuals subject to some ecological rules that affect individuals' birth and death rates. I will neglect any potential factor that could lead to two individuals becoming non-identical [in their birth and death rates](#), like mutations, etc. Since all individuals are identical [in their birth and death rates](#), we only really need to track the total population size through time to know everything there is to know about the population. The population as a whole at any time t can thus be characterized by a single number - its population size (Figure 2.1). Imagine further that if a population has n identical individuals, then, from the ecological rules, we can determine a *birth rate* $b(n)$, which gives us a measure of the probability that a new individual will be born and the population size becomes $n + 1$ 'per unit time'. One must be slightly precise about what exactly they mean when they say 'per unit time' since there are no discrete 'time steps' for individuals to be born. Here, by 'birth rate', I mean the probability that there will be a birth (and no death) per an *infinitesimal* amount of time. More formally, letting N_t denote the random variable representing the population size at time t and letting $\mathbb{P}(E)$ denote the probability (in the common-sense usage) of an event E , the birth rate⁶ $b(n)$ of individuals a population with population size n is the quantity

$$b(n) := \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P}(N_{t+\epsilon} = n + 1 | N_t = n) \quad (2.5)$$

⁶Note that unlike usual ecology convention, this is *not* a per-capita birth rate. The ecology per capita birth and death rates can be found by dividing my birth/death rates by the current population size $n(t)$

Exactly analogously, we can also define the *death rate* $d(n)$ of individuals in a population of n individuals as the quantity

$$d(n) := \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P}(N_{t+\epsilon} = n - 1 | N_t = n) \quad (2.6)$$

An alternative, perhaps more intuitive characterization, of these same quantities is the following: If we have a population of size n , and we know that *either a birth or a death* is about to occur, then, the probability that the event that occurs is a birth is

$$\mathbb{P}[\text{ birth } | \text{ something happened }] = \frac{b(n)}{b(n) + d(n)}$$

and the probability that the event is instead a death is

$$\mathbb{P}[\text{ death } | \text{ something happened }] = \frac{d(n)}{b(n) + d(n)}$$

Example 1. Consider the case where the per-capita birth rate is a constant $\lambda > 0$, *i.e.*, $b(n) = \lambda n$, and the per-capita death rate has the linear density-dependence $(\mu + (\lambda - \mu)\frac{n}{K})$, *i.e.* the total death rate is $d(n) = (\mu + (\lambda - \mu)\frac{n}{K})n$, where μ and K are positive constants. Taking the difference between the birth and death rates, we obtain $b(n) - d(n) = (\lambda - \mu)n(1 - \frac{n}{K})$, where, identifying $r = \lambda - \mu$, we obtain the familiar [functional form of the density dependence of the logistic equation on the RHS](#). Note, however, that the population itself is stochastic, whereas the logistic equation is a deterministic description.

Now, let $P(n, t)$ be the probability that the population size is n at time t . We wish to have an equation to describe how $P(n, t)$ changes with time - this will provide a probabilistic description of how we expect the population size to change over time.

To do this, we imagine a large ensemble of populations. In a large ensemble of copies evolving independently, a fraction $P(n, t)$ will have population size n at time t by definition of probability. We can now simply measure the ‘inflow’ and ‘outflow’ of copies of the population from each state (Figure 2.1). If a population has n individuals, it could either have gotten there from a population of $n+1$ individuals, with a death rate of $d(n+1)$, or from a population of $n-1$ individuals, with a birth rate of $b(n-1)$. Thus, the rate of ‘inflow’ to the state n is given by

$$R_{\text{in}}(n, t) = b(n-1)P(n-1, t) + d(n+1)P(n+1, t) \quad (2.7)$$

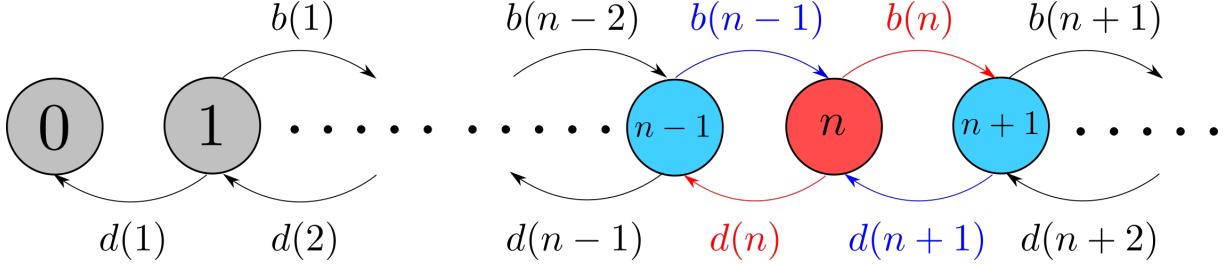


Figure 2.1: Schematic description of a one-dimensional birth-death process. Consider a population of identical individuals. The state of the system can be described by a single number, the population size (numbers within the circles). Births and deaths result in changes in the total population size, and the birth and death rates (arrows) are dependent on the current population size. For a given state n , the blue arrows depict the rate of ‘inflow’ to the state (from the blue states), whereas the red arrows depict the rate of ‘outflow’.

Similarly, if the population has n individuals, it could obtain a different state in two ways: With rate $b(n)$, the population witnesses a birth, and with rate $d(n)$, it witnesses a death. Thus, the rate of ‘outflow’ is given by

$$R_{\text{out}}(n, t) = b(n)P(n, t) + d(n)P(n, t) \quad (2.8)$$

The rate of change of the probability of the system being in state n is given by the rate of inflow minus the rate of outflow. Thus, we have

$$\begin{aligned} \frac{\partial P}{\partial t}(n, t) &= R_{\text{in}}(n, t) - R_{\text{out}}(n, t) \\ &= b(n-1)P(n-1, t) + d(n+1)P(n+1, t) - b(n)P(n, t) - d(n)P(n, t) \end{aligned} \quad (2.9)$$

For convenience, let us define two ‘step operators’ \mathcal{E}^\pm , which act on any functions of populations to their right by either adding or removing an individual, *i.e*

$$\mathcal{E}^\pm f(n, t) = f(n \pm 1, t)$$

Rearranging the RHS of (2.9) to write in terms of these step operators, we obtain the compact expression

$$\frac{\partial P}{\partial t}(n, t) = (\mathcal{E}^- - 1)b(n)P(n, t) + (\mathcal{E}^+ - 1)d(n)P(n, t) \quad (2.10)$$

This is the so-called ‘master equation’, and completely describes our system. However, in general, $b(n)$ and $d(n)$ may be rather complicated, in which case it may not be possible to

solve (2.10) directly.

2.2.2 The system-size expansion

The system-size expansion arises from noting that in many systems, the interactions are governed not by population size, but by population *density*. However, the population jumps themselves are discretized at the scale of the individual, which becomes negligibly small if we have a large population density. Thus, we assume that there exists a system-size parameter $K > 0$ such that the discrete jumps between states happen in units of $1/K$, and we make the substitutions

$$\begin{aligned} x &= \frac{n}{K} \\ b_K(x) &= \frac{1}{K}b(n) \\ d_K(x) &= \frac{1}{K}d(n) \end{aligned}$$

As K grows very large, the discontinuous jumps in n thus appear like ‘continuous’ transitions in our new variable x , which can be thought of as the ‘density’ of organisms. A system-size parameter K often naturally emerges in ecological systems through resource-limiting factors such as habitat size or carrying capacity. Under these substitutions, equation (2.10) becomes

$$\frac{\partial P}{\partial t}(x, t) = (\Delta^- - 1)Kb_K(x)P(x, t) + (\Delta^+ - 1)Kd_K(x)P(x, t) \quad (2.11)$$

where we now have the new step operators

$$\Delta^\pm f(x, t) = f\left(x \pm \frac{1}{K}, t\right) \quad (2.12)$$

If K is large, then we can now taylor-expand the action of these step operators as:

$$\Delta^\pm f(x, t) = f\left(x \pm \frac{1}{K}, t\right) = f(x, t) \pm \frac{1}{K} \frac{\partial f}{\partial x}(x, t) + \frac{1}{2K^2} \frac{\partial^2 f}{\partial x^2}(x, t) + \mathcal{O}(K^{-3})$$

Substituting these expansions into (2.11) and neglecting terms of $\mathcal{O}(K^{-3})$ and higher, we obtain

$$\frac{\partial P}{\partial t}(x, t) = -\frac{\partial}{\partial x}\{A^-(x)P(x, t)\} + \frac{1}{2K} \frac{\partial^2}{\partial x^2}\{A^+(x)P(x, t)\}$$

(2.13)

where

$$A^\pm(x) = b_K(x) \pm d_K(x)$$

Equation (2.13) is a Fokker-Planck equation and corresponds to the SDE:

$$dX_t = A^-(X_t)dt + \sqrt{\frac{A^+(X_t)}{K}}dW_t \quad (2.14)$$

Note that the deterministic component of this process depends on the difference between birth and death rates (a mechanistic measure of Malthusian fitness), whereas the stochastic part depends on their sum (a measure of total turnover rate) and scales inversely with \sqrt{K} and thus vanishes for infinitely large populations. As we will see, this general pattern will turn up repeatedly.

2.2.3 Stochastic fluctuations and the weak noise approximation

If we assume the noise is weak (this will be made precise shortly), then we can go still further with analytic techniques by measuring fluctuations from the deterministic expectations, albeit with some slightly cumbersome calculations to arrive at the final expressions. We will grit our teeth and get through the algebra below, with my promise that the final answer is neat and easy to handle. It is clear that as $K \rightarrow \infty$, equation (2.14) describes a deterministic process, obtained as the solution to

$$\frac{dx}{dt} = A^-(x) = b_K(x) - d_K(x) \quad (2.15)$$

This is a very intuitive equation, saying that the rate of change of the population is equal to the birth rate minus the death rate. Let the solution of this equation be given by $\alpha(t)$, so that $\frac{d\alpha}{dt}(t) = A^-(\alpha(t))$.

We can now measure (scaled) fluctuations from the deterministic solution α through a new variable $y = \sqrt{K}(x - \alpha(t))$. For notational clarity, I will also introduce a new time variable $s = t$ which is equal to the original time variable (this is just so the equations look clearer). Let the probability density function of this new variable be given by $\tilde{P}(y, s)$. In summary, I have introduced the variables:

$$\begin{aligned} y &= \sqrt{K}(x - \alpha(t)) \\ s &= t \end{aligned}$$

$$\tilde{P}(y, s) = \frac{1}{\sqrt{K}} P(x, t)$$

Note that by ordinary rules of variable substitution, we have:

$$\begin{aligned} \frac{\partial \tilde{P}}{\partial t} &= \frac{\partial \tilde{P}}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial \tilde{P}}{\partial s} \frac{\partial s}{\partial t} \\ &= \frac{\partial \tilde{P}}{\partial y} \left(-\sqrt{K} \frac{d\alpha}{dt} \right) + \frac{\partial \tilde{P}}{\partial s} \\ &= -\sqrt{K} A^-(\alpha(s)) \frac{\partial \tilde{P}}{\partial y} + \frac{\partial \tilde{P}}{\partial s} \end{aligned} \quad (2.16)$$

and

$$\frac{\partial}{\partial y} = \frac{1}{\sqrt{K}} \frac{\partial}{\partial x} \quad (2.17)$$

Reformulating (2.13) in terms of y, s and \tilde{P} and substituting (2.16) and (2.17) yields:

$$\begin{aligned} -A^-(\alpha) \frac{\partial \tilde{P}}{\partial x} + \frac{\partial \tilde{P}}{\partial s} &= -\sqrt{K} \frac{\partial}{\partial y} \left(A^-(\alpha + \frac{y}{\sqrt{K}}) \tilde{P} \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left(A^+(\alpha + \frac{y}{\sqrt{K}}) \tilde{P} \right) \\ \Rightarrow \frac{\partial \tilde{P}}{\partial s} &= -\frac{\partial}{\partial y} \left[\sqrt{K} \left(A^-(\alpha + \frac{y}{\sqrt{K}}) - A^-(\alpha) \right) \tilde{P} \right] + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left(A^+(\alpha + \frac{y}{\sqrt{K}}) \tilde{P} \right) \end{aligned} \quad (2.18)$$

We are now ready to make a weak noise ‘expansion’ (This is a special case of a more general idea called a ‘perturbative expansion’ or ‘perturbation theory’ in physics and related fields). We do so by assuming that \tilde{P} , $A^-(\alpha + \frac{y}{\sqrt{K}})$, and $A^+(\alpha + \frac{y}{\sqrt{K}})$ can be approximated by series expansions in $\frac{1}{\sqrt{K}}$ of the form:

$$\begin{aligned} \tilde{P} &= \sum_{n=0}^{\infty} \tilde{P}_n \left(\frac{1}{\sqrt{K}} \right)^n \\ A^- \left(\alpha(s) + \frac{y}{\sqrt{K}} \right) &= \sum_{n=0}^{\infty} A_n^-(s) \left(\frac{y}{\sqrt{K}} \right)^n \\ A^+ \left(\alpha(s) + \frac{y}{\sqrt{K}} \right) &= \sum_{n=0}^{\infty} A_n^+(s) \left(\frac{y}{\sqrt{K}} \right)^n \end{aligned}$$

with $A_0^-(s) = A^-(\alpha(s))$, $A_0^+(s) = A^+(\alpha(s))$. These could be Taylor expansions, for example, but the exact form of the coefficients is irrelevant as long as it is known to us, so any expansion

will work. We can now substitute these series expansions into (2.18) to obtain:

$$\begin{aligned} \sum_{n=0}^{\infty} \left(\frac{1}{\sqrt{K}} \right)^n \frac{\partial \tilde{P}_n}{\partial s} &= -\frac{\partial}{\partial y} \left[\sqrt{K} \left(\sum_{n=1}^{\infty} A_n^-(s) \left(\frac{y}{\sqrt{K}} \right)^n \right) \left(\sum_{m=0}^{\infty} \tilde{P}_m \left(\frac{1}{\sqrt{K}} \right)^m \right) \right] \\ &\quad + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left[\left(\sum_{n=0}^{\infty} A_n^+(s) \left(\frac{y}{\sqrt{K}} \right)^n \right) \left(\sum_{m=0}^{\infty} \tilde{P}_m \left(\frac{1}{\sqrt{K}} \right)^m \right) \right] \end{aligned} \quad (2.19)$$

We can now compare the coefficients of $K^{-n/2}$ for each n in order to arrive at approximations in the series expansion, the idea being that you neglect all terms which are of order greater than $\mathcal{O}(K^{-m/2})$ for some m according to the desired precision.

We observe that for any fixed r , the coefficient of $K^{-r/2}$ on the LHS is $\frac{\partial \tilde{P}_r}{\partial s}$. On the RHS, the coefficients of $K^{-r/2}$ in the second term have the form $\tilde{P}_m A_n^+ y^n$, subject to the constraint that $m + n = r$. Furthermore, all such terms (and only such terms) are coefficients of $K^{-r/2}$. Thus, after grouping, the coefficient of $K^{-r/2}$ from the second terms of the RHS of (2.19) is precisely

$$\frac{1}{2} \frac{\partial^2}{\partial y^2} \sum_{m=0}^r \tilde{P}_m A_{r-m}^+ y^{r-m}$$

Exactly analogous reasoning reveals that the contribution of the first term of the RHS is:

$$-\frac{\partial}{\partial y} \sum_{m=0}^r \tilde{P}_m A_{r-m+1}^- y^{r-m+1}$$

Thus, we find that the r th term of the expansion satisfies:

$$\frac{\partial \tilde{P}_r}{\partial s} = -\frac{\partial}{\partial y} \left(\sum_{m=0}^r \tilde{P}_m A_{r-m+1}^- y^{r-m+1} \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left(\sum_{m=0}^r \tilde{P}_m A_{r-m}^+ y^{r-m} \right) \quad (2.20)$$

We will now make the meaning of ‘weak’ precise. Let us assume that the fluctuations are weak enough that we can obtain a reasonable approximation of the dynamics by retaining only the highest order (in $1/K$) term in equation (2.20) and neglecting all higher-order terms⁷. We are then left with the expression:

$$\frac{\partial \tilde{P}_0}{\partial s} = -A_1^-(s) \frac{\partial}{\partial y} (y \tilde{P}_0) + \frac{A_0^+(s)}{2} \frac{\partial^2 \tilde{P}_0}{\partial y^2} \quad (2.21)$$

⁷For example, we may imagine this is reasonable if the deterministic trajectory is at a stable fixed point and subject to weak fluctuations

which is simply the Fokker-Planck equation for the Itô process

$$dY_t = A_1^-(t)Y_t dt + \sqrt{A_0^+(t)}dW_t$$

This is the ‘weak noise approximation’, (sometimes also called the ‘linear noise approximation’ because the resulting Fokker-Planck equation is linear). This equation describes a so-called ‘Ornstein-Uhlenbeck process’, and is easily solved by using $\exp(-\int A_1^-(s)ds)$ as an ‘integrating factor’. In particular, multiplying both sides by $\exp(-\int A_1^-(s)ds)$ yields

$$\begin{aligned} \exp\left(-\int_0^t A_1^-(s)ds\right) dY_t - Y_t A_1^-(t) \exp\left(-\int_0^t A_1^-(s)ds\right) dt &= \sqrt{A_0^+(t)} \exp\left(-\int_0^t A_1^-(s)ds\right) dW_t \\ \Rightarrow d\left(\exp\left(-\int_0^t A_1^-(s)ds\right) Y_t\right) &= \sqrt{A_0^+(t)} \exp\left(-\int_0^t A_1^-(s)ds\right) dW_t \end{aligned}$$

Integrating both sides and noting that $A_0^+(s) = A^+(\alpha(s))$, we thus obtain the final expression

$$Y_t = Y_0 \exp\left(\int_0^t A_1^-(s)ds\right) + \int_0^t \exp\left(-\int_s^t A_1^-(v)dv\right) \sqrt{A^+(\alpha(s))} dW_s \quad (2.22)$$

as the zeroth-order weak noise approximation for stochastic fluctuations from the deterministic trajectory due to demographic noise. ~~Note that this is an exact equation~~ This equation can be exactly solved via analytical techniques and one can get many insights from the solution. For example, if $Y_0 = 0$ (*i.e.* we start at the deterministic steady state, a natural assumption for measuring fluctuations from it), then we can show by taking expectations in (2.22) and using results presented in 2.1.2 that we must have $\mathbb{E}[Y_t|Y_0] = 0$. In other words, the fluctuations have zero expectation and are expected to occur symmetrically about $\alpha(t)$, with no bias. The variance (spread) of the fluctuations Y_t , as well as higher moments, can also be exactly calculated from (2.22) using some tools from stochastic calculus, but I will not demonstrate this here.

Importantly, higher order terms do not form Fokker-Planck equations, and in general, \tilde{P}_r for $r > 0$ may be negative and therefore does not even describe a probability. As such, formulating the solution as the solution to an SDE only works for \tilde{P}_0 . If noise is large enough that it is not well-approximated by \tilde{P}_0 , this approximation method is not very useful.

Chapter 3

Population dynamics from stochastic first principles

Somewhere [...] between the specific that has no meaning and the general that has no content there must be, for each purpose and at each level of abstraction, an optimum degree of generality

Kenneth Boulding (1956)

Let us now consider the situation we are actually interested in. Assume that our population is *not* composed of identical organisms, but instead can contain up to m different kinds of organisms - for example, individuals may come in one of m colors, or a gene may have m different alleles. The specific interpretation of the different variants is irrelevant to our formalism, and I, therefore, refer to each distinct variant of an organism simply as a ‘type’. Unlike many classic stochastic formulations in evolutionary theory (Fisher, 1930; S. Wright, 1931; Moran, 1958; Kimura, 1957; Kimura, 1964; Kimura and Crow, 1964; Crow and Kimura, 1970; Lande, 1976), I do not assume a fixed (effective) population size and instead allow the total population size to fluctuate naturally over time.

3.1 Description of the process and the Master Equation

Given a population that can contain up to m different (fixed) kinds of organisms, it can be entirely characterized by specifying the number of organisms of each type (Figure 3.1A). Thus, the state of the population at a given time t is an m -dimensional *vector* of the form $\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_m(t)]^T$, where $n_i(t)$ is the number of individuals of type i in the population at time t .

Given a state $\mathbf{n}(t)$, we also need to describe how this vector can change over time due to births and deaths (ecology). In this case, a birth or death could result in an individual belonging to one of m different types. Thus, whereas before we had two functions $b(n)$ and $d(n)$ which take in a number as an input, we now require $2m$ functions that take in a vector as an input (Figure 3.1B). In other words, for each type $i \in \{1, 2, \dots, m\}$, we must specify a birth rate $b_i(\mathbf{n})$ and a death rate $d_i(\mathbf{n})$. By ‘rates’, I mean that if we know that *either a birth or a death* occurs, then the probability that this event is the birth of an individual of type i is given by

$$\mathbb{P}[\text{Birth of a type } i \text{ individual} \mid \text{something happened}] = \frac{b_i(\mathbf{n})}{\sum_{j=1}^m (b_j(\mathbf{n}) + d_j(\mathbf{n}))}$$

and the probability that the event is the death of an individual of type i is

$$\mathbb{P}[\text{Death of a type } i \text{ individual} \mid \text{something happened}] = \frac{d_i(\mathbf{n})}{\sum_{j=1}^m (b_j(\mathbf{n}) + d_j(\mathbf{n}))}$$

As before, we can describe the rate of change of $P(\mathbf{n}, t)$, the probability of finding the population in a state \mathbf{n} at time t , by measuring the inflow and outflow rates. Since the population changes in units of exactly one individual (by definition of a birth-death process; see section 2.1), we know that these inflow and outflow rates must only involve populations that are a single individual away from our focal population. In other words, for a population $\mathbf{n} = [n_1, \dots, n_m]^T$, the ‘inflow’ is from all populations of the form $[n_1, \dots, n_i - 1, \dots, n_m]^T$ through a birth of a type i individual, and from all populations of the form $[n_1, \dots, n_i + 1, \dots, n_m]^T$

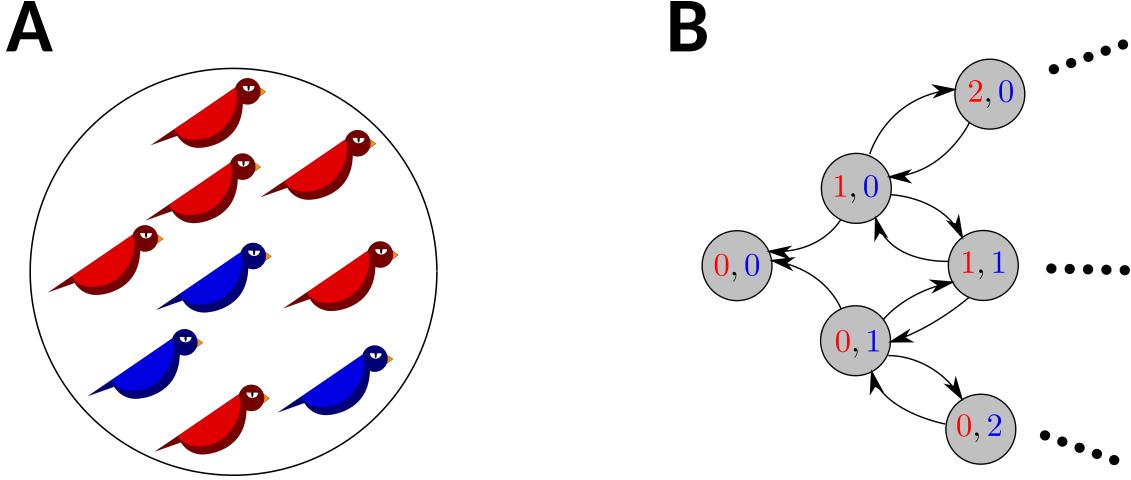


Figure 3.1: Schematic description of a multi-dimensional birth-death process. (A) Consider a population of birds in which individuals are either red or blue. In this case, we have $m = 2$, since there are two types of individuals in the population. (B) The state of the system can be described by a vector containing the number of individuals of each discrete type (in this case, the number of red and blue birds in the population). Births and deaths result in changes in the elements of the state vector.

through the death of a type i individual. Thus, we have the inflow rate

$$\begin{aligned} R_{\text{in}}(\mathbf{n}, t) &= \sum_{j=1}^m b_j([n_1, \dots, n_j - 1, \dots, n_m]^T) P([n_1, \dots, n_j - 1, \dots, n_m]^T, t) \\ &\quad + \sum_{j=1}^m d_j([n_1, \dots, n_j + 1, \dots, n_m]^T) P([n_1, \dots, n_j + 1, \dots, n_m]^T, t) \end{aligned} \quad (3.1)$$

Outflow is through births and deaths of individuals in the population \mathbf{n} itself, and thus we have:

$$R_{\text{out}}(\mathbf{n}, t) = \sum_{j=1}^m b_j(\mathbf{n}) P(\mathbf{n}, t) + \sum_{j=1}^m d_j(\mathbf{n}) P(\mathbf{n}, t) \quad (3.2)$$

We will now define step operators, both for notational ease and in anticipation of the system size expansion. For each $i \in \{1, \dots, m\}$, let us define two step operators \mathcal{E}_i^\pm by their action on any function $f([n_1, \dots, n_m], t)$ as:

$$\mathcal{E}_i^\pm f([n_1, \dots, n_i, \dots, n_m]^T, t) = f([n_1, \dots, n_i \pm 1, \dots, n_m]^T, t) \quad (3.3)$$

In other words, \mathcal{E}_i^\pm just changes the population through the addition or removal of one type i

individual. We can now the rate of change of $P(\mathbf{n}, t)$ as

$$\frac{\partial P}{\partial t}(\mathbf{n}, t) = R_{\text{in}}(\mathbf{n}, t) - R_{\text{out}}(\mathbf{n}, t) \quad (3.4)$$

Substituting (3.1), (3.2), and (3.3) into equation (3.4), we obtain:

$$\frac{\partial P}{\partial t}(\mathbf{n}, t) = \sum_{j=1}^m [(\mathcal{E}_j^- - 1)b_j(\mathbf{n})P(\mathbf{n}, t) + (\mathcal{E}_j^+ - 1)d_j(\mathbf{n})P(\mathbf{n}, t)] \quad (3.5)$$

This is the master equation of our m -dimensional process.

3.2 The system-size expansion

As explained in Section 2.1.3, we will now assume (on ecological grounds) that there exists a system-size parameter $K > 0$ such that the discrete jumps between states happen in units of $1/K$ and the total population size is controlled by K , with $K = \infty$ corresponding to an infinitely large population. In particular, we assume that the birth and death rates scale such that we can make the substitutions

$$\begin{aligned} \mathbf{x} &\coloneqq \frac{\mathbf{n}}{K} \\ b_i^{(K)}(\mathbf{x}) &\coloneqq \frac{1}{K}b_i(\mathbf{n}) \\ d_i^{(K)}(\mathbf{x}) &\coloneqq \frac{1}{K}d_i(\mathbf{n}) \end{aligned}$$

where $\mathbf{x} = \mathbf{n}/K$ measures population *density* instead of population numbers. We now define new step operators Δ_i^\pm by their action on any real-valued function $f(\mathbf{x}, t)$ as

$$\Delta_i^\pm f([x_1, \dots, x_m]^T, t) = f([x_1, \dots, x_i \pm \frac{1}{K}, \dots, x_m]^T, t) \quad (3.6)$$

In terms of these new variables, (3.5) becomes

$$\frac{\partial P}{\partial t}(\mathbf{x}, t) = K \sum_{j=1}^m [(\Delta_j^- - 1)b_j^{(K)}(\mathbf{x})P(\mathbf{x}, t) + (\Delta_j^+ - 1)d_j^{(K)}(\mathbf{x})P(\mathbf{x}, t)] \quad (3.7)$$

If K is large, we can once again Taylor expand the action of the step operators as

$$f([x_1, \dots, x_i \pm \frac{1}{K}, \dots, x_m]^T, t) = f(\mathbf{x}, t) \pm \frac{1}{K} \frac{\partial f}{\partial x_i}(\mathbf{x}, t) + \frac{1}{2K^2} \frac{\partial^2 f}{\partial x_i^2}(\mathbf{x}, t) + \mathcal{O}(K^{-3})$$

which, after substituting into (3.7) and neglecting $\mathcal{O}(K^{-3})$ terms, yields the equation

$$\frac{\partial P}{\partial t}(\mathbf{x}, t) = \sum_{j=1}^m \left[-\frac{\partial}{\partial x_j} \{A_j^-(\mathbf{x})P(\mathbf{x}, t)\} + \frac{1}{2K} \frac{\partial^2}{\partial x_j^2} \{A_j^+(\mathbf{x})P(\mathbf{x}, t)\} \right] \quad (3.8)$$

where

$$A_i^\pm(\mathbf{x}) = b_i^{(K)}(\mathbf{x}) \pm d_i^{(K)}(\mathbf{x})$$

Equation (3.8) is an m -dimensional Fokker-Planck equation, and corresponds to the m -dimensional Itô process

$$d\mathbf{X}_t = \mathbf{A}^-(\mathbf{X}_t)dt + \frac{1}{\sqrt{K}} \mathbf{D}(\mathbf{X}_t)d\mathbf{W}_t \quad (3.9)$$

where $\mathbf{A}^-(\mathbf{X}_t)$ is an m -dimensional vector with i^{th} element $= A_i^-(\mathbf{X}_t)$. $\mathbf{D}(\mathbf{X}_t)$ is an $m \times m$ matrix with ij^{th} element $(\mathbf{D}(\mathbf{X}_t))_{ij} = \delta_{ij} (A_i^+ A_j^+)^{\frac{1}{4}}$, where δ_{ij} is the Kronecker delta symbol, defined by

$$\delta_{ij} = \begin{cases} 1 & i = j \\ 0 & i \neq j \end{cases}$$

Finally, \mathbf{W}_t is the m -dimensional Wiener process and can be thought of as a vector of independent one-dimensional Wiener processes (which have been defined in 2.1.2). This is the ‘mesoscopic’ description of our process.

3.3 Functional forms of the birth and death rates

I assume that the birth and death rate functions have the functional form

$$\begin{aligned} b_i^{(K)}(\mathbf{x}) &= x_i b_i^{(\text{ind})}(\mathbf{x}) + \mu Q_i(\mathbf{x}) \\ d_i^{(K)}(\mathbf{x}) &= x_i d_i^{(\text{ind})}(\mathbf{x}) \end{aligned} \quad (3.10)$$

where $b_i^{(\text{ind})}(\mathbf{x})$ and $d_i^{(\text{ind})}(\mathbf{x})$ are non-negative functions that respectively describe the per-capita birth and death rate of type i individuals, $\mu \geq 0$ is a constant describing the mutation rate in the population, and $Q_i(\mathbf{x})$ is a non-negative function that describes the additional birth rate

of type i individuals due to mutations in the population \mathbf{x} that cannot be captured in the per-capita birth rate¹. My assumptions on the functional forms (3.10) thus amount to saying that birth and death rates can be separated into mutational and non-mutational components, and furthermore that the density dependence of the birth and death rates due to non-mutational effects is in a form that allows us to write down per-capita birth and death rates for each type. Let us define the *Malthusian fitness* of the i^{th} type as $w_i(\mathbf{x}) := b_i^{(\text{ind})}(\mathbf{x}) - d_i^{(\text{ind})}(\mathbf{x})$, and the *per-capita turnover rate* of the i^{th} type as $\tau_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) + d_i^{(\text{ind})}(\mathbf{x})$. The quantity $w_i(\mathbf{x})$ describes the per-capita growth rate of type i individuals in a population \mathbf{x} discounting mutation. Ecologists often denote this quantity by the symbol r_i and simply call it the (exponential) growth rate of type i , but I will stick to w_i and ‘fitness’ here. τ_i is a measure of the number of events (birth events + death events) that a type i individual experiences in a given time interval — [individuals](#) [Populations of types](#) with higher turnover rates experience more events (on average) than those with lower turnover rates. This can be thought of as a measure of the ‘pace of life’ [of a type](#). I briefly note that the quantity τ_i has also been called the ‘variability in the reproductive output’ in the literature (Gillespie, 1974). It is notable that both w_i and τ_i depend on the state of the population as a whole (*i.e.* \mathbf{x}) and not just on the density of the focal type. Thus, in general, the fitness and the turnover rate in our model are both frequency and density-dependent.

3.4 Statistical measures for population-level quantities

Though the causes of evolution are generally described in terms of ecological phenomena affecting birth, death, and interactions, all of which operate at the *individual* level like we have been working with, evolution itself is typically described at a *population* level, in terms of type frequencies (Bourrat, 2019). We are also often interested in describing the effect of evolutionary forces on population-level quantities, such as the mean fitness or the mean phenotype in the population. Furthermore, the relevant quantities at the individual level, such as individual fitness or phenotype, are typically equal for all individuals of the same type (in some sense this is our basis for defining different types in the first place). I use the term ‘type-level quantities’ henceforth to refer to such quantities that are equal for all individuals that are of the same type. To facilitate the description of such quantities, given

¹When $x_i = 0$, *i.e.* there are no type i individuals in the population, individuals of type i may still be born through mutations during births of the other types. This cannot be captured in $b_i^{(\text{ind})}(\mathbf{x})$ because the term $x_i b_i^{(\text{ind})}(\mathbf{x})$ vanishes when $x_i = 0$. Note that no analogous problem exists for the death rate, since the death rate of type i individuals must be 0 when x_i is 0 to ensure that we never have negative population densities.

any state $\mathbf{x}(t)$ that describes our population at time t , let us first define the total (scaled) population size and the frequency of each type in the population as:

$$\begin{aligned} N_K(t) &:= \sum_{i=1}^m x_i(t) = \frac{1}{K} \sum_{i=1}^m n_i(t) \\ p_i(t) &:= \frac{x_i(t)}{N_K(t)} = \frac{n_i(t)}{\sum_{j=1}^m n_j(t)} \end{aligned} \quad (3.11)$$

Now, let f be any type-level quantity with (possibly time-dependent) value $f_i \in \mathbb{R}$ for the i^{th} type. For example, if each type is a phenotype for a trait such as height, which can be assigned a numerical value, then setting $f_i = \text{value of } i^{\text{th}} \text{ phenotype}$ gives us the mean trait value in the population. We can compute the statistical mean value of any such quantity in the population as

$$\bar{f}(t) := \sum_{i=1}^m f_i p_i \quad (3.12)$$

the statistical covariance of two such quantities f and g as

$$\text{Cov}(f, g) := \overline{fg} - \bar{f}\bar{g} \quad (3.13)$$

and the statistical variance of a quantity f as $\sigma_f^2 := \text{Cov}(f, f)$.

It is important to recognize that these quantities are distinct from and independent of the *probabilistic* expectation, variance, and covariance obtained by integrating over realizations in the underlying probability space. Indeed, for finite populations, the statistical mean, statistical variance, and statistical covariance are all themselves stochastic processes: For each instant of time, these population-level quantities are a random variable (*i.e.* a *function* and not just a number) depending on \mathbf{p} , the (random) vector of type frequencies in the population. For infinite populations, the statistical mean, variance, and covariance are entirely deterministic time-dependent quantities that simply describe how f is distributed across the population. Previous authors, such as Frank (Frank, 1997; Frank, 2012) and van Veelen (Van Veelen, 2005; Van Veelen, 2020), have pointed out that failure to clearly make this distinction between statistical operations and probabilistic operations has led to much confusion in the population biology literature with regard to the infinite population Price equation, which is entirely deterministic and does not incorporate features of finite populations such as drift in its original formulation (Frank, 1997; Van Veelen, 2005; Frank,

2012).

3.5 Stochastic Trait Frequency Dynamics

In appendix B, I use a multivariate version of Itô's formula to derive a general stochastic equation for the frequencies of each type in the population. Letting $\bar{w} = \sum w_i p_i$ and $\bar{\tau} = \sum \tau_i p_i$ be the average population fitness and turnover respectively, I show in appendix B that the frequency of the i^{th} type in the population $\mathbf{x}(t)$ changes according to the equation:

$$\begin{aligned} dp_i(t) = & \left[(w_i(\mathbf{x}) - \bar{w})p_i + \mu \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right\} \right] dt \\ & - \frac{1}{K} \frac{1}{N_K(t)} \left[(\tau_i(\mathbf{x}) - \bar{\tau})p_i + \mu \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right\} \right] dt \\ & + \frac{1}{\sqrt{K}} \frac{1}{N_K(t)} \left[(A_i^+(\mathbf{x}))^{1/2} dW_t^{(i)} - p_i \sum_{j=1}^m (A_j^+(\mathbf{x}))^{1/2} dW_t^{(j)} \right] \end{aligned} \quad (3.14)$$

where $W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(m)}$ are m independent one-dimensional Wiener processes and I have used the notation $Q_i(\mathbf{p}) = Q_i(\mathbf{x})/N_K(t)$ for notational clarity. I will show below that the first term in this expression describes directional changes in the population composition due to ‘classical’ evolutionary forces such as selection and mutation that occur in deterministic infinite population models. The second term is an additional directional force on population composition that is only seen in finite populations and can be thought of as a biasing ‘selection’ for reduced turnover rate due to an effect similar to gambler’s ruin in probability theory. The consequences of this term, as well as connections with previous studies, are discussed in detail in part III. Finally, the last term of equation (3.14) describes non-directional stochastic effects due to fluctuations and has a ‘spreading effect’ (Also see equation (E.7) in Appendix E for a more elegant representation of these noise terms as an integral against a single Wiener process). For the case $m = 2, \mu = 0$ (two interacting types of individuals with no mutations between types), after using the representation of noise terms presented in Appendix E and letting $p = p_1$, equation (3.14) becomes the considerably simpler expression

$$\begin{aligned} dp = & \left[(w_1 - w_2)p(1-p) - \frac{1}{KN_K}(\tau_1 - \tau_2)p(1-p) \right] dt \\ & + \frac{1}{\sqrt{KN_K}} (p(1-p)[\tau_1 + (\tau_2 - \tau_1)p])^{1/2} dW_t \end{aligned} \quad (3.15)$$

In this equation, if we start with an initial condition $p_0 \in [0, 1]$, the system defined by equation (3.15) will always remain in $[0, 1]$, and in that sense, it is ‘well-behaved’. To see this, one can observe that the RHS of equation (3.15) identically vanishes at both $p = 0$ and $p = 1$, since every term on the RHS contains the product $p(1 - p)$. Further, since there is no mutation between types, $p = 0$ (extinction of type 1 organisms) and $p = 1$ (extinction of type 2 organisms) are both ‘absorbing states’, *i.e.* a system which reaches these states can never leave them. Well-behavedness at the boundaries for the more general equation (3.14) can be checked directly, and has been carried out in Appendix F.

3.6 The infinite population limit

Like in 2.2, we can once again take $K \rightarrow \infty$ in (3.9) to obtain a deterministic expression. Here, the expression reads

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}^-(\mathbf{x}) = \mathbf{b}^{(K)}(\mathbf{x}) - \mathbf{d}^{(K)}(\mathbf{x}) \quad (3.16)$$

where the m -dimensional vector-valued functions $\mathbf{b}^{(K)}(\mathbf{x})$ and $\mathbf{d}^{(K)}(\mathbf{x})$ on the RHS are defined as having i^{th} element $b_i^{(K)}(\mathbf{x})$ and $d_i^{(K)}(\mathbf{x})$ respectively. For the trait frequencies, by taking $K \rightarrow \infty$ in (3.14), we obtain a deterministic equation that reads:

$$\frac{dp_i}{dt} = (w_i(\mathbf{x}) - \bar{w})p_i + \mu \left[Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right] \quad (3.17)$$

The first term of (3.17) describes changes due to faithful (non-mutational) replication, and the second describes changes due to mutation. For this reason, equation (3.17) is called the *replicator-mutator equation* in the evolutionary game theory literature, where the individual ‘types’ are interpreted to be pure strategies (Hofbauer and Sigmund, 1998; Page and Nowak, 2002; Cressman and Tao, 2014). If in addition, each $w_i(\mathbf{x})$ is linear in \mathbf{x} , meaning we can write $w_i(\mathbf{x}) = \sum_j a_{ij}x_j$ for some set of constants a_{ij} , then we recover the more well-known replicator-mutator equation for matrix games in which the constants a_{ij} form the ‘payoff matrix’ (See the example presented in Appendix D). As is well-known, the replicator equation (without mutation) for matrix games with m pure strategies is equivalent to the generalized Lotka-Volterra equations for a community with $m - 1$ species (Hofbauer and Sigmund, 1998), providing the connection to community ecology. Equation (3.17) is also equivalent to Eigen’s *quasispecies equation* from molecular evolution if each ‘type’ is interpreted as a genetic

sequence and each $w_i(\mathbf{x})$ is a constant function² (Page and Nowak, 2002). We can now calculate how the mean of any ‘type level’ quantity f , defined as f_i for the i^{th} type, changes in the population (For example, if each type is a phenotype for a trait such as height, which can be assigned a numerical value, then setting $f_i = \text{value of } i^{\text{th}} \text{ phenotype}$ gives us the mean trait value in the population). The product rule of calculus tells us that we have the relation

$$\frac{d}{dt} \left(\sum_{i=1}^m f_i p_i \right) = \sum_{i=1}^m \left(f_i \frac{\partial p_i}{\partial t} + p_i \frac{\partial f_i}{\partial t} \right) = \sum_{i=1}^m f_i \frac{\partial p_i}{\partial t} + \overline{\left(\frac{\partial f}{\partial t} \right)} \quad (3.18)$$

Multiplying both sides of equation (3.17) by f_i and summing over all i , we obtain

$$\begin{aligned} \sum_{i=1}^m f_i \frac{\partial p_i}{\partial t} &= \sum_{i=1}^m f_i w_i(\mathbf{x}) p_i - \bar{w} \sum_{i=1}^m f_i p_i + \mu \left[\sum_{i=1}^m Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^m Q_j(\mathbf{p}) \sum_{i=1}^m p_i f_i \right) \right] \\ \Rightarrow \frac{d\bar{f}}{dt} &= \bar{w}\bar{f} - (\bar{w})(\bar{f}) + \mu \left[\sum_{i=1}^m Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \bar{f} \right] \end{aligned}$$

Using the definition of statistical covariance from (3.13), we obtain

$$\sum_{i=1}^m f_i \frac{\partial p_i}{\partial t} = \text{Cov}(w, f) + \mu \left[\sum_{i=1}^m Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \bar{f} \right] \quad (3.19)$$

Thus, substituting this into (3.18), we get

$$\frac{d\bar{f}}{dt} = \text{Cov}(w, f) + \mu \left[\sum_{i=1}^m Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \bar{f} \right] + \overline{\left(\frac{\partial f}{\partial t} \right)} \quad (3.20)$$

This is a Price equation for quantities f_i which can vary over time (Lion, 2018; Day et al., 2020). To obtain the more familiar Price equation seen in textbooks, we can consider time-independent f_i , *i.e.* situations in which each f_i is constant over time, and thus changes in \bar{f} are purely due to changes in the composition of the population. For such quantities, we have $\frac{\partial f_i}{\partial t} = 0 \forall i$ and thus obtain the dynamic version of the famous Price equation (Page

²Mutational effects are often additionally assumed to act through direct ‘transmission probabilities’ of mutating from one type to another. This means that we can write $Q_i(\mathbf{p}) = \sum_j Q_{ij} p_j$, where $Q_{ii} = 0$, and for each $j \neq i$, $Q_{ij} \geq 0$ is a constant describing the probability of a $j \rightarrow i$ mutation (conditioned on the occurrence of a mutation). Substituting this into (3.17) yields an equation in terms of ‘ Q -matrices’ or ‘mutation matrices’ that may be more familiar to some.

and Nowak, 2002; Lion, 2018):

$$\boxed{\frac{d\bar{f}}{dt} = \text{Cov}(w, f) + \mu \left[\sum_{i=1}^m Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \bar{f} \right]} \quad (3.21)$$

The first term of the RHS describes the statistical covariance between the quantity f and the fitness w . The second term describes ‘transmission bias’ due to mutational effects - the first summation is the ‘inflow’ of f due to mutations, and the second is the ‘outflow’.

3.7 Stochastic fluctuations and the weak noise approximation

As in the one-dimensional case, we can go a little further if the noise is sufficiently weak. Let the deterministic trajectory obtained by solving (3.16) be given by $\boldsymbol{\alpha}(t) = [\alpha_1(t), \alpha_2(t), \dots, \alpha_m(t)]^T$. We can once again track stochastic fluctuations from the deterministic trajectory by introducing the new variables

$$\begin{aligned} \mathbf{y} &= \sqrt{K}(\mathbf{x} - \boldsymbol{\alpha}(t)) \\ s &= t \\ \tilde{P}(\mathbf{y}, s) &= \frac{1}{\sqrt{K}} P(\mathbf{x}, t) \end{aligned} \quad (3.22)$$

Then, after some algebra that follows the exact same steps as in section 2.2.3 and retaining only the highest order terms in \sqrt{K} , we obtain the equation:

$$\frac{\partial \tilde{P}_0}{\partial s}(\mathbf{y}, s) = \sum_{j=1}^m \left(-\frac{\partial}{\partial y_j} \left\{ (A_j^-)_1(s) \tilde{P}_0(\mathbf{y}, s) \right\} + \frac{1}{2} A_j^+(\boldsymbol{\alpha}(s)) \frac{\partial^2}{\partial y_j^2} \{ \tilde{P}_0(\mathbf{y}, s) \} \right) \quad (3.23)$$

where $(A_j^-)_1(s)$ is the $\mathcal{O}(1/\sqrt{K})$ term of the power series expansion

$$A_j^-(\boldsymbol{\alpha} + \frac{\mathbf{y}}{\sqrt{K}}) = \sum_{n=0}^{\infty} (A_j^-)_n(s) \left(\frac{\mathbf{y}}{\sqrt{K}} \right)^n$$

In the case where the series expansion is a Taylor expansion, then the first-order term of this expansion is given by

$$(A_j^-)_1(s) = \sum_{i=1}^m y_i \left(\frac{\partial A_j^-(\mathbf{x})}{\partial x_i} \Big|_{\mathbf{x}=\alpha(s)} \right) \quad (3.24)$$

In multi-variable calculus, the directional derivative³ $D_{\mathbf{v}}(f(\mathbf{x}))$ of a multidimensional function $f : \mathbb{R}^m \rightarrow \mathbb{R}$ along a vector \mathbf{v} is the function defined by:

$$D_{\mathbf{v}}(f(\mathbf{x})) := \sum_{i=1}^m \left(\frac{\partial f(\mathbf{x})}{\partial x_i} \right) v_i = \lim_{h \rightarrow 0} \frac{f(\mathbf{x} + h\mathbf{v}) - f(\mathbf{x})}{h} \quad (3.25)$$

Comparing with (3.24), we see that the weak-noise approximation of our process is:

$$\frac{\partial P}{\partial t}(\mathbf{y}, t) = \sum_{j=1}^m \left(-\frac{\partial}{\partial y_j} \{ D_{\mathbf{y}}(A_j^-(\boldsymbol{\alpha}))(t) P(\mathbf{y}, t) \} + \frac{1}{2} A_j^+(\boldsymbol{\alpha}(t)) \frac{\partial^2}{\partial y_j^2} \{ P(\mathbf{y}, t) \} \right) \quad (3.26)$$

where we have dropped the tildes and gone back from s to t for notational clarity. The directional derivative of the population turnover rate A_j^- ‘in the direction’ of the stochastic fluctuation \mathbf{y} at the deterministic point $\boldsymbol{\alpha}(s)$ is precisely the multidimensional analogue of the derivative we had in (2.21). The meaning of equation (3.26) is clearer if we compute how the moments of the fluctuation y_i in the density of type i individuals (for some i) change over time. Let $n > 0$. We have:

$$\frac{d}{dt} \mathbb{E}[y_i^n] = \frac{d}{dt} \int_{\mathbb{R}^m} y_i^n P(\mathbf{y}, t) d\mathbf{y} \quad (3.27)$$

$$= \int_{\mathbb{R}^m} y_i^n \frac{\partial P}{\partial t}(\mathbf{y}, t) d\mathbf{y} \quad (3.28)$$

where I have assumed that y_i^n and $P(\mathbf{y}, t)$ vary sufficiently smoothly to allow us to interchange the order of derivatives and integrals. By the Leibniz integral rule, this only requires the map $(\mathbf{y}, t) \rightarrow y_i^n P(\mathbf{y}, t)$ to be bounded and C^1 in an open subset of $\mathbb{R}^m \times [0, \infty)$. We have also used the notation $\int_{\mathbb{R}^m} f(\mathbf{y}) d\mathbf{y} := \int_{\mathbb{R}} \int_{\mathbb{R}} \dots \int_{\mathbb{R}} f(\mathbf{y}) dy_1 dy_2 \dots dy_m$. The one-dimensional integrals are over the entire real line and not just over $[0, \infty)$ because fluctuations can be either positive (greater than $\boldsymbol{\alpha}(t)$) or negative (lesser than $\boldsymbol{\alpha}(t)$). For notational brevity, let us use the shorthand $D_j := D_{\mathbf{y}}(A_j^-(\boldsymbol{\alpha}))(t)$. We can now substitute (3.26) into (3.28) to

³Some authors use the notation $\partial_{\mathbf{v}} f(\mathbf{x})$ or $\mathbf{v} \cdot \nabla f(\mathbf{x})$ for this object.

obtain

$$\frac{d}{dt} \mathbb{E}[y_i^n] = \int_{\mathbb{R}^m} y_i^n \left(\sum_{j=1}^m \left(-\frac{\partial}{\partial y_j} \{D_j P(\mathbf{y}, t)\} + \frac{1}{2} A_j^+(\boldsymbol{\alpha}(t)) \frac{\partial^2}{\partial y_j^2} \{P(\mathbf{y}, t)\} \right) \right) d\mathbf{y} \quad (3.29)$$

$$= \sum_{j=1}^m \left[- \int_{\mathbb{R}^m} y_i^n \frac{\partial}{\partial y_j} \{D_j P(\mathbf{y}, t)\} d\mathbf{y} + \frac{A_j^+(\boldsymbol{\alpha}(t))}{2} \int_{\mathbb{R}^m} y_i^n \frac{\partial^2}{\partial y_j^2} \{P(\mathbf{y}, t)\} d\mathbf{y} \right] \quad (3.30)$$

We will evaluate the integrals on the RHS of (3.30) using integration by parts. Recall that for any two functions u and v defined on a domain Ω , the general formula for integration by parts is given by:

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v d\mathbf{x} = - \int_{\Omega} u \frac{\partial v}{\partial x_i} d\mathbf{x} + \int_{\partial\Omega} uv \gamma_i dS(\mathbf{x}) \quad (3.31)$$

where $\partial\Omega$ is the boundary of Ω , dS is the surface element of this boundary, and γ_i is the i^{th} component of the unit outward normal to the boundary. In our case, we have $\Omega = \mathbb{R}^m$, and thus the boundary conditions are evaluated as $\|y\| \rightarrow \infty$. I assume that the magnitude of stochastic fluctuations is bounded, and therefore impose the condition $\lim_{\|y\| \rightarrow \infty} P(\mathbf{y}, t) = 0$.

Further, I assume that this decay is fast enough that $\lim_{\|y\| \rightarrow \infty} D_j P(\mathbf{y}, t) = 0 \forall j$. Under these conditions, we can evaluate the two integrals in the RHS of (3.30) by using integration by parts and discarding the boundary term (The second term on the RHS of (3.31)). Note that since the y_i s are orthogonal to each other, we have the relation:

$$\frac{\partial y_i^n}{\partial y_j} = \delta_{ij} n y_i^{n-1}$$

Using this relation and then using integration by parts on the RHS of (3.30) (once for the first term and twice for the second term), we obtain the considerably simpler expression

$$\frac{d}{dt} \mathbb{E}[y_i^n] = n \int_{\mathbb{R}^m} y_i^{n-1} D_i P(\mathbf{y}, t) d\mathbf{y} + \frac{n(n-1)}{2} A_i^+(\boldsymbol{\alpha}(t)) \int_{\mathbb{R}^m} y_i^{n-2} P(\mathbf{y}, t) d\mathbf{y} \quad (3.32)$$

$$\Rightarrow \frac{d}{dt} \mathbb{E}[y_i^n] = n \mathbb{E}[y_i^{n-1} D_i] + \frac{n(n-1)}{2} A_i^+(\boldsymbol{\alpha}(t)) \mathbb{E}[y_i^{n-2}] \quad (3.33)$$

Of particular interest are the cases $n = 1$ (corresponding to the expected value of y_i) and $n = 2$ (which can be used along with the expected value to compute the variance of y_i). We

have:

$$\frac{d}{dt} \mathbb{E}[y_i] = \mathbb{E}[D_i] \quad (3.34)$$

$$\frac{d}{dt} \mathbb{E}[y_i^2] = 2\mathbb{E}[y_i D_i] + A_i^+(\boldsymbol{\alpha}(t)) = 2\langle y_i, D_i \rangle + 2\mathbb{E}[y_i]\mathbb{E}[D_i] + A_i^+(\boldsymbol{\alpha}(t)) \quad (3.35)$$

Where $\langle X, Y \rangle$ is the *probability* covariance between two random variables X and Y , defined as $\langle X, Y \rangle := \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$. This is not to be confused with the *statistical* covariance defined by (3.13) that appears in the deterministic Price equation (3.21). Thus, whether stochastic fluctuations are expected to grow or decay is controlled by D_i , a measure of how the growth rate ($b_i - d_i$) changes along the direction of the fluctuation, whereas the spread of the fluctuations (the variance) has contributions from the net turnover rate ($A_i^+ = b_i + d_i$) and the (probability) covariance between the fluctuation and D_i . In the case of the functional forms given by (3.10), we have:

$$A_i^-(\mathbf{x}) = w_i(\mathbf{x})x_i + \mu Q_i(\mathbf{x}) \quad (3.36)$$

and thus, from (3.24), we can calculate the directional derivative D_i as

$$D_i = \sum_{k=1}^m y_k \left(\frac{\partial A_i^-(\mathbf{x})}{\partial x_k} \Big|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) \quad (3.37)$$

$$= \sum_{k=1}^m y_k \left(\frac{\partial}{\partial x_k} (w_i(\mathbf{x})x_i + \mu Q_i(\mathbf{x})) \Big|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) \quad (3.38)$$

$$= \sum_{k=1}^m y_k \left(\alpha_i \frac{\partial w_i}{\partial x_k} \Big|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) + y_i w_i(\boldsymbol{\alpha}) + \mu \sum_{k=1}^m y_k \left(\frac{\partial Q_i}{\partial x_k}(\mathbf{x}) \Big|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) \quad (3.39)$$

$$= y_i w_i(\boldsymbol{\alpha}) + \alpha_i D_{\mathbf{y}}(w_i(\boldsymbol{\alpha})) + \mu D_{\mathbf{y}}(Q_i(\boldsymbol{\alpha})) \quad (3.40)$$

Using this in (3.34), we see that the expected change of a fluctuation in the density of type i individuals evolves as:

$$\frac{d}{dt} \mathbb{E}[y_i] = \underbrace{w_i(\boldsymbol{\alpha}) \mathbb{E}[y_i]}_{\begin{array}{l} \text{Current fitness of type } i \\ \text{at deterministic trajectory } \boldsymbol{\alpha} \\ (\text{scaled by expected fluctuation } \mathbb{E}[y_i]) \end{array}} + \underbrace{\alpha_i \mathbb{E}[D_{\mathbf{y}}(w_i(\boldsymbol{\alpha}))]}_{\begin{array}{l} \text{Expected change in fitness} \\ \text{of type } i \text{ in going from } \boldsymbol{\alpha} \text{ to } \mathbf{y} \\ (\text{scaled by deterministic density } \alpha_i) \end{array}} + \underbrace{\mu \mathbb{E}[D_{\mathbf{y}}(Q_i(\boldsymbol{\alpha}))]}_{\text{Expected effect of mutations}} \quad (3.41)$$

Thus, the expected behavior of fluctuations in the weak noise limit is controlled purely by fitness differences and mutational effects. If $\mathbb{E}[y_i] \equiv 0 \forall i$ is a stable fixed point for the system

of equations defined by (3.41), then stochastic fluctuations are expected to decay, meaning that the deterministic point α is a stable configuration for the complete dynamics when fluctuations are weak. In the case of 2-strategy matrix games (*i.e.* when $m = 2$ and the fitness functions $w_i(p)$ are of the form $w_i = \sum_j a_{ij} p_j$ for some constants a_{ij}), it has been shown that $\mathbb{E}[y_i] \equiv 0 \forall i$ is a stable fixed point for (3.41) if and only if the population state α is an evolutionarily stable strategy (ESS) for the deterministic game defined by (3.16) (which of course reduces to the replicator dynamics defined by (3.17)), thus recovering a stochastic version of a classic result in evolutionary game theory (Tao and Cressman, 2007).

Chapter 4

Stochastic field equations for the evolution of quantitative traits

The result has been forty years of bewilderment about what he meant, whereas if he had been willing to make a slight sacrifice of strict mathematical propriety (as I have done) he could have expressed himself in a way that everyone would have understood

George Price (1972), speaking about Fisher

So far, we have dealt with populations in which individuals come in countably many different kinds. While developing these models, we have been on mathematically solid ground that is well understood by statistical physicists and mathematicians. However, things become more complicated when we deal with ‘quantitative’ traits. Traits like body size, body weight, or beak length, often take on uncountably many values (say, all values in the interval $[0, 1]$, for example). In this case, we cannot describe the population using a vector as we did before, but instead require a function. More precisely, if the set of all possible trait values is \mathcal{T} , we will characterize the population using a special kind of function $\phi^{(t)}$ such that the

quantity $\int_A \phi^{(t)}(x)dx$ gives us the number of individuals that are in any ‘nice’ region $A \subset \mathcal{T}$ of the possible trait space¹ at time t . The state space of the stochastic process thus becomes infinite-dimensional, which complicates matters slightly. The principal objects of interest here are *functionals* $F[x, \phi^{(t)}]$ which take in a scalar x representing the trait value of interest, and a function $\phi^{(t)}$ representing the population at time t . Thus, whereas in the previous section we were interested in how a function $f(x(t))$ changes based on the change in an input variable $x(t)$ (the population), we are now interested in how a functional $F[\phi^{(t)}]$ changes with the change in an input function $\phi^{(t)}$.

The mathematics for these sorts of processes is an active area of research and is comparatively far from well developed. The mathematically rigorous formulation of the kinds of processes I study here falls in the realm of measure-valued branching processes, and is highly technical and rather inaccessible unless one is already comfortable with advanced measure-theoretic notions (Champagnat et al., 2006; Champagnat et al., 2008). This means that the existing formalism, while admirable in its generality and mathematical rigor, is rather unusable for most biologists, who do not have formal training in analysis (but see Week et al. (2021) for a very friendly introduction to the major ideas through heuristics). One can, however, make progress if they are willing to take some mathematical leaps of faith and sacrifice rigor for the sake of accessibility and heuristic understanding. I adopt this attitude below and hope that all questions of well-posedness, existence, etc. will be sorted out by some clever mathematicians in the future.

Physicists use the term ‘field’ for functions of the form $f(x, t) : \mathbb{R}^n \times [0, \infty) \rightarrow \mathbb{R}^m$, where \mathbb{R}^n represents space and $[0, \infty)$ represents time. They then call models which describe such functions ‘field theories’. In physics jargon, the stochastic process I will formulate when viewed as a sequence of functions thus describes a (scalar) ‘stochastic field’, and the formalism

¹The mathematically informed reader may notice that this sounds like I am trying to dance around the word ‘measure’. Indeed, we are really looking to construct branching processes that take values in some nice space of measures that can be endowed with sufficient mathematical structure for notions like convergence and integration to make sense and are absolutely continuous with respect to the Lebesgue measure. All the Dirac deltas that will turn up shortly are ‘properly’ viewed as measures, integrals with Dirac deltas in the integrand are ‘properly’ interpreted as integration with respect to the Dirac measure, and the functional derivative is to be ‘properly’ interpreted as a Gateaux or Fréchet derivative assuming that the relevant function space has enough structure for these notions to make sense. If one tries to be careful about these things, they will quickly find themselves drowning in a quagmire of mathematical formalism. If you know and care about enough mathematics for this to really bother you, see Champagnat et al. (2006) or Champagnat et al. (2008) for a much more rigorous treatment that avoids using informal tools such as functional derivatives and functional equivalents of Fokker-Planck equations in favor of a probabilistic approach grounded in (measure-theoretic) Markov and martingale theory.

I will develop below is a ‘stochastic field theory’ of evolution, where physical space has been replaced by an abstract trait space. This is closely related to the area of physics called ‘statistical field theory’, the analog of quantum field theory for systems with a large number of classical particles. Stochastic field theories over physical space have recently been used in biology to model brain function (Bressloff, 2010) and collective motion (Ó Laighléis et al., 2018).

In the following sections, I will rely heavily on a heuristic object called the functional derivative $\delta F/\delta\phi$. The functional derivative is an *ad hoc*, somewhat informal notion, defined indirectly as the unique object that obeys, for any ‘nice’ function ρ

$$\int \frac{\delta F}{\delta\phi(x)} \rho(x) dx = \lim_{h \rightarrow 0} \frac{F[\phi + h\rho] - F[\phi]}{h} \quad (4.1)$$

This definition is formulated in analogy to directional derivatives in multi-variable calculus: Noting that a function can be thought of as an infinite-dimensional vector, informally ‘taking the limit’ $m \rightarrow \infty$ in (3.25) yields (4.1).

4.1 Description of the process and the Master Equation

We envision a population of individuals with a ‘trait’ that takes values in some one-dimensional set $\mathcal{T} \subseteq \mathbb{R}$. Since the trait of any given individual is fixed, and since each individual can only have one exact trait value, an individual with a trait value $x \in \mathcal{T}$ can be characterized as a Dirac delta mass centered at x , defined indirectly as the object which satisfies, for any one-dimensional function f ,

$$\int_A f(y) \delta_{x_i} dy = \begin{cases} f(x_i) & x_i \in A \\ 0 & x_i \notin A \end{cases}$$

for every subset $A \subset \mathcal{T}$. The Dirac mass is often written $\delta_{x_i} = \delta(y - x_i)$ as a ‘function’ of a dummy variable y which will be integrated over, since in this view the Dirac mass is a ‘function’ that can only occur inside an integral. Though I will indeed (kind of) use the Dirac mass as a function, I will stick to the notation δ_{x_i} because it emphasizes that δ_{x_i} is supposed to represent an individual with a trait value of x_i (the dummy variable y can be confusing in this regard). Note that by choosing $f(x) \equiv 1$, we get an ‘indicator’ that is 1 if the individual is within the set A and 0 otherwise. If the population at any time t consists of

$N(t)$ individuals with trait values $\{x_1, x_2, \dots, x_{N(t)}\}$, then it can be completely characterized (Figure 4.1) by the ‘distribution’

$$\nu^{(t)} = \sum_{i=1}^{N(t)} \delta_{x_i}$$

which in physics notation would be a function $\nu^{(t)}(y) = \sum_{i=1}^{N(t)} \delta(y - x_i)$. Thus, the state space of our process is

$$\mathcal{M}(\mathcal{T}) = \left\{ \sum_{i=1}^n \delta_{x_i} \mid n \in \mathbb{N}, x_i \in \mathcal{T} \right\}$$

Note that for any set $A \subset \mathcal{T}$, $\int_A \nu^{(t)} dx$ gives the number of individuals that have trait values that lie within the set A and that integrating over \mathcal{T} gives the population size $N(t)$ at time t . Given the population $\nu^{(t)} = \sum_{i=1}^{N(t)} \delta_{x_i}$ and a real function $f(x)$, we have $\int_{\mathcal{T}} f(y) \nu^{(t)} dy = \sum_{i=1}^{N(t)} f(x_i)$. Now that we have described the population, we must define the rules for how it changes. I will do this through two non-negative functionals $b(x|\nu)$ and $d(x|\nu)$ from $\mathcal{T} \times \mathcal{M}(\mathcal{T})$ to $[0, \infty)$ that describe the rate at which individuals with trait value x are born and die respectively in a population ν . Again, we must be careful about what exactly we mean when we speak about ‘rates’. In this case, I mean that if we know that the population is currently described by the function ν , and we know that *either a birth or a death* occurs, then the probability that this event is the birth of an individual whose phenotype is within the set $A \subset \mathcal{T}$ is given by

$$\mathbb{P}[\text{Birth with offspring in } A \mid \text{something happened}] = \frac{1}{\mathcal{N}} \int_A b(x|\nu) dx$$

and the probability that the event is the death of an individual whose phenotype is within the set A is

$$\mathbb{P}[\text{Death of an individual in } A \mid \text{something happened}] = \frac{1}{\mathcal{N}} \int_A d(x|\nu) dx$$

where $\mathcal{N} = \int_{\mathcal{T}} b(x|\nu) + d(x|\nu) dx$ is the normalizing constant in both cases. Note that we assume \mathcal{N} is always finite and non-zero.

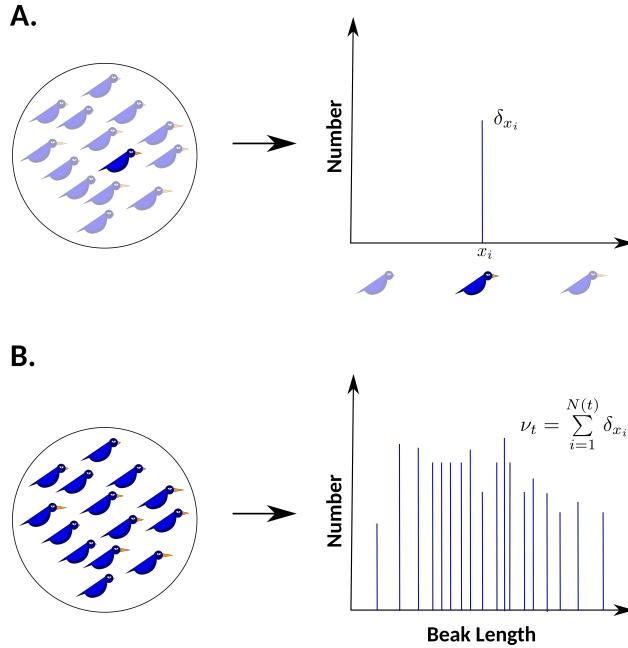


Figure 4.1: Schematic description of a function valued birth-death process. Consider a population of birds in which individuals have varying beak lengths. **(A)** Each individual in the population can be described as a Dirac delta mass centered at its beak length. This is because each individual has exactly one fixed beak length, and therefore, can be thought of as a distribution centered at that particular beak length and with zero spread. **(B)** The population as a whole is thus described as a sum of Dirac masses. $N(t)$ here is the size of the population at time t . Birth and death of individuals would correspond to the addition and removal of Dirac masses respectively. Note that if we had a large number of individuals, this distribution begins to look like a continuous distribution.

Example 2. Consider the birth and death functionals:

$$\begin{aligned} b(x|\nu) &= r \int_{\mathcal{T}} m(x, y) \nu(y) dy; \quad m(x, y) = \exp\left(\frac{-(x-y)^2}{\sigma_m^2}\right) \\ d(x|\nu) &= \frac{\nu(x)}{Kn(x)} \int_{\mathcal{T}} \alpha(x, y) \nu(y) dy; \quad \alpha(x, y) = \exp\left(\frac{-(x-y)^2}{\sigma_\alpha^2}\right) \end{aligned} \tag{4.2}$$

This choice corresponds to an asexual population having a constant (per-capita) birth rate r . Birth is sometimes with mutation, and the extent of the mutations is controlled by a Gaussian kernel $m(x, y)$. The death rate is density-dependent, mediated by a Gaussian competition kernel $\alpha(x, y)$, and also contains a phenotype-dependent carrying capacity controlled by $n(x)$, scaled by a constant K . The biological interpretation of the death rate is through

ecological specialization for limiting resources - individuals have different intrinsic advantages (controlled by $n(x)$), and experience greater competition from conspecifics that are closer to them in phenotype space (controlled by $\alpha(x, y)$).

Let us now define, for each $x \in \mathcal{T}$, two *step operators* \mathcal{E}_x^\pm that satisfy

$$\mathcal{E}_x^\pm[f(y, \nu)] = f(y, \nu \pm \delta_x)$$

In other words, the step operators \mathcal{E}_x^\pm simply describe the effect of adding or removing a single individual with trait value x from the population. We will assume that our process is nice enough that we can find a function $P(\nu, t)$ such that the probability that the process takes value $\nu^{(t)}$ at time t is given by² $\int_{\mathcal{T}} P(\nu, t) dx$. In one dimension, this is often automatically satisfied for biologically reasonable choices of the birth and death functionals (Dawson, 1975; Walsh, 1986; Konno and Shiga, 1988; Reimers, 1989; Dawson et al., 2000; Also see section 2.3 in Etheridge, 2000 and proposition 3.1 in Champagnat et al., 2008).

We can now use the same trick as in chapter 3 and obtain a master equation by counting inflow and outflow of states. Any change to a state must be through the addition or subtraction of a single individual (now a single Dirac mass). For any state $\nu \in \mathcal{M}(\mathcal{T})$, the transition rate from $\nu - \delta_x$ to ν is simply $\mathcal{E}_x^- b(x|\nu)$, and similarly, the transition rate from $\nu + \delta_x$ to ν is $\mathcal{E}_x^+ d(x|\nu)$. The transition rate out of ν to a state $\nu + \delta_x$ is just $b(x|\nu)$, and transition out to a state $\nu - \delta_x$ is just $d(x|\nu)$. Thus, integrating over all possible x to obtain the total inflow and

²If you know some measure theory, note that this amounts to saying the following (I think?). Let $(\mathcal{M}(\mathcal{T}), \mathcal{F}, \{\mathcal{F}_t\}_{t \geq 0}, \mathbb{P})$ be the filtered probability space in which our adapted stochastic process $\{X_t\}_{t \geq 0}$ lives. Let m be the Lebesgue measure on $(\mathcal{T}, \mathcal{B}(\mathcal{T}))$. Then, we require $\mathbb{P}(X_t \ll m \ \forall t) = 1$, i.e. we need the process to be $[\mathbb{P}]$ -a.s. absolutely continuous with respect to the Lebesgue measure for all time. Since each $\nu \in \mathcal{M}(\mathcal{T})$ is a finite measure, it is obviously σ -finite, and we can now thus use the Radon-Nikodym theorem to find a density function $R_m(x, t)$ such that for any $t \geq 0$ and any $B \in \mathcal{B}(\mathcal{T})$, we have $X_t(B) = \int_B R_m(x, t) dm$ $[\mathbb{P}]$ -a.s.. We also require being able to find a second function $Q_{\mathbb{P}}(\nu, t)$ on $\mathcal{F} \times [0, \infty)$ such that we can write $\mathbb{P}(X_t = \nu) = \int_{\mathcal{T}} Q_{\mathbb{P}}(\nu, t) \nu(dx)$. The function $Q_{\mathbb{P}}(\nu, t)$ is similar to a density function and thus essentially amounts to another absolute continuity condition, except on the stochastic process itself (and thus the measure \mathbb{P} defined on \mathcal{F}) rather than on the elements of its state space (i.e. the measures $\nu \in \mathcal{M}(\mathcal{T})$, which are defined on \mathcal{T}). We can now use the $[\mathbb{P}]$ -a.s. assured absolute continuity with respect to the Lebesgue measure of any measure ν attained by our process X_t to write $\int_{\mathcal{T}} Q_{\mathbb{P}}(\nu, t) \nu(dx) = \int_{\mathcal{T}} Q_{\mathbb{P}}(\nu, t) R_m(x, t) dm$. Changing notation from dm to dx to revert to the usual abuse of notation for integration with respect to the Lebesgue measure, we can now define $P(\nu, t) dx := Q_{\mathbb{P}}(\nu, t) R_m(x, t) dx$. This allows us finally to write $\mathbb{P}(X_t = \nu) = \int_{\mathcal{T}} P(\nu, t) dx$ as desired.

outflow rate for a state ν , we see that $P(\nu, t)$ must satisfy:

$$\frac{\partial P}{\partial t}(\nu, t) = \int_{\mathcal{T}} [(\mathcal{E}_x^- - 1)b(x|\nu)P(\nu, t) + (\mathcal{E}_x^+ - 1)d(x|\nu)P(\nu, t)] dx \quad (4.3)$$

This is the master equation of our infinite-dimensional process.

4.2 The functional system-size expansion

To proceed, as before, I assume that there exists a system-size parameter $K > 0$ such the total population size is controlled by K , with $K = \infty$ corresponding to an infinitely large population. This allows us to obtain a new process $\{\phi^{(t)}\}_{t \geq 0}$ such that for any set $A \subset \mathcal{T}$, $\int_A \phi^{(t)} dx$ gives the ‘density’ of individuals that have trait values that lie within the set A . Note that we expect this stochastic process to evolve continuously if K is large since the contribution of each individual is negligible. Specifically, I assume we can make the substitutions:

$$\begin{aligned}\phi^{(t)} &:= \frac{1}{K} \nu^{(t)} = \frac{1}{K} \sum_{i=1}^{N(t)} \delta_{x_i} \\ b_K(x|\phi^{(t)}) &:= \frac{1}{K} b(x|\nu^{(t)}) \\ d_K(x|\phi^{(t)}) &:= \frac{1}{K} d(x|\nu^{(t)})\end{aligned}$$

$\{\phi^{(t)}\}_{t \geq 0}$ takes values in

$$\mathcal{M}_K(\mathcal{T}) := \left\{ \frac{1}{K} \sum_{i=1}^n \delta_{x_i} \mid n \in \mathbb{N}, x_i \in \mathcal{T} \right\}$$

In terms of these new variables, we obtain the master equation:

$$\frac{\partial P}{\partial t}(\phi, t) = K \int_{\mathcal{T}} [(\Delta_x^- - 1)b_K(x|\phi)P(\phi, t) + (\Delta_x^+ - 1)d_K(x|\phi)P(\phi, t)] dx \quad (4.4)$$

where I have introduced new step operators Δ_x^\pm that satisfy:

$$\Delta_x^\pm[F(y, \phi)] = F\left(y, \phi \pm \frac{1}{K} \delta_x\right)$$

We can now conduct a system-size expansion as before by using a functional ‘Taylor expansion’ of the step operators. Recall that the functional version of the Taylor expansion of a functional $F[\rho]$ about a function ρ_0 defined on a domain $\Omega \subseteq \mathbb{R}$ is given by:

$$F[\rho_0 + \rho] = F[\rho_0] + \int_{\Omega} \rho(x) \frac{\delta F}{\delta \rho_0(x)} dx + \frac{1}{2!} \int_{\Omega} \int_{\Omega} \rho(x) \rho(y) \frac{\delta^2 F}{\delta \rho_0(x) \delta \rho_0(y)} dx dy + \dots$$

Since $\Delta_x^{\pm}[F[\phi]] = F[\phi \pm \delta_x/K]$, we can Taylor expand the RHS to see that our step operators obey

$$\begin{aligned} \Delta_x^{\pm}[F[\phi]] &= F[\phi] \pm \frac{1}{K} \int_{\mathcal{T}} \frac{\delta F}{\delta \phi(y)} \delta_x dy + \frac{1}{2K^2} \int_{\mathcal{T}} \int_{\mathcal{T}} \frac{\delta^2 F}{\delta \phi(y) \delta \phi(z)} \delta_x dy \delta_x dz + \mathcal{O}(K^{-3}) \\ &= F[\phi] \pm \frac{1}{K} \frac{\delta F}{\delta \phi(x)} + \frac{1}{2K^2} \frac{\delta^2 F}{\delta \phi(x)^2} + \mathcal{O}(K^{-3}) \end{aligned} \quad (4.5)$$

Neglecting terms of $\mathcal{O}(K^{-3})$, we can now substitute (4.5) into (4.4) to obtain:

$$\begin{aligned} \frac{\partial P}{\partial t}(\phi, t) &= K \int_{\mathcal{T}} \left[\left(-\frac{1}{K} \frac{\delta}{\delta \phi(x)} + \frac{1}{2K^2} \frac{\delta^2}{\delta \phi(x)^2} \right) \{b_K(x|\phi)P(\phi, t)\} \right] dx \\ &\quad + K \int_{\mathcal{T}} \left[\left(\frac{1}{K} \frac{\delta}{\delta \phi(x)} + \frac{1}{2K^2} \frac{\delta^2}{\delta \phi^2(x)} \right) \{d_K(x|\phi)P(\phi, t)\} \right] dx \end{aligned}$$

Rearranging these terms, we obtain a ‘functional Fokker-Planck equation’:

$$\frac{\partial P}{\partial t}(\phi, t) = \int_{\mathcal{T}} \left[-\frac{\delta}{\delta \phi(x)} \{\mathcal{A}^-(x|\phi)P(\phi, t)\} + \frac{1}{2K} \frac{\delta^2}{\delta \phi(x)^2} \{\mathcal{A}^+(x|\phi)P(\phi, t)\} \right] dx \quad (4.6)$$

where

$$\mathcal{A}^{\pm}(x|\phi) = b_K(x|\phi) \pm d_K(x|\phi) = \frac{1}{K} (b(x|\nu) \pm d(x|\nu))$$

Equation (4.6) constitutes the ‘mesoscopic’ description of our process. We can once again appeal to the link between Fokker-Planck equations and Itô processes to say that (4.6) corresponds to the SPDE:

$$\frac{\partial \phi}{\partial t}(x, t) = \mathcal{A}^-(x|\phi) + \sqrt{\frac{\mathcal{A}^+(x|\phi)}{K}} \dot{W}(x, t) \quad (4.7)$$

where $\dot{W}(x, t)$ is the so-called ‘spacetime white noise’ process. If one wishes to be mathematically careful, the connection between (4.6) and (4.7) becomes somewhat tenuous. In particular, while the equivalence between Fokker-Planck equations and SDEs (Langevin equations) for finite-dimensional stochastic processes is part of the standard mathematical canon, the corresponding equivalence is much less well understood for the infinite-dimensional measure-valued processes we are dealing with - we may need $\mathcal{M}_K(\mathcal{T})$ to have a considerable amount of structure (Ex: separable Hilbert space) for things to work out (Da Prato and Zabczyk, 2014; Balan, 2018), and I do not know whether our domain is nice enough. For example, it is not immediately clear what conditions on $\mathcal{A}^\pm(x|\phi)$ are needed for (4.7) to even admit a solution. Nevertheless, as I show below, we can recover some well-known deterministic equations from equation (4.7) in the infinite population limit, illustrating consistency with known models. Under certain assumptions on the domain $\mathcal{M}_K(\mathcal{T})$ and the birth-death operators \mathcal{A}^\pm , equivalences between functional Fokker-Planck equations and SPDEs are rigorously studied and understood (See chapter 10 in Bogachev et al., 2015).

4.3 The infinite population limit

Taking $K \rightarrow \infty$ in equation (4.7) yields a PDE:

$$\frac{\partial \psi}{\partial t}(x, t) = \mathcal{A}^-(x|\psi) = b_K(x|\psi) - d_K(x|\psi) \quad (4.8)$$

where I have used a different symbol ψ simply to highlight that $\psi(x, t)$ as the solution to equation (4.8) is a deterministic function, whereas $\phi(x, t)$ as defined in equation (4.7) is really a stochastic process $\{\phi^{(t)}\}_{t \geq 0}$ in which each $\phi^{(t)}$ is a finite measure. Equation (4.8) simply says that in the absence of stochasticity, the change in the density of individuals with trait values x is given by the difference between the birth and death rates of these individuals in the population. Models of this form are precisely the ‘PDE models’ discussed in studies of Adaptive Diversification (Doebeli, 2011). Equation (4.8) is also the basic equation of ‘oligomorphic dynamics’ (Sasaki and Dieckmann, 2011; Lion et al., 2022) if one assumes the population is composed of a small number of ‘morphs’, *i.e.* $\psi(x, t) = \sum_{k=1}^S n_k(t)\psi_k(x, t)$, where $n_k \geq 0$ is the abundance of the k^{th} morph, $\psi_k(x, t)$ is the phenotypic distribution of the k^{th} morph (often assumed a normal distribution with narrow variance) and S is the total number of distinct morphs in the population. Models of the form (4.8) are also used to study intraspecific trait variation in community ecology (Nordbotten et al., 2020). A prominent

recent example is Wickman et al.'s (2022) 'trait space equations' in their framework for eco-evolutionary community dynamics.

As before, I assume that the birth and death functions take the form:

$$\begin{aligned} b_K(x|\psi) &= \psi(x, t)b^{(\text{ind})}(x|\psi) + \mu Q(x|\psi) \\ d_K(x|\psi) &= \psi(x, t)d^{(\text{ind})}(x|\psi) \end{aligned} \quad (4.9)$$

As in chapter 3, $Q(x|\psi)$ describes birth due to mutations and $\mu \geq 0$ is a constant mutation rate. The functions $b^{(\text{ind})}(x|\psi)$ and $d^{(\text{ind})}(x|\psi)$ describe the per-capita birth rate and death rate of type x individuals in a population ψ . These functions could in principle model several ecological factors. For example, $b^{(\text{ind})}(x|\psi)$ may incorporate the effects of mate choice in the sexual case or intrinsic duplication rates in the asexual case, and $d^{(\text{ind})}(x|\psi)$ may model death due to intraspecific competition for resources. Note that the definition of the mean value (3.12) of a type level quantity $f(x)$ now becomes

$$\bar{f}(t) = \int_{\mathcal{T}} f(x)p(x, t)dx$$

Substituting equation (4.9) into (4.8), we obtain

$$\frac{\partial \psi}{\partial t}(x, t) = w(x|\psi)\psi(x, t) + \mu Q(x|\psi) \quad (4.10)$$

where I have defined $w(x|\psi) := b^{(\text{ind})}(x|\psi) - d^{(\text{ind})}(x|\psi)$, the (Malthusian) 'fitness' of the phenotype x . To track population numbers and trait frequencies, let us define as before, the scaled population size and trait frequency as

$$\begin{aligned} N_K(t) &:= \int_{\mathcal{T}} \psi(x, t)dx = \frac{1}{K} \int_{\mathcal{T}} \nu(x, t)dx \\ p(x, t) &:= \frac{\psi(x, t)}{N_K(t)} = \frac{\nu(x, t)}{\int_{\mathcal{T}} \nu(y, t)dy} \end{aligned} \quad (4.11)$$

The population mean fitness is:

$$\bar{w}(t) = \int_{\mathcal{T}} w(x|\psi)p(x, t)dx \quad (4.12)$$

Using the chain rule in the definition of $p(x, t)$, we can calculate:

$$\begin{aligned}\frac{\partial p}{\partial t} &= \frac{1}{N_K(t)} \frac{\partial \psi}{\partial t}(x, t) - \frac{\psi(x, t)}{N_K^2(t)} \frac{dN_K}{dt} \\ &= \frac{1}{N_K(t)} \frac{\partial \psi}{\partial t}(x, t) - \frac{\psi(x, t)}{N_K^2(t)} \int_{\mathcal{T}} \frac{\partial \psi}{\partial t}(y, t) dy\end{aligned}$$

Where I have used the definition of $N_K(t)$ and assumed that integrals and derivatives commute in the second line. Substituting (4.10), we now obtain

$$\begin{aligned}\frac{\partial p}{\partial t} &= \frac{1}{N_K(t)} [w(x|\psi)\psi(x, t) + \mu Q(x|\psi)] - \frac{\psi(x, t)}{N_K^2(t)} \int_{\mathcal{T}} w(y|\psi)\psi(y, t) + \mu Q(y|\psi) dy \\ &= w(x|\psi)p(x, t) + \frac{\mu}{N_K(t)} Q(x|\psi) - p(x, t) \left(\int_{\mathcal{T}} w(y|\psi)p(y, t) dy + \frac{\mu}{N_K(t)} \int_{\mathcal{T}} Q(y|\psi) dy \right)\end{aligned}$$

where I have used the definition of $p(x, t)$ in the second line. Using (4.12) and rearranging the terms gives us:

$$\frac{\partial p}{\partial t}(x, t) = [w(x|\psi) - \bar{w}(t)] p(x, t) + \frac{\mu}{N_K(t)} \left[Q(x|\psi) - p(x, t) \int_{\mathcal{T}} Q(y|\psi) dy \right] \quad (4.13)$$

This is a version of the replicator-mutator equation for continuous strategy spaces when each x is viewed as a strategy (Cressman and Tao, 2014).

Equation (4.13) also recovers Kimura's (1965) continuum-of-alleles model when each x is viewed as an allele, $Q(x|\psi)$ takes the form of a convolution of $\psi(x, t)$ with a mutation kernel, and the trait space is the entire real line, *i.e.* $\mathcal{T} = \mathbb{R}$. To see this, let $Q(y|\psi) = \int_{\mathbb{R}} m(y - z)\psi(z, t) dz$, where $m : \mathbb{R} \rightarrow [0, \infty)$ is a mutation kernel, which by definition is normalized such that $\int_{\mathbb{R}} m(x) dx = 1$. Let us further note that I have implicitly been assuming³ that even though the total number of individuals $\int \nu(x) dx \rightarrow \infty$ as $K \rightarrow \infty$, the scaled population size $N_K = \int \nu(x) dx / K$ remains finite at all times, *i.e.* $N_K(t) = \int_{\mathbb{R}} \psi(x, t) dx < \infty \forall t$. Thus, for

³This actually need not be an assumption, and can be derived as a theorem under biologically reasonable conditions on the birth/death functionals upon taking the rigorous approach to both the system size expansion (rescaling) and the infinite population limit. See, for example, Etheridge (2000) or proposition 4.1 in Champagnat et al. (2008)

any fixed $t > 0$, we have

$$\begin{aligned} \int_{\mathbb{R}} \int_{\mathbb{R}} |m(y - z)\psi(z, t)| dy dz &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} |m(y - z)| dy \right) |\psi(z, t)| dz \\ &= \left(\int_{\mathbb{R}} |m(x)| dx \right) \left(\int_{\mathbb{R}} |\psi(z, t)| dz \right) < \infty \end{aligned}$$

We can therefore conclude that $m(y - z)\psi(z, t) \in L^1(\mathbb{R} \times \mathbb{R})$ for any given $t > 0$, meaning we can use the Fubini-Tonnelli theorem to interchange the order of integration of iterated integrals of $m(y - z)\psi(y)$. We are now ready to evaluate the rightmost integral of (4.13).

We have:

$$\begin{aligned} \int_{\mathbb{R}} Q(y|\psi) dy &= \int_{\mathbb{R}} \int_{\mathbb{R}} m(y - z)\psi(z, t) dz dy \\ &= \int_{\mathbb{R}} \int_{\mathbb{R}} m(y - z)\psi(z, t) dy dz \\ &= \int_{\mathbb{R}} \psi(z, t) \left(\int_{\mathbb{R}} m(y - z) dy \right) dz \\ &= \int_{\mathbb{R}} \psi(z, t) \int_{\mathbb{R}} m(u) du dz \\ &= \int_{\mathbb{R}} \psi(z, t) dz \int_{\mathbb{R}} m(u) du \\ &= N_K(t) \int_{\mathbb{R}} m(u) du \end{aligned} \tag{4.14}$$

where I have used the Fubini-Tonnelli theorem to go from the first step to the second, and have made the substitution $u = y - z$ to go from the third to the fourth step. We then note that since m is a kernel, it satisfies $\int_{\mathbb{R}} m(u) du = 1$, and (4.14) therefore becomes $\int_{\mathbb{R}} Q(y|\psi) dy = N_K(t)$. Substituting this in (4.13), we have

$$\frac{\partial p}{\partial t}(x, t) = [w(x|\psi) - \bar{w}(t)] p(x, t) + \frac{\mu}{N_K(t)} \left[\int_{\mathbb{R}} m(x - z)\psi(z, t) dz - p(x, t)N_K(t) \right]$$

Substituting our definition $p(z, t) = \psi(z, t)/N_K(t)$ now yields

$$\boxed{\frac{\partial p}{\partial t}(x, t) = [w(x|\psi) - \bar{w}(t)] p(x, t) + \mu \left[\int_{\mathbb{R}} m(x-z)p(z, t)dz - p(x, t) \right]} \quad (4.15)$$

which is Kimura's continuum of alleles model (Kimura, 1965; Crow and Kimura, 1970).

We can now use the same trick we used in deriving (3.20) from (3.17). By multiplying both sides of equation (4.13) by a type level quantity⁴ $f(x, t)$ and integrating over the trait space, we obtain

$$\begin{aligned} \frac{d\bar{f}}{dt} &= \int_{\mathcal{T}} f(x, t) w(x|\psi) p(x, t) dx - \bar{w}(t) \int_{\mathcal{T}} f(x, t) p(x, t) dx \\ &\quad + \frac{\mu}{N_K(t)} \int_{\mathcal{T}} f(x, t) \left[Q(x|\psi) - p(x, t) \int_{\mathcal{T}} Q(y|\psi) dy \right] dx \\ &= \bar{f}\bar{w} - \bar{w} \cdot \bar{f} + \frac{\mu}{N_K(t)} \int_{\mathcal{T}} f(x, t) \left[Q(x|\psi) - p(x, t) \int_{\mathcal{T}} Q(y|\psi) dy \right] dx \end{aligned} \quad (4.16)$$

We now observe that

$$\text{Cov}(f, w(x|\psi)) = \bar{f}\bar{w} - \bar{f} \cdot \bar{w} \quad (4.17)$$

is the statistical covariance of the quantity f with the Malthusian fitness function. The second term, which I will denote by

$$M_{\bar{f}}(x|\psi) := \frac{\mu}{N_K(t)} \left[\int_{\mathcal{T}} f(x, t) Q(x|\psi) dx - \left(\bar{f} \int_{\mathcal{T}} Q(x|\psi) dx \right) \right] \quad (4.18)$$

reflects the transmission bias of mutations. Thus, we see that equation (4.16) reads

$$\boxed{\frac{d\bar{f}}{dt} = \text{Cov}(f, w(x|\psi)) + M_{\bar{f}}(x|\psi) + \left(\frac{\partial f}{\partial t} \right)} \quad (4.19)$$

⁴This is now a real function $f(x, t) : \mathcal{T} \times [0, \infty) \rightarrow \mathbb{R}$ of two variables, the trait value and time. I assume this function is nice enough for all the below operations to make sense

from which it is clear that we have obtained a version of the Price equation (3.20) for quantitative traits (Note that (4.19) is precisely the equation obtained by informally taking $m \rightarrow \infty$ in (3.20)). For the special case $f(x) = x$, we have $\partial f / \partial t = 0$ and thus:

$$\frac{d\bar{x}}{dt} = \text{Cov}(x, w(x|\psi)) + M_{\bar{x}}(x|\psi) \quad (4.20)$$

We can also recover some more familiar dynamics under the following additional assumptions:

- Rare mutations, *i.e.* $\mu \rightarrow 0$.
- Small mutational effects with ‘almost faithful’ reproduction, meaning $Q(x|\psi) \rightarrow 0$, and the distribution $\psi(x, t)$ tends to stay very ‘sharp’ (*i.e* strongly peaked about its mean value).
- Separation of ecological and evolutionary timescales, meaning that the system is always at ecological equilibrium. Thus, the expected rate of change of resident numbers in a resident population is 0, and we have $w(y|\delta_{y(t)}) = 0$.

Under these assumptions, if we supply an initial condition $\psi(x, 0) = N_K(0)\delta_{y_0}$ for some constants $N_K(0) > 0$ and $y_0 \in \mathcal{T}$ (meaning we start with a completely monomorphic population of size $N_K(0)$ in which all individuals have trait value y_0), it is reasonable to assume that the population remains sufficiently well clustered for some (possibly small) time $t > 0$ that we can continue to approximate the distribution $\psi(x, t)$ as a Dirac Delta mass $N_K(t)\delta_{y(t)}$ that is moving across the trait space in a deterministic manner dictated by a function $y(t)$ (to be found). Note that we have $p(x, t) = \delta_{y(t)}$, $\bar{x}(t) = y(t)$, and $\bar{w}(t) = 0$. Thus, from equation (4.20), we have

$$\begin{aligned} \frac{d\bar{x}}{dt} &= \int_{\mathcal{T}} (x - \bar{x}(t))(w(x|\psi) - \bar{w}(t))p(x, t)dx \\ \Rightarrow \frac{dy}{dt} &= \int_{\mathcal{T}} (x - \bar{x}(t))w(x|N_K\delta_{y(t)})\delta_{y(t)}dx \end{aligned} \quad (4.21)$$

Our assumptions on mutation rate and mutational effects imply that the population will be concentrated in an infinitesimal neighborhood around the mean value $y(t)$ (*i.e* that the distribution of traits in the population is sharply peaked). We can thus Taylor expand

$w(x|N_K\delta_{y(t)})$ about $y(t)$ as:

$$w(x|N_K\delta_{y(t)}) = \underbrace{w(y|N_K\delta_{y(t)})}_{=0} + (x - y(t)) \frac{d}{dz} w(z|N_K\delta_{y(t)}) \Big|_{z=y} + \dots$$

Thus, substituting in (4.21), to first order, we obtain

$$\frac{dy}{dt} = \left(\int_{\mathcal{T}} (x - \bar{x}(t))^2 p(x, t) dx \right) \frac{d}{dz} w(z|N_K\delta_{y(t)}) \Big|_{z=y}$$

where I have used $\bar{x}(t) = y(t)$. We can define the shorthand $B(y) = \int_{\mathcal{T}} (x - y(t))^2 p(x, t) dx = \int_{\mathcal{T}} (x - \bar{x}(t))^2 p(x, t) dx$ to obtain:

$$\frac{dy}{dt} = B(y) \left(\frac{d}{dz} w(z|N_K\delta_{y(t)}) \Big|_{z=y} \right) \quad (4.22)$$

Note that by the definition of statistical variance, the function $B(y(t))$ is numerically equal to $\sigma_x^2(t)$, the statistical variance of the trait in the population at time t . The term $w(z|N_K\delta_{y(t)})$ is the expected growth rate of an individual with trait value z in a population of size N_K in which (almost) every individual has trait value y . This quantity is referred to as the ‘invasion fitness’ of a ‘mutant’ trait z in a population of ‘resident’ y individuals. Equation (4.22) is the canonical form of a broad class of systems called ‘gradient equations’ or ‘gradient dynamics’ in quantitative genetics (Lande, 1982; Abrams et al., 1993; Lehtonen, 2018; Lion, 2018), and captures the approximate evolutionary dynamics of quantitative traits under certain mutation limits. It is also deeply related (Lehtonen, 2018; Lion, 2018) to the canonical equation of adaptive dynamics first formulated in Dieckmann and Law (1996). The major difference is that in the ‘proper’ canonical equation of adaptive dynamics (as formulated in Dieckmann and Law (1996)), the function $B(y)$ explicitly relies on mutations as a continual source of variation, whereas in gradient dynamics and our equation (4.22), $B(y) = \sigma_x^2(t)$ captures the standing genetic variation in the population but does not specify the source of this variation. Note that strictly speaking, if $\psi(x, t) = \delta_{y(t)}$ exactly, then $B(y) \equiv 0$. This just reflects our assumption that mutations are vanishingly rare and mutants are sampled from infinitesimally close to the resident value. More detailed mathematical arguments are required to ensure that this convergence ‘makes sense’ and that $B(y)$ does not actually equal 0. This has been

proved rigorously using much more sophisticated mathematical tools grounded in martingale theory (Champagnat et al., 2006). A heuristic derivation of the canonical equation of adaptive dynamics is provided in the classic article by Dieckmann and Law (1996).

4.4 Stochastic trait frequency dynamics in the infinite-dimensional case

In chapter 3, I also derived SDEs for the trait frequency dynamics of the complete stochastic case using Itô’s formula. Doing the same for quantitative traits in our framework is tricky because it requires us to find a version of Itô’s formula that holds for SPDEs of the form (4.7). The formulation of Itô formulas and/or an infinite-dimensional stochastic calculus for general function (measure) valued stochastic processes is an active area of research in pure mathematics (Da Prato and Zabczyk, 2014), and without further information on the nature of the domain $\mathcal{M}_K(\mathcal{T})$, it is not clear (to me at least) whether an Itô’s formula exists for our case in general. However, it bears noting that the ‘intuitive’ Itô’s formula one would expect does indeed hold true for broad classes of Hilbert space valued processes (Prévôt and Röckner, 2007; Da Prato and Zabczyk, 2014; Liu and Röckner, 2015). If $\mathcal{A}^\pm(x|\phi)$ in (4.6) are Gaussian functions, then it has been proven (Week et al., 2021) that the SDEs ‘work as expected’ if we take $m \rightarrow \infty$ in (3.14) (see section 5.1 in chapter 5). However, carrying out a general derivation is beyond the scope of this thesis.

4.5 Stochastic fluctuations and the weak noise approximation

We can also formally carry out a functional analogue of the weak noise expansion as we did in chapter 3. Assume that $\psi(x, t)$ is the deterministic trajectory obtained as the solution to (4.8). We introduce a new process $\{\zeta^{(s)}\}_{s \geq 0}$ which measures the fluctuations of $\phi^{(t)}$ from the deterministic trajectory $\psi(x, t)$. More precisely, let us introduce the new variables:

$$\begin{aligned}\zeta^{(s)}(x) &= \sqrt{K}(\phi^{(t)}(x) - \psi(x, t)) \\ s &= t \\ \tilde{P}(\zeta, s) &= \frac{1}{\sqrt{K}}P(\phi, t)\end{aligned}\tag{4.23}$$

Note that the following relations hold:

$$\frac{\delta F[\zeta]}{\delta \phi(x)} = \int_{\mathcal{T}} \frac{\delta F[\zeta]}{\delta \zeta(y)} \frac{\delta \zeta(y)}{\delta \phi(x)} dy = \sqrt{K} \frac{\delta F[\zeta]}{\delta \zeta(x)} \quad (4.24)$$

$$\frac{\partial}{\partial s} = \frac{\partial}{\partial t} \quad (4.25)$$

Furthermore, for any $\zeta \in \mathcal{M}_K(\mathcal{T})$, we have:

$$\begin{aligned} \frac{\partial \tilde{P}}{\partial t}(\zeta, s) &= \frac{\delta \tilde{P}}{\delta \zeta} \frac{\partial \zeta}{\partial t} + \frac{\partial \tilde{P}}{\partial s} \frac{\partial s}{\partial t} \\ &= \frac{\delta \tilde{P}}{\delta \zeta} \left(-\sqrt{K} \frac{\partial \psi}{\partial t} \right) + \frac{\partial \tilde{P}}{\partial s} \\ &= -\sqrt{K} \frac{\delta}{\delta \zeta} \{ \mathcal{A}^-(x|\psi) \tilde{P}(\zeta, s) \} + \frac{\partial \tilde{P}}{\partial s} \end{aligned} \quad (4.26)$$

Reformulating equation (4.6) in terms of the new variables (4.23) and using the relations (4.24), (4.25) and (4.26), we obtain:

$$\begin{aligned} -\sqrt{K} \frac{\delta}{\delta \zeta(x)} \{ \mathcal{A}^-(x|\psi) \tilde{P}(\zeta, s) \} + \frac{\partial \tilde{P}}{\partial s} &= \int_{\mathcal{T}} \left[- \left(\sqrt{K} \frac{\delta}{\delta \zeta(x)} \right) \{ \mathcal{A}^- \left(x \middle| \psi + \frac{\zeta}{\sqrt{K}} \right) \tilde{P}(\zeta, s) \} \right] dx \\ &\quad + \int_{\mathcal{T}} \left[\frac{1}{2K} \left(K \frac{\delta^2}{\delta \zeta(x)^2} \right) \{ \mathcal{A}^+ \left(x \middle| \psi + \frac{\zeta}{\sqrt{K}} \right) \tilde{P}(\zeta, s) \} \right] dx \end{aligned}$$

and rearranging gives us:

$$\begin{aligned} \frac{\partial \tilde{P}}{\partial s} &= -\sqrt{K} \int_{\mathcal{T}} \frac{\delta}{\delta \zeta(x)} \left\{ \left(\mathcal{A}^- \left(x \middle| \psi + \frac{\zeta}{\sqrt{K}} \right) - \mathcal{A}^-(x|\psi) \right) \tilde{P}(\zeta, s) \right\} dx \\ &\quad + \frac{1}{2} \int_{\mathcal{T}} \frac{\delta^2}{\delta \zeta(x)^2} \{ \mathcal{A}^+ \left(x \middle| \psi + \frac{\zeta}{\sqrt{K}} \right) \tilde{P}(\zeta, s) \} dx \end{aligned} \quad (4.27)$$

We will now Taylor expand our functionals about ψ (we assume that this is possible). Thus, we have the expansions:

$$\mathcal{A}^- \left(x \middle| \psi + \frac{\zeta}{\sqrt{K}} \right) = \mathcal{A}^- (x|\psi) + \frac{1}{\sqrt{K}} \int_{\mathcal{T}} \zeta(y) \frac{\delta}{\delta \psi(y)} \{ \mathcal{A}^-(y|\psi) \} dy + \dots$$

$$\mathcal{A}^+ \left(x \middle| \psi + \frac{\zeta}{\sqrt{K}} \right) = \mathcal{A}^+(x|\psi) + \frac{1}{\sqrt{K}} \int_{\mathcal{T}} \zeta(y) \frac{\delta}{\delta \psi(y)} \{ \mathcal{A}^+(y|\psi) \} dy + \dots$$

We also assume that \tilde{P} can be expanded as

$$\tilde{P} = \sum_{n=0}^{\infty} \tilde{P}_n \left(\frac{1}{\sqrt{K}} \right)^n$$

substituting these expansions into equation (4.27), equating coefficients of powers of $1/K$, and truncating at the lowest order term, we have:

$$\frac{\partial \tilde{P}_0}{\partial s}(\zeta, s) = \int_{\mathcal{T}} \left[-\frac{\delta}{\delta \zeta(x)} \left\{ \int_{\mathcal{T}} \zeta(y) \frac{\delta}{\delta \psi(y)} \{ \mathcal{A}^-(y|\psi) \} dy \tilde{P}_0(\zeta, s) \right\} + \frac{1}{2} \mathcal{A}^+(x|\psi) \frac{\delta^2}{\delta \zeta(x)^2} \{ \tilde{P}_0(\zeta, s) \} \right] dx$$

We thus arrive at the functional Fokker-Planck equation:

$$\frac{\partial \tilde{P}_0}{\partial s}(\zeta, s) = \int_{\mathcal{T}} \left(-\frac{\delta}{\delta \zeta(x)} \left\{ \mathcal{D}_{\zeta}[\mathcal{A}^-](x) \tilde{P}_0(\zeta, s) \right\} + \frac{1}{2} \mathcal{A}^+(x|\psi) \frac{\delta^2}{\delta \zeta(x)^2} \{ \tilde{P}_0(\zeta, s) \} \right) dx \quad (4.28)$$

where

$$\mathcal{D}_{\zeta}[\mathcal{A}^-](x) = \int_{\mathcal{T}} \zeta(y) \frac{\delta}{\delta \psi(y)} \{ \mathcal{A}^-(y|\psi) \} dy = \frac{d}{d\epsilon} \mathcal{A}^-(x|\psi + \epsilon \zeta) \Big|_{\epsilon=0}$$

can be thought of now as the functional analogue of a directional derivative of $\mathcal{A}^-(x|\psi)$ in the direction of the ‘fluctuation’ ζ , now a function. In Appendix D, we will see how this linear approximation can be used to study phenotypic clustering/evolutionary branching using spectral methods first introduced for various specific models by Rogers et al. (2012a) and further expanded upon in Rogers and McKane (2015).

Part III

Major Takeaways & Discussion

Chapter 5

A unified view of population dynamics

Not only is algebraic reasoning exact;
it imposes an exactness on the verbal
postulates made before algebra can start
which is usually lacking in the first verbal
formulations of scientific principles.

J.B.S. Haldane (1964)

So far, we have seen a lot of relatively abstract formalism for describing evolution in finite populations. Some concrete and hopefully illustrative examples are also presented in Appendix D. Have we gained anything from this (re)formulation? I show below that the **formalism developed stochastic equation for trait frequencies** I derived in part II naturally yields some ‘fundamental equations’ for evolutionary population dynamics that are very general, help us clearly understand how evolution operates in finite fluctuating populations, and recover well-known results in the infinite population limit, thus showing that I have generalized these infinite population results. We will also see rigorous, quantitative formulations of some well-known ideas in evolutionary biology, such as mutation as a ‘source’ of variation or drift causing continual erosion of variation.

5.1 Fundamental equations for evolution in finite populations

5.1.1 The fundamental equation for changes in type frequencies

Equation (3.14), which we derived in chapter 3, is a very general equation for how frequencies change over time in stochastic populations. To recap, we started with a population which can contain up to m different (fixed) types of individuals. The population as a whole is characterized by a vector $\mathbf{n}(t) = [n_1(t), \dots, n_m(t)]^T$ indexing the number of individuals of each type in the population. Changes of the population are through either birth or death of single individuals. On ecological grounds, we postulated the existence of a ‘system-size’ parameter $K > 0$ that leads to density-dependent growth and prevents the population from growing infinitely large. We then moved from numbers $\mathbf{n}(t)$ to population densities $\mathbf{x}(t) = \mathbf{n}/K$ by assuming that birth and death rates depend on population densities \mathbf{x} and not only on population numbers \mathbf{n} . In the regime where K is not too small (corresponding to ‘medium sized’ populations), we found a continuous description of how \mathbf{x} changes stochastically. We further assumed that each type has a per-capita birth rate $b^{(\text{ind})}(\mathbf{x})$, a per-capita death rate $d^{(\text{ind})}(\mathbf{x})$, and an additional term $\mu Q_i(\mathbf{x})$ representing additional birth of type i individuals due to mutations of other types. All three of these functions may in general vary in an arbitrarily complicated frequency-dependent manner (or be frequency independent constants, of course). We then moved from type densities $\mathbf{x}(t)$ to type frequencies $\mathbf{p}(t)$ and saw that under these assumptions, p_i , the frequency of the i^{th} type in the population, obeys:

$$\begin{aligned}
dp_i(t) = & \underbrace{\left[(w_i(\mathbf{x}) - \bar{w})p_i + \mu \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right\} \right]}_{\text{Infinite population predictions: selection-mutation balance for higher fitness}} dt \\
& - \underbrace{\frac{1}{K} \frac{1}{N_K(t)} \left[(\tau_i(\mathbf{x}) - \bar{\tau})p_i + \mu \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right\} \right]}_{\text{Directional noise-induced effects: selection-mutation balance for lower turnover rates}} dt \\
& + \underbrace{\frac{1}{\sqrt{K} N_K(t)} \left[\sqrt{A_i^+(\mathbf{x})} dW_t^{(i)} - p_i \sum_{j=1}^m \sqrt{A_j^+(\mathbf{x})} dW_t^{(j)} \right]}_{\text{Non-directional noise-induced effects due to stochastic fluctuations}}
\end{aligned} \tag{5.1}$$

where $N_K = \sum x_i = \sum n_i/K$ is the total population size scaled by K (and thus KN_K is the total population size), $w_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) - d_i^{(\text{ind})}(\mathbf{x})$ and $\tau_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) + d_i^{(\text{ind})}(\mathbf{x})$ are respectively the Malthusian fitness and per-capita turnover rate of the i^{th} type, and $A_i^+ := x_i \tau_i(\mathbf{x}) + \mu Q_i(\mathbf{x})$. Each $W_t^{(i)}$ is an independent one-dimensional Wiener process (Also see equation (E.7) in Appendix E for a more elegant representation of this term as an integral against a single Wiener process). Letting $K \rightarrow \infty$ in (5.1) recovers the replicator-mutator equation in the infinite population limit. The first term of (5.1) represents the direct effects of forces captured in classic deterministic models, and reflects a selection-mutation balance. However, finite populations experience a new directional force dependent on $\tau_i(\mathbf{x})$, the per-capita turnover rate of type i , that cannot be captured in infinite population models (Kuosmanen et al., 2022). Remarkably, this term acts in a way that is mathematically identical to the classical action of selection and mutation in infinite population models as captured by the first term in (5.1), but in the opposite direction - A higher relative τ_i leads to a decrease in frequency (Notice the minus sign before the second term in (5.1)).

5.1.2 The fundamental equation for the mean value of a type-level quantity

We can also calculate how the statistical mean value of any type-level quantity (see section 3.4) changes over time. Let f be any type level quantity, with value $f_i(t) \in \mathbb{R}$ for the i^{th} type. I allow the value f_i to possibly vary over time (for example, due to plasticity or a changing environment). By multiplying both sides of equation (5.1) by f_i and summing over all i (The same steps as going from (3.17) to (3.20)), we see that the statistical mean \bar{f} of the quantity in the population varies as:

$$d\bar{f} = \underbrace{\text{Cov}(w, f)dt}_{\text{Classical selection}} - \underbrace{\frac{1}{KN_K(t)}\text{Cov}(\tau, f)dt}_{\text{Noise-induced selection}} + \underbrace{\overline{\left(\frac{\partial f}{\partial t}\right)}dt}_{\text{Ecological feedbacks}} + \underbrace{M_{\bar{f}}(\mathbf{p}, N_K)dt}_{\text{Mutational effects}} + \underbrace{\frac{1}{\sqrt{KN_K(t)}}dW_{\bar{f}}}_{\text{Stochastic fluctuations}} \quad (5.2)$$

where all covariances are understood in the statistical sense (see section 3.4). Here,

$$M_{\bar{f}}(\mathbf{p}, N_K) := \mu \left(1 - \frac{1}{KN_K(t)}\right) \left(\sum_{i=1}^m f_i Q_i(\mathbf{p}) - \bar{f} \sum_{i=1}^m Q_i(\mathbf{p}) \right) \quad (5.3)$$

is a term capturing mutational effects/transmission fidelity that vanishes in the low mutation rate ($\mu \rightarrow 0$) limit, and

$$dW_{\bar{f}} := \sum_{i=1}^m (f_i - \bar{f}) \sqrt{A_i^+} dW_t^{(i)} \quad (5.4)$$

is a stochastic integral term describing un-directed stochastic fluctuations and vanishes upon taking expectation values (Also see equation (E.9) in Appendix E for a more elegant representation of this term as an integral against a single Wiener process). Note that the $1/\sqrt{K}$ factor outside this term in equation (5.2) means that these stochastic fluctuations also disappear in the infinite population ($K \rightarrow \infty$) limit.

Taking $K \rightarrow \infty$ in equation (5.2) recovers the standard Price equation as the infinite population limit (either (3.20) or (3.21) based on whether f_i varies with time; also see Rice, 2020 for a stochastic Price equation in a discrete-time setting).

Each term in equation (5.2) lends itself to a simple biological interpretation. The first term, $\text{Cov}(w, f)$, is well-understood in the classical Price equation and represents the effect of natural selection in the infinite population setting. In the stochastic Price equation (5.2), the effects of the second term of equation (5.1) decompose into a selection term $\text{Cov}(\tau, f)$ for reduced turnover rates and a transmission bias term that vanishes in the weak mutation ($\mu \rightarrow 0$) limit. Following Constable et al., 2016 and Week et al., 2021, I refer to the effect of the covariance term (the second term of equation (5.2)) as *noise-induced selection* since it occurs exactly analogously to classical natural selection (but for lower τ) and is induced purely by the finiteness of the population. Biologically, the $\text{Cov}(\tau, f)$ term (with a negative sign) describes a biasing effect due to differential turnover rates and can intuitively be understood as being similar to gambler's ruin in probability theory through the following reasoning: If a type i has a higher τ_i , it must have a higher birth and death rate and thus experiences more births and deaths in a given time interval than an otherwise equivalent species with a lower τ_i . More events mean greater demographic stochasticity, and types with a higher τ_i thus tend to be eliminated by a stochastic analog of selection because they experience more chance events (births and deaths) in a given time period and thus have greater stochastic fluctuations. This effect is less visible if the total population size is higher because larger populations generally experience less stochasticity, which is reflected in the $1/N_K$ factor in this term. Noise-induced selection for reduced turnover rates has also been interpreted as a selection for 'lower reproductive variance' (Gillespie, 1974; Wang et al., 2023). This latter interpretation is on purely mathematical grounds — the total turnover $\tau_i x_i$ determines the

variance of the distribution obtained as the solution to the SDE for population densities given by equation (3.9).

~~Note that types that tend to increase the total population size $KN_K(t)$ (such as altruists in evolutionary theory and mutualists in ecological communities) will reduce the magnitude of this effect compared to types that do not facilitate such an increase, such as cheaters and highly competitive species. Further, if altruists/mutualists act by reducing the death rate (rather than increasing the birth rate) of other individuals, their presence causes higher w and lower τ in the beneficiary individuals, both of which are favored by selection (but note that if they act by increasing the birth rate, they increase the magnitude of negative noise induced selection disfavouring the beneficiary individual), which explains why this effect preferentially favors mutualists in reversing the direction of deterministic selection in finite population models with fluctuating population sizes if interaction effects are on the death rate (McLeod and Day, 2019). Thus, selection for reduced turnover rate could help explain why cooperation often persists in fluctuating populations in laboratory experiments and finite population IBMs of social evolution that allow for fluctuating populations despite infinite population models predicting their extinction (Houchmandzadeh and Vallade, 2012; Sanchez and Gore, 2013; Houchmandzadeh, 2015; Chotibut and Nelson, 2015; Behar et al., 2016; McAvoy et al., 2018; McLeod and Day, 2019).~~

~~The fact that total population size controls the strength of noise induced selection also explains why cooperation is favored in the early transient period of population growth (Mebinger et al., 2010) when simulations are initiated from a small population size. In the early transient period, $N_K(t)$ is small, and the biasing effect of differential turnover rates is stronger, thus favoring cooperation. The fact that the entire term scales inversely with the total population size $KN_K(t)$ suggests that the effect of this force is weak for large populations, which explains why the persistence of cooperators is often only observed in restrictive sounding conditions such as quasi neutrality, timescale separation, or a weak selection + weak mutation limit (McLeod and Day, 2019). In all three of these cases, the effects of classical selection vanish, and thus the first term on the RHS of (5.2) becomes identically 0. It therefore no longer contributes to the trait frequency dynamics, allowing us to see the (otherwise weak) contributions of noise induced selection.~~

The third term of (5.2) is relevant in both finite and infinite populations whenever f_i can vary over time and represents feedback effects on the quantity f_i of the i^{th} species over

short ('ecological') time-scales. Such feedback could be through a changing environment, phenotypic/behavioral plasticity, or any manner of other 'ecological' phenomena. This is the term that captures eco-evolutionary feedback loops.

The fourth term of (5.2) is a transmission bias term, with a correction factor due to noise-induced selection. Finally, the last term of (5.2) describes the role of stochastic fluctuations. The contributions of this last term are 'directionless' due to the dW_t factors, and this term vanishes when we take a conditional expectation value over the underlying probability space. I denote this probabilistic expectation value operation by $\mathbb{E}[\cdot]$ to distinguish it from the statistical mean (3.12). Note that this expectation is conditioned on the initial state of the population, and thus $\mathbb{E}[\cdot]$ is really shorthand for $\mathbb{E}[\cdot \mid \mathbf{X}_0 = \mathbf{x}_0]$.

Two particularly interesting implications of (5.2) are realized upon ignoring mutations by setting $\mu = 0$ and then substituting $f = w$ and $f = \tau$. We first note that:

$$\text{Cov}(w, \tau) = \text{Cov}\left(b^{(\text{ind})}(\mathbf{x}) - d^{(\text{ind})}(\mathbf{x}), b^{(\text{ind})}(\mathbf{x}) + d^{(\text{ind})}(\mathbf{x})\right) \quad (5.5)$$

$$= \sigma_{b^{(\text{ind})}(\mathbf{x})}^2 - \sigma_{d^{(\text{ind})}(\mathbf{x})}^2 \quad (5.6)$$

It is important to remember once again that just like the statistical mean, the statistical variance $\sigma_f^2(t)$ of a type-level quantity f is a random variable obtained by calculating the variance of the quantity *in the population* at time t , and is not to be confused with the probabilistic/ensemble variance obtained by calculating the variance of a quantity *over different realizations* of the stochastic process (see section 3.4). Upon substituting $f = w$ in (5.2) and taking expectations over the underlying probability space, we obtain:

$$\mathbb{E}\left[\frac{dw}{dt}\right] = \underbrace{\mathbb{E}\left[\sigma_w^2\right]}_{\text{Fisher's fundamental theorem}} - \underbrace{\mathbb{E}\left[\frac{\sigma_{b^{(\text{ind})}}^2 - \sigma_{d^{(\text{ind})}}^2}{KN_K(t)}\right]}_{\text{Noise-induced selection}} + \underbrace{\mathbb{E}\left[\overline{\frac{\partial w}{\partial t}}\right]}_{\text{Eco-evolutionary feedbacks to fitness}} \quad (5.7)$$

Taking $K \rightarrow \infty$ in (5.7) recovers a well-known equation in population genetics upon noting that the process tends to a deterministic process as $K \rightarrow \infty$, as noted in section 3.6, and thus the expectation value in the infinite population case is superfluous. The first term, σ_w^2 , is the subject of Fisher's fundamental theorem (Fisher, 1930; Price, 1972; Frank and Slatkin, 1992; Kokko, 2021). The second term of equation (5.7) is a manifestation of noise-induced selection and vanishes in the infinite population limit, and is thus particular to finite populations.

Finally, the last term arises in both finite and infinite populations whenever w_i can vary over time (Frank and Slatkin, 1992; Kokko, 2021; Baez, 2021), be it through frequency-dependent selection, phenotypic plasticity, varying environments, or other ecological mechanisms, and represents feedback effects on the fitness w_i of the i^{th} species over short ('ecological') time-scales. The fact that Fisher (1930) appears to have been rather vague and dismissive of this feedback has led to much discussion, debate, and confusion about the interpretation, importance, and implications of his 'fundamental theorem' (see Kokko, 2021 and sources cited therein).

Carrying out the same steps with $f = \tau$ in (5.2) yields a new equation due to Kuosmanen et al. (2022). The result is an analog of Fisher's fundamental theorem for the turnover rates, and reads:

$$\mathbb{E} \left[\frac{d\bar{\tau}}{dt} \right] = \underbrace{\mathbb{E} \left[\sigma_{b^{(\text{ind})}}^2 - \sigma_{d^{(\text{ind})}}^2 \right]}_{\text{Classical selection effects}} - \underbrace{\mathbb{E} \left[\frac{\sigma_\tau^2}{KN_K(t)} \right]}_{\text{Noise-induced selection effects}} + \underbrace{\mathbb{E} \left[\overline{\frac{\partial \tau}{\partial t}} \right]}_{\text{Eco-evolutionary feedbacks to } \tau_i} \quad (5.8)$$

The implications of this equation have been extensively discussed in Kuosmanen et al., 2022, which is where I refer the interested reader.

5.1.3 The fundamental equation for the variance of a type-level quantity

Equation (5.2) is a general equation for the mean value of an arbitrary type level quantity f in the population. In many real-life situations, especially those pertaining to finite populations, we are interested in not just the mean, but also the variance of a type-level quantity. In Appendix C, I show that under the same assumptions used to derive (5.1), the statistical variance of any type level quantity f obeys

$$\begin{aligned} d\sigma_f^2 &= \text{Cov} (w, (f - \bar{f})^2) dt - \frac{1}{KN_K} [\bar{\tau}\sigma_f^2 + 2\text{Cov} (\tau, (f - \bar{f})^2)] dt \\ &\quad + 2\text{Cov} \left(\frac{\partial f}{\partial t}, f \right) dt + M_{\sigma_f^2}(\mathbf{p}, N_K) dt + \frac{1}{\sqrt{KN_K(t)}} dW_{\sigma_f^2} \end{aligned} \quad (5.9)$$

where

$$M_{\sigma_f^2}(\mathbf{p}, N_K) := \mu \left[\left(1 - \frac{2}{KN_K}\right) \left(\sum_{i=1}^m (f_i - \bar{f})^2 Q_i(\mathbf{p}) \right) + \sigma_f^2 \left(1 - \frac{1}{KN_K}\right) \sum_{i=1}^m Q_i(\mathbf{p}) \right] \quad (5.10)$$

is a mutational term that vanishes in the $\mu \rightarrow 0$ limit and

$$dW_{\sigma_f^2} := \sum_{i=1}^m (f_i - \bar{f})^2 \sqrt{A_i^+} dW_t^{(i)} \quad (5.11)$$

is a stochastic integral term measuring the (non-directional) effect of stochastic fluctuations that vanishes upon taking an expectation over the probability space (Also see equation (E.11) in Appendix E for a more elegant representation of this term as an integral against a single Wiener process). In the case of one-dimensional quantitative traits, an infinite-dimensional version of (5.9) has recently been rigorously derived (Week et al., 2021) using measure-theoretic tools under certain additional assumptions (See equation (21c) in Week et al., 2021). Taking expectations over the probability space in (5.9) and substituting mutation as acting via a Gaussian kernel also recovers equations previously derived (Débarre and Otto, 2016) in the context of evolutionary branching in finite populations as a special case. An infinite population ($K \rightarrow \infty$) version of equation (5.9) also appears in Lion, 2018.

Once again, terms of equation (5.9) lend themselves to straightforward biological interpretation. The quantity $(f_i - \bar{f})^2$ is a measure of the distance of f_i from the population mean value \bar{f} , and thus covariance with $(f - \bar{f})^2$ quantifies the type of selection operating in the system: A negative correlation is indicative of stabilizing selection, and a positive correlation is indicative of disruptive (*i.e.* diversifying) selection. An extreme case of diversifying selection for fitness occurs if the mean fitness is at a local minimum for fitness but $f_i \not\equiv \bar{f}$ (*i.e.* the population still exhibits some variation in f). In this case, if the variation in f is associated with a variation in fitness, then $\text{Cov}(w, (f - \bar{f})^2)$ is strongly positive and the population experiences a sudden explosion in variance, causing the emergence of polymorphism in the population. If $\text{Cov}(w, (f - \bar{f})^2)$ is still positive after the initial emergence of multiple morphs, evolution eventually leads to the emergence of stable coexisting polymorphisms in the population - this phenomenon is a slight generalization of the idea of evolutionary branching that occurs in frameworks such as adaptive dynamics (Geritz et al., 1998), and a special case of the exact equation (5.9) has been used before to study evolutionary branching in finite populations (Débarre and Otto, 2016).

The $\text{Cov}(\partial f/\partial t, f)$ term once again represents the effect of eco-evolutionary feedback loops due to rapid change in f that is not solely due to changes in \mathbf{p} . The $M_{\sigma_f^2}(\mathbf{p}, N_K)$ term quantifies the effect of mutations on the variance of f . Note that each $Q_i(\mathbf{p}) \geq 0$ by its definition in (3.10) and thus $\sum_i Q_i(\mathbf{p}) > 0$ if there are any mutational effects (and = 0 otherwise). Furthermore, the total population size $KN_K > 2$ for most interesting evolutionary questions. Thus, from (5.10), it is clear that when $\mu > 0$ (*i.e.* there is mutation in the system), we have $M_{\sigma_f^2}(\mathbf{p}, N_K) > 0$, meaning that mutations always increase the variance of f in the population. Since σ_f^2 is a measure of the amount of variation of f in the population, the $M_{\sigma_f^2}(\mathbf{p}, N_K)$ term is thus a rigorous formulation of the well-known idea that mutation acts as a ‘source’ of variation for evolution to act on.

The $\bar{\tau}\sigma_f^2$ term quantifies the loss of variation due to stochastic extinctions (*i.e.* demographic stochasticity) and thus represents the classic effect of neutral genetic (or ecological) drift in finite populations. To see this, it is instructive to consider the case in which this is the only force at play. Let us imagine a population of asexual organisms in which each f_i is simply a label or mark arbitrarily assigned to individuals in the population at the start of an experiment/observational study and subsequently passed on to offspring - for example, a neutral genetic tag in a part of the genome that experiences a negligible mutation rate. Let us set $\mu = 0$ so that the labels cannot change between parents to offspring. This means that we have $M_{\sigma_f^2}(\mathbf{p}, N_K) \equiv 0$. Further, since the labels are arbitrary and have no effect whatsoever on the biology of the organisms, each label has the same fitness $w_i \equiv w$ and per-capita turnover $\tau_i \equiv \tau$, and thus $\bar{w} = w$ and $\bar{\tau} = \tau$. Note that since every type has the same fitness and turnover rate, we have $\text{Cov}(w, (f - \bar{f})^2) \equiv \text{Cov}(\tau, (f - \bar{f})^2) \equiv 0$. Since the labels do not change over time, we also have $\text{Cov}(\partial f/\partial t, f) = 0$. From (5.9), we see that in this case, the variance changes as

$$d\sigma_f^2 = -\frac{\tau\sigma_f^2}{KN_K}dt + \frac{1}{\sqrt{KN_K(t)}}dW_{\sigma_f^2} \quad (5.12)$$

Where I have written $\bar{\tau} = \tau$, since every type has the same turnover rate. On taking expectations, the second term on the RHS vanishes and we see that the expected variance in the population obeys

$$\frac{d\mathbb{E}[\sigma_f^2]}{dt} = -\left(\mathbb{E}\left[\frac{\tau}{KN_K}\right]\right)\mathbb{E}[\sigma_f^2] \quad (5.13)$$

where I have decomposed the expectation of the product on the RHS into a product of expectations, which is admissible since the label f is completely arbitrary and thus independent

of both $\bar{\tau}$ and $N_K(t)$. Equation (5.13) is a simple linear ODE and has the solution

$$\mathbb{E}[\sigma_f^2](t) = \sigma_f^2(0)e^{-\mathbb{E}[\frac{\tau}{KN_K}]t} \quad (5.14)$$

which tells us that the expected diversity (variance) of labels in the population decreases exponentially over time. The rate of loss is $\mathbb{E}[\tau(KN_K)^{-1}]$, and thus, populations with higher turnover rate τ and/or lower population size KN_K lose diversity faster. This is because populations with higher τ experience more stochastic events per unit time (a gambler's ruin type scenario), while extinction is 'easier' in smaller populations because a smaller number of deaths is required to eliminate a label from the population completely. Note that *which* labels/individuals are lost is entirely random (since all labels are arbitrary), but nevertheless, labels tend to be stochastically lost until only a small number of labels remain in the population.

5.2 A stochastic field theory for quantitative traits

In chapter 4, I formulated a 'field equation' for the evolution of one-dimensional quantitative traits in populations. To recap, given a one-dimensional quantitative trait that takes values in a trait space $\mathcal{T} \subseteq \mathbb{R}$, I defined the set $\mathcal{M}(\mathcal{T}) = \{\sum_{i=1}^n \delta_{x_i} \mid n \in \mathbb{N}, x_i \in \mathcal{T}\}$, where each δ_{x_i} is a Dirac mass centered at x_i . I then formulated a model for the evolutionary dynamics of finite fluctuating populations of individuals that vary in the relevant quantitative trait as an $\mathcal{M}(\mathcal{T})$ valued birth-death process. The population at time t is characterized by a randomly varying density distribution ('stochastic field') $\nu^{(t)} \in \mathcal{M}(\mathcal{T})$ such that for any subset $A \subseteq \mathcal{T}$, the number of individuals that have trait value in A is given by integrating $\nu^{(t)}$ over A . The change of this field is determined entirely by two functionals, $b(x|\nu)$ and $d(x|\nu)$ from $\mathcal{T} \times \mathcal{M}(\mathcal{T})$ to $[0, \infty)$, that respectively describe the birth rate and death rate of individuals of type x in a population ν . Under the assumption that there exists a suitable system size parameter $K > 0$, I moved from the space of 'number' distribution functions $\mathcal{M}(\mathcal{T})$ to the space of 'density' distribution functions $\mathcal{M}_K(\mathcal{T}) = \{\sum_{i=1}^n \delta_{x_i}/K \mid n \in \mathbb{N}, x_i \in \mathcal{T}\}$ via a functional analog of the system size expansion. By appropriately rescaling the birth and death rate functionals, I then determined that $P(\phi, t)$, the probability that the population is described by the distribution $\phi \in \mathcal{M}_K(\mathcal{T})$ at time t , (approximately) satisfies the 'stochastic

field equation':

$$\frac{\partial P}{\partial t}(\phi, t) = \int_{\mathcal{T}} \left[-\frac{\delta}{\delta\phi(x)} \{ [\phi(x)w(x|\phi) + \mu Q(x|\phi)] P(\phi, t) \} + \frac{1}{2K} \frac{\delta^2}{\delta\phi(x)^2} \{ [\phi(x)\tau(x|\phi) + \mu Q(x|\phi)] P(\phi, t) \} \right] dx \quad (5.15)$$

where $w(x|\phi)$, $\tau(x|\phi)$, and $Q(x|\phi)$ are functionals that respectively describe the Malthusian fitness, per-capita turnover rate, and birth rate due to mutations (with mutation rate μ) of type $x \in \mathcal{T}$ individuals (now a continuous variable) in a population ϕ . We then saw that this equation yields some well-known frameworks of quantitative genetics in the infinite population ($K \rightarrow \infty$) limit, thus illustrating consistency with known theories.

If the ‘intuitive’ version of Itô’s formula for $\mathcal{M}_K(\mathcal{T})$ valued stochastic processes obtained by informally ‘taking the limit’ $m \rightarrow \infty$ in the Itô’s formula for m -dimensional stochastic processes holds, then the exact same steps carried out in Appendices B and C will ‘go through’ essentially unchanged for quantitative traits and yield the equations obtained by simply informally taking $m \rightarrow \infty$ in (5.1), (5.2) and (5.9) as the ‘fundamental equations’ for the evolution of quantitative traits. In their supplementary sections S1 and S2, Week et al. (2021) have recently proposed exactly the Itô formula we would need for $L^2(\mathbb{R}, m)$ valued processes in which the $\mathcal{A}^\pm(x|\phi)$ (in my notation as defined in chapter 4) are Hilbert-Schmidt operators via a heuristic Itô multiplication table [and](#). They have also shown that their heuristics are equivalent to the rigorous infinite-dimensional stochastic calculus proposed by Da Prato and Zabczyk (2014) for more general Hilbert space valued processes¹. Furthermore, using these ‘spacetime white noise heuristics’ together with a particular functional form for (weak) solutions to an SPDE describing populations with Gaussian trait distributions, Week et al. (2021) have arrived at precisely the infinite-dimensional version of equations (5.2) and (5.9) obtained by ‘taking the limit’ $m \rightarrow \infty$ in (5.2) and (5.9) for the special case in which the type-level quantity f is simply the value of the quantitative trait under study from a completely different starting point(!). Equation (5.15) is the ‘functional Fokker-Planck view’ of the stochastic processes for the evolutionary ecology of quantitative traits that Week et al. (2021) study from the ‘SPDE view’. Indeed, the two approaches are exactly complementary,

¹If this sentence sounds like abstract gibberish to you, don’t worry too much about the details -the essence is that Week et al. (2021) have proven the ‘correct’ formula we need holds whenever $b(x|\nu)$, $d(x|\nu)$ and $\mathcal{M}_K(\mathcal{T})$ together fulfill certain technical requirements. My heuristics here are only a ‘first step’ and have focused on accessibility and actually formulating the problem in formal terms rather than on mathematical propriety, and thus make no attempt to verify whether and when these technical requirements are satisfied. The required properties will likely be fulfilled only for some ‘nice’ subspace of $\mathcal{M}_K(\mathcal{T})$ and not the entire space, and future, more rigorous work should focus on figuring out which functions are included in this subspace.

and my formulation in chapter 4 provides an alternate method of attack for the study of quantitative traits in finite fluctuating populations that may be more appropriate for some particular problems, while Week et al.’s (2021) approach may be more appropriate for others.

More generally, the well-known intimate relation between Itô SDEs and Fokker-Planck equations (see section 2.1) is extremely useful and has been extensively exploited in the literature (Van Kampen, 1981; Øksendal, 1998; Gardiner, 2009), because some problems are much more easily attacked using tools from probability theory that operate on Itô processes, whereas others are more amenable to study using tools from PDE theory that operate on the Fokker-Planck equations. In the infinite-dimensional case, working with the field theory through functional Fokker-Planck equations such as (5.15) allows us to use tools with an essentially ‘differential equation’ flavor such as spectral methods (Rogers et al., 2012a; Rogers and McKane, 2015; See also Appendix D for a general model-independent pedagogical example), fast-mode (‘adiabatic’) elimination/slow manifold approximation (Constable et al., 2013; Parsons and Rogers, 2017), and WKB style expansions to detect extreme events (Rogers et al., 2012a; Assaf and Meerson, 2017; though it bears noting that WKB theory is closely related to large deviation principles that occur in the study of Markov processes and often come in a distinctly ‘probabilistic’ flavor, see Dembo and Zeitouni, 1998). On the other hand, many biologically interesting questions are much easier to approach from the branching processes/SPDE side using tools that have an essentially ‘probabilistic’ flavor such as duality (Dawson, 1975; Greenman, 2020), infinitesimal generators/semi-group theory (Ethier and Kurtz, 1986; Finkelshtein et al., 2012), and the martingale problem (Ethier and Kurtz, 1986; Dawson et al., 2000; Champagnat et al., 2006; Champagnat et al., 2008). Rigorously establishing and systematically exploiting relations between functional Fokker-Planck equations of the form (5.15) and the sort of general SPDEs studied by Week et al. (2021) for measure-valued branching processes could thus prove very fruitful for developing more integrative eco-evolutionary theory since it allows us to seamlessly transition between complementary views of the same object to suit the problem at hand².

To the best of my knowledge, a general formulation of stochastic field equations for the population dynamics of quantitative traits from the functional Fokker-Planck side in the manner I have carried out here has not been done before. Similar equations have been formulated for some specific stochastic models of quantitative trait evolution (Rogers et al., 2012a;

²The physics literature has long recognized this. Physicists regularly exploit the connections between stochastic/quantum field equations and SPDEs, often studying SPDEs in completely ‘classical’ systems using tools and language from quantum field theory (Hochberg et al., 1999; Baez and Biamonte, 2018).

Rogers and McKane, 2015) and population ecology of size-structured communities (O'Dwyer et al., 2009). Stochastic field equations are also known in mathematical neurobiology (Buice and Cowan, 2007; Bressloff, 2010; Coombes et al., 2014), and have recently been proposed in a model of collective motion (Ó Laighléis et al., 2018). Broadly similar deterministic field theoretic approaches have also been proposed to study bio-geography and spatial biodiversity patterns such as the species-area relationship (O'Dwyer and Green, 2010; But also see Grilli et al., 2012 - the actual details of the calculations presented in O'Dwyer and Green, 2010 are not correct). The ubiquity of such approaches in diverse areas of theoretical biology suggests that the formal systematic study and analysis of such equations could have wide-spread applications.

Currently, (stochastic) field equations are primarily used by physicists working in areas such as statistical field theory and quantum field theory, and are largely attacked using ingenious and often somewhat system-specific tools that may not necessarily be mathematically well understood (Carmona and Rozovskii, 1999; but see Bogachev et al., 2015 for a rigorous treatment of some infinite-dimensional Fokker-Planck equations). Equation (5.15) opens up the study of quantitative trait dynamics in finite fluctuating populations to analysis using some of these tools - for example, a careful reading of the literature provides many hints (albeit couched in physics language) for how (quantum) field theoretic ideas such as the Fock space representation (Del Razo et al., 2022), ‘loop diagrams/expansions’ (Hochberg et al., 1999; Dodd and Ferguson, 2009), and the often closely related path integral formalism (Doi, 1976; Peliti, 1985; Hochberg et al., 1999; Chow and Buice, 2015; Weber and Frey, 2017), can all be co-opted to coax biological insights from the processes modeled by equation (5.15). The formalism developed in Chapter 4 is intended to encourage the use of such techniques and make some of the models and ideas more accessible to theorists without formal training in mathematical areas such as measure theory and functional analysis.

Importantly, the formalism I develop here in terms of functional Fokker-Planck equations likely does *not* carry over to the study of higher dimensional quantitative traits (or populations which vary in two or more one-dimensional quantitative traits) because these processes are routinely badly behaved in higher dimensions: In particular, a probability density $P(\phi, t)$ with respect to the reals frequently does not even exist in higher dimensions if one has any biologically non-trivial features such as interactions between types (Fleming and Viot, 1979; Walsh, 1986; Konno and Shiga, 1988; Evans and Perkins, 1994; Etheridge, 2000), rendering equation (5.15) entirely meaningless. My (admittedly limited) understanding is that such

processes are also rather difficult to study in two or more dimensions from the SPDE side for similar reasons (Etheridge, 2000), where the lack of a density leads to distribution-valued (rather than function-valued) solutions at best and thus demands a considerable amount of advanced functional analysis to establish even ‘basic’ existence/uniqueness results (Walsh, 1986; Carmona and Rozovskii, 1999; Prévôt and Röckner, 2007; Liu and Röckner, 2015; Balan, 2018). It may well be the case that concrete biologically useful progress in this direction requires new mathematics altogether, a situation increasingly also encountered in other areas of mathematical biology (Vittadello and Stumpf, 2022). Just as theoretical physics has done in the past, the pursuit of general mathematical descriptions of biological theory may therefore also inspire new ideas that are of independent interest to ‘pure’ mathematicians (Cohen, 2004).

Chapter 6

Discussion and Outlook

Writing in the last months of this millennium, it is clear that the prime intellectual task of the future lies in constructing an appropriate theoretical framework for biology

Sydney Brenner (1999)

In this thesis, we have seen how stochastic birth-death processes can be used to construct and analyze mechanistic individual-based models for the dynamics of finite populations. In doing so, we have also seen that various well-known equations of evolutionary dynamics can be recovered in the infinite population size limit. In the finite-dimensional case, the infinite population limit corresponds to the equations of population genetics and evolutionary game theory. In the infinite-dimensional case, we instead obtain the equations of quantitative genetics. In both cases, the mean value of the trait in the population changes according to an equation resembling the Price equation. My derivation further highlights the natural connections between various equations of population dynamic and extends these similarities to finite, fluctuating populations (Table 6.1).

Number of possible distinct types/trait variants (m)	State Space description	Model parameters	Mesoscopic description	Infinite population limit
$m = 1$ (Identical individuals)	$[0, 1, 2, 3, \dots]$ (Population size)	Two non-negative functions, $b(n)$ and $d(n)$, describing the birth and death rate of individuals when the population size is n	Univariate Fokker-Planck equation (one-dimensional SDEs)	Dynamics of populations of identical individuals
$1 < m < \infty$ (Discrete traits)	$[0, 1, 2, 3, \dots]^m$ (Number of individuals of each trait variant)	$2m$ non-negative functions, $b_i(\mathbf{n})$ and $d_i(\mathbf{n})$ (for $1 \leq i \leq m$) describing the birth and death rate of trait variant i when the population is $\mathbf{n} = [n_1, \dots, n_m]$	Multivariate Planck equation (m -dimensional SDEs)	Fokker- Planck equation Lotka-Volterra competition Quasispecies equation Price equation (discrete traits)
$m = \infty$ (Quantitative traits)	$\left\{ \sum_{i=1}^n \delta_{x_i} \mid n \in \mathbb{N}, x_i \in \mathcal{T} \subseteq \mathbb{R} \right\}$ (Each Dirac mass δ_{x_i} is an individual with trait value x_i in the trait space \mathcal{T})	Two non-negative functionals $b(x \nu)$ and $d(x \nu)$ describing the birth and death rate of trait variant x when the population is the (stochastic) field/function ν	Functional Fokker-Planck equation/Field theory (SPDEs)	Kimura's continuum-of-alleles model Sasaki and Dieckmann's (2011) Oligomeric Dynamics Wickman et al.'s (2022) Trait Space Equations for intraspecific trait variation Gradient Dynamics Price equation (quantitative traits)

Table 6.1: Summary of the various birth-death processes studied in this thesis

Stochastic models in biology often exhibit behaviors that are markedly different from their deterministic limits (Jafarpour et al., 2017; Boettiger, 2018; Jhawar et al., 2020; Coomer et al., 2022). Since real-life populations are stochastic and finite, it is thus imperative that modellers work with stochastic models instead of their deterministic limits, lest they risk missing important phenomena that are unique to stochastic systems (Black and McKane, 2012; Hastings, 2004; Shoemaker et al., 2020; Schreiber et al., 2022). In the context of population biology, the finiteness of populations is an important source of stochasticity that can lead to behavior not captured in corresponding infinite population limits (Black and McKane, 2012; Rogers et al., 2012a; Débarre and Otto, 2016; DeLong and Cressler, 2023). When actually incorporating this stochasticity into eco-evolutionary theory, it is important to build up the stochastic theory from first principles instead of adding noise to existing deterministic models in an ad-hoc fashion, since the latter procedure can very easily lead to ~~inconsistent models~~ ill-behaved models with un-biological properties (Strang et al., 2019). Indeed, several theorists have called for a ‘bottom-up’ reformulation of eco-evolutionary dynamics from the first principles of stochastic birth-death processes on the grounds that such a formulation is more fundamental and mechanistic (Metcalf and Pavard, 2007a; Lambert, 2010; Geritz and Kisdi, 2012; Doebeli et al., 2017).

In this thesis, I present a mathematical framework for such a reformulation. Part II of this thesis develops a formalism for both populations of individuals that vary in discrete characters (Chapter 3) as well as populations of individuals that vary in a single one-dimensional quantitative character (Chapter 4). The central result of this reformulation is a series of stochastic differential equations derived in Chapter 5. To begin, I derive an equation for change in type frequencies in the population (equation (5.1)) that generalizes the replicator-mutator equation to finite, fluctuating, closed populations evolving in continuous time. From this, I show how one can derive an equation for changes in the population mean of an arbitrary type-level quantity (equation (5.2)) that generalizes the dynamic, continuous time version of the Price equation to finite, fluctuating populations. I also derive an equation for changes in the population variance of an arbitrary type-level quantity (equation (5.9)) in such populations that generalizes a recent infinite population formulation of the same (Lion, 2018). My work thus generalizes some fundamental formal structures of eco-evolutionary population dynamics to finite, fluctuating populations. ~~In particular, my equations generalize the unifying formalism described in Lion, 2018 to finite, fluctuating populations—taking $K \rightarrow \infty$ in equations (5.1), (5.2), and (5.9) recover equations (6), (11), and (14) in Lion, 2018 respectively as their infinite population limits~~ The equations I derive deal with

biologically important quantities in a ‘coarse-grained’ manner, lend themselves to simple biological interpretation, and are very general. They thus fulfill the criteria to be called ‘fundamental theorems’ (in the sense of Queller, 2017) and present a ‘unifying perspective’ (in the sense of Lion, 2018) for the dynamics of finite populations. While my equations thus recover standard results such as the Price equation and the replicator-mutator equation in the infinite population limit, they also predict that these results do not completely capture the behavior of finite populations.

In Chapter 4, I also postulate a ‘stochastic field theory’ approach to modelling the evolution of quantitative traits in finite, fluctuating populations. I then show that this approach is consistent with known frameworks in theoretical population biology, reproducing results from quantitative genetics in the infinite population limit. This formulation also highlights a (somewhat heuristic) approach to the study of spacetime stochastic processes and related mathematical objects that avoids measure-theoretic tools and may be of independent interest to applied mathematicians. Broadly similar stochastic field theoretic approaches have also been proposed in mathematical neurobiology (Buice and Cowan, 2007; Bressloff, 2010; Coombes et al., 2014) and models of collective motion (Ó Laighléis et al., 2018). I conjecture that the stochastic ‘fundamental equations’ I derive in section 5.1 for discrete traits should also hold for quantitative traits under mild assumptions on the trait space of the quantitative traits, though no proof is attempted in this thesis. We can, however, take encouragement from the observation that Week et al., 2021 have already derived some special cases of these equations¹ for quantitative traits using an approach grounded in the theory of measure-valued branching processes under certain technical assumptions on the stochastic process under study, the most biologically important assumption being that the traits are normally distributed in the population. My formalism and Week et al., 2021’s formalism are complementary to each other, and reflect a deep duality between related mathematical structures that is well-appreciated in the applied mathematics literature (see section 5.2). These two complementary approaches to modeling stochastic processes have long been recognized in population genetics (Lambert, 2006), where similar analogs of the broad idea of the approach for models of fixed total population size go by the names ‘branching processes approach’ (Week et al., 2021’s formalism; In the classic literature, seen in work like Haldane, 1927; Fisher, 1931) and ‘diffusion theory approach’ (My formalism; In the

¹In particular, equations (21b) and (21c) in Week et al., 2021 are precisely the $m \rightarrow \infty$ limit of my equations for changes in the mean value of a type-level quantity (equation 5.2) and changes in the variance of a type-level quantity (equation 5.9) respectively for the special case in which the type-level quantity is the value of the quantitative trait being studied. See section 5.2 for a more detailed technical discussion.

classic literature, seen in work like S. Wright, 1931; Kimura, 1957).

6.1 Noise-induced selection is a generic outcome of evolution in finite, fluctuating populations

Several specific finite population models illustrate that evolution in finite population can proceed in a direction different from that predicted by the infinite population limit (Parsons et al., 2010; Melbinger et al., 2010; Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Chotibut and Nelson, 2015; Débarre and Otto, 2016; Behar et al., 2016; Constable et al., 2016; Veller et al., 2017; Abu Awad and Coron, 2018; Parsons et al., 2018; McAvoy et al., 2018; McLeod and Day, 2019). Indeed, in some stochastic models, the outcome of finite population models can be exactly opposite to that of the infinite population limit, a phenomenon sometimes referred to as ‘reversing the direction of deterministic evolution’ (Constable et al., 2016; McLeod and Day, 2019; Wang et al., 2023). These effects have been recognized to play an important role in specific finite population models of epidemiology (Kogan et al., 2014; Humplík et al., 2014; Parsons et al., 2018; Day et al., 2020), heterogamety (Veller et al., 2017; Saunders et al., 2018), life-history evolution (Gillespie, 1974; Kuosmanen et al., 2022), and social evolution (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Chotibut and Nelson, 2015; Constable et al., 2016; McLeod and Day, 2019; Wang et al., 2023).

Chapter 5 formulates a set of stochastic differential equations derived from our general birth-death formalism which reveal that the results of all these studies can be explained through a single set of equations. In particular, the equations of section 5.1 generically predict a directional evolutionary force — noise-induced selection — acting on variation in per-capita turnover rate τ in the population and favoring types whose per-capita turnover rate is lower than that of the population average. Noise-induced selection is only seen in finite populations, is seen whenever there is differential turnover rate τ in the system, and arises due to different types of individuals in the population experiencing a different number of stochastic events (birth and death) in a given time interval. This thesis is, to the best of my knowledge, the first time that noise-induced selection has been explicitly formulated and derived in full generality using a general ‘model-independent’ evolutionary framework such as the Price equation (or my generalization thereof). My results thus unify several previous studies focused on specific systems under a single conceptual banner and show that their predictions are a generic outcome in finite, fluctuating populations of any nature as long as

there is some heritable variation in per-capita turnover rates.

Populations consisting mostly of types that tend to increase the *total* population size $KN_K(t)$ (such as altruists in evolutionary theory and mutualists in ecological communities) will experience a reduced magnitude of noise-induced selection acting in the system compared to Populations consisting mostly of types that do not facilitate such an increase, such as cheaters and highly competitive species. Further, if altruists/mutualists act by reducing the death rate (rather than increasing the birth rate) of other individuals, their presence causes higher w and lower τ in the beneficiary individuals, both of which are favored by evolution through classical selection and noise-induced selection respectively (but note that if they act by increasing the birth rate, they increase the magnitude of negative noise-induced selection disfavouring the beneficiary individual). This is why the direction of deterministic selection is specifically reversed to favor mutualists in finite population models with fluctuating population sizes if interaction effects are on the death rate (McLeod and Day, 2019). Thus, selection for reduced turnover rate could help explain why cooperation often persists in fluctuating populations in laboratory experiments (Sanchez and Gore, 2013) and finite population IbMs (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Chotibut and Nelson, 2015; Behar et al., 2016; McAvoy et al., 2018; McLeod and Day, 2019) of social evolution despite infinite population models predicting their extinction.

The fact that total population size controls the strength of noise-induced selection also explains why cooperation is favored in the early transient period of population growth (Mebbingen et al., 2010) when simulations are initiated from a small population size — In the early transient period, $N_K(t)$ is small, and the biasing effect of differential turnover rates is stronger, thus favoring cooperation. The fact that the entire term scales inversely with the total population size $KN_K(t)$ suggests that the effect of this force is weak for large populations, which explains why the persistence of cooperators is often only observed in restrictive sounding conditions such as quasi-neutrality, timescale separation, or a weak selection + weak mutation limit (McLeod and Day, 2019). In all three of these cases, the effects of classical selection vanish, and thus the first term on the RHS of (5.2) becomes identically 0. It therefore no longer contributes to the trait frequency dynamics, allowing us to see the (otherwise weak) contributions of noise-induced selection.

6.2 Noise-induced selection has concrete implications for finite populations

The existence of noise-induced selection directly implies (from equation (5.1) or (5.2)) that evolution is not expected to maximize fitness in finite populations even if fitness is entirely frequency independent as long as there is some (heritable) variation in the turnover rates τ_i , further underscoring the now well-appreciated point that the view of evolution as ‘climbing a hill’ on a fitness landscape and thereby maximizing fitness is rather limited (Grodwohl, 2017). Further, from our generalization of Fisher’s fundamental theorem (equation (5.7)), it is clear that unlike in infinite populations, the mean fitness of a finite population can systematically decrease even without any frequency-dependence in fitness, and perhaps more surprisingly, can change even when there is no standing variation in the mean fitness as long as there is some variation in either the birth rates or the death rates.

The equations of section 5.1 also imply that for the evolution of a trait to be *truly* neutral in finite populations (in the sense of all m types in a system having equal fixation/extinction probability if we start with an initial state in which every type has frequency $1/m$), it is not sufficient for the trait in question to be neutral with respect to fitness w . Instead, we also require the trait to be neutral with respect to turnover rate τ . Though this is clear from examining equations (5.1) or (5.2), I also illustrate a deviation from neutrality explicitly for the $m = 2$ case using a simple resource-competition model in Appendix D.3. Indeed, in ecological models, previous work in the ‘quasi-neutral’ regime has shown that in finite, fluctuating populations, the equal growth rate of types is not sufficient to ensure equal fixation probabilities and that there is a slight biasing for types with lower turnover rates, sometimes interpreted as a selection ‘for longevity’ (Lin et al., 2012; Oliveira and Dickman, 2017; Balasekaran et al., 2022). In models of evolutionary game theory in fluctuating finite populations, individuals with lower death rates have higher fixation probability even when growth rates are equalized (Huang et al., 2015; Czuppon and Traulsen, 2018). Similarly, models of cell cycle dynamics find that selection favors cell types that periodically arrest their cell cycle relative to non-arresting cells even when their growth rates are equal (Wodarz et al., 2017). In the language of the birth-death formalism I develop in this thesis, all of these studies equalize the growth rates w of competing types but allow the turnover to vary (by arresting the cell cycle and thus reducing turnover, for example), thus allowing noise-induced selection for reduced turnover to operate in the system. Since noise-induced

selection is directional, the seemingly small systematic deviations from neutrality exhibited by these models can have significant consequences over long evolutionary times (For example, see Veller et al., 2017). My derivations show analytically that such deviations from neutrality in finite populations are a generic phenomenon explained by noise-induced selection and should be expected whenever there is variation in turnover rates. Selection on turnover rates also leads to insights on life-history evolution, and these insights have been extensively reported in a recent pre-print (Kuosmanen et al., 2022) ~~that independently arrives at my equations for type frequencies (equation (5.1)) and the change of mean fitness and turnover in the population (equations (5.7) and (5.8) respectively) using certain discrete time stochastic processes and their approximation via techniques reminiscent of numerical stochastic integration (Kuosmanen et al., 2022).~~

On the practical side, the existence of noise-induced selection implies that simulation studies working with evolutionary individual-based or agent-based models should be careful about whether interaction effects are incorporated into birth rates or death rates since this seemingly arbitrary choice can have unintended consequences due to noise-induced selection, thus potentially biasing results (McLeod and Day, 2019; Kuosmanen et al., 2022). My results also indicate that measuring the growth rate of populations is not, in general, sufficient for accurate prediction/inference of future trajectories of the relative abundance of a species (or phenotype, allele, etc.) from empirical data even in completely controlled environments. The growth rate $w_i = b_i^{(ind)} - d_i^{(ind)}$ of a species i only specifies the difference between its per-capita birth and death rates. In contrast, the complete stochastic dynamics also depend on the total turnover $\tau_i = b_i^{(ind)} + d_i^{(ind)}$ (*i.e.* the sum of the per-capita birth and death rates). This discrepancy can lead to systematic deviations from the trajectories predicted using the growth rate alone, especially in smaller populations. **This deviation should be empirically measurable.** Recent improvements in statistical inference methods for birth and death rates suggest that we should also be able to estimate the birth and death rates of real populations (Huynh et al., 2023) and thus the quantitative predictions of our SDEs, providing a direct empirical test of theoretical predictions. However, since the strength of noise-induced selection scales inversely with the total population size, we can expect such deviations to quickly become too small to detect in medium to large populations, though this is purely due to limits on the precision of experimental measurements rather than due to any theoretical limitations.

Being mindful of noise-induced selection is also important for applied fields that regularly deal with manipulating small populations, such as conservation and population management.

For example, when trying to increase the population of a hypothetical desired species in a multi-species community, increasing the birth rate is *not* equivalent to reducing the death rate even though both result in an increase in the Malthusian fitness (growth rate) w_i . Decreasing the death rate leads to a decrease in τ_i , which is favored by noise-induced selection, whereas increasing the birth rate leads to an increase in τ_i , which leads to noise-induced selection acting to reduce the abundance of the focal species from the community (See Appendix D.3 for a simple example that illustrates this). If the total community size is small, increasing the birth rate of a species can thus lead to noise-induced selection completely eliminating the focal species from the community despite the fact that we *increased* the growth rate of this species. Indeed, numerical investigations of ‘burst-death’ dynamics in fluctuating viral populations show that increasing Malthusian fitness by boosting the survival rate (*i.e.* reducing the death rate) leads to a greater increase in fixation probability than if an exactly equivalent increase in Malthusian fitness is achieved via increasing the burst rate (*i.e.* increasing the birth rate) of viral particles (Alexander and Wahl, 2008). A recent finite population birth-death model for cancer treatment provides another concrete example of the consequences of the asymmetry between changing birth rates and death rates: [This study shows that in their model](#), due to the presence of noise-induced selection, the potential of a tumorous growth to adapt to treatments and experience evolutionary rescue depends inversely on the per-capita turnover τ_i of the constituent cancer cells, with obvious implications for optimal treatment strategies (Raatz and Traulsen, 2023).

Lastly, noise-induced selection is particular to fluctuating populations and does not occur in models with fixed population sizes² such as Wright-Fisher or Moran models, suggesting that working with such constant population frameworks is not sufficient to accurately capture the dynamics of real populations. Several theoretical studies have pointed out that the empirically common practice of approximating the size of fluctuating populations through a constant ‘effective population size’ (often obtained as the harmonic mean of population size over time) is of limited applicability (Gillespie, 1974; Sjödin et al., 2005; Parsons et al., 2010; Iizuka, 2010; Abu Awad and Coron, 2018; Kuosmanen et al., 2022). Recent experimental evolution studies have also directly shown that the harmonic mean of population size need not be a good proxy for capturing and/or predicting evolutionary dynamics in fluctuating

²If $\sum_j x_j$ is a constant, the map $x_i \rightarrow x_i / \sum_j x_j$ becomes a linear map and we no longer need Itô’s formula to move from densities to frequencies in the derivation I conduct in Appendix B; Thus, simply dividing equation (3.9) by the (now constant) total population size provides the complete dynamics of the system in frequency space: Note that the directional terms in equation (3.9) depend only on \mathbf{A}^- , which in turn depends only on the fitness w_i and the mutation terms, and this system thus has no noise-induced selection.

populations (Chavhan et al., 2019). In our case, noise-induced selection cannot be captured by any constant effective population size, harmonic mean or otherwise, since noise-induced selection effects also depend on the turnover τ_i of each type and are thus a property of each *type* in a fluctuating population (and not merely a function of the population size as a whole), meaning that each type would demand its own effective population size for the entire population. My results thus lend further support to one of the general messages of these previous studies - approximating fluctuating populations via a simpler constant (effective) population size requires some careful justification, since it may inadvertently remove important evolutionary properties of the systems under study.

6.3 Connections with other theoretical frameworks

Lion, 2018 has recently proposed a reformulation of the relatively well-known unification of eco-evolutionary dynamics via the Price equation (Frank, 2012; Queller, 2017; Luque and Baravalle, 2021) in a dynamically sufficient, continuous time framework using ‘feedback’ as a ‘unifying perspective’. Our equations generalize the unifying framework described in Lion, 2018 to finite, fluctuating populations — taking $K \rightarrow \infty$ in equations (5.1), (5.2), and (5.9) recover equations (6), (11), and (14) in Lion, 2018 respectively. Lion, 2018 has pointed out that in the dynamic setting (for infinite populations), the replicator-mutator equation (3.17) is in some sense the ‘most fundamental’ of the lot, and equations like the Price equation are best viewed as a hierarchy of moment equations for the population mean, population variance, etc. of a type-level quantity. This is also true in our framework - equation (5.1) is the fundamental equation for population dynamics, and equations like (5.2) and (5.9) can then be derived from (5.1) through repeated application of Ito’s formula (in principle for any moment of the distribution of f in the population, though this quickly becomes too tedious to actually carry out in practice). If we assume that the quantity f follows a Gaussian distribution, then the mean and variance completely characterize the distribution, and equations (5.1), (5.2), and (5.9) together specify the complete stochastic dynamics of the system.

The equivalent of our stochastic equations has recently been derived for quantitative traits from a very different starting point using the theory of measure-valued branching processes (Week et al., 2021) — Equations (21b) and (21c) in Week et al., 2021 are exactly the $m \rightarrow \infty$ version of our equations for changes in the mean value of a type-level quantity and changes in the variance of a type-level quantity respectively for the special case in which

the type-level quantity is the value of the quantitative trait being studied. The methods used by these authors is exactly complementary to the field equations we formulate in Chapter 4 (See section 5.2 for a more detailed technical discussion). A recent preprint has also independently arrived at our equations for type frequencies (equation (5.1)) and the change of mean fitness and turnover in the population (equations (5.7) and (5.8)) in the context of life-history evolution using certain discrete time stochastic processes and their approximation via techniques reminiscent of numerical stochastic integration (Kuosmanen et al., 2022).

A recent study showed that in a very broad class of density-dependent competition models, model-specific details can be ignored in favor of a universal ‘coarse-grained’ description that only includes a small number of very generally defined quantities (Mazzolini and Grilli, 2022). Among the generic predictions of this general description was the presence of noise-induced selection in general competition models. Our work generalizes this particular result of Mazzolini and Grilli, 2022 to models with arbitrary interaction types and presents the relevant equations in a formalism that is more in accordance with standard biological models such as the Price equation. Our results also show that equations like the replicator-mutator equation and Price equation are asymptotically ‘universal’, in the sense that an arbitrary density-dependent birth-death process with functional forms given by (3.10) will satisfy the replicator-mutator or Price equation as we move to infinitely large populations ($K \rightarrow \infty$). Thus, these equations provide coarse-grained ‘fundamental theorems’ that are always satisfied for biological populations that are sufficiently large.

Rice has proposed a stochastic version of the Price equation in a series of papers (Rice, 2008; Rice and Papadopoulos, 2009; Rice, 2020) that follow the classic derivation by Price himself (Rice, 2004). Like the original Price equation, these equations are formulated in a very general manner that relates the phenotypic change between two given populations. As such, they are more general than my equations, but, just like the original Price equation, are dynamically insufficient and cannot be phrased in the language of dynamical systems (to the best of my understanding). They are thus the true stochastic analog of the original Price equation, whereas my version (equation (5.2)) is the analog of Lion’s (2018) version of the Price equation in a continuous time, dynamically sufficient setting. The two formulations are complementary — Rice’s equations (and the original Price equation) provide a very general description for partitioning phenotypic change between two time points, whereas my approach (and Lion’s (2018) version of the Price equation) provides equations that are slightly less

general, but are formulated in a continuous time, dynamically sufficient, predictive manner. Rice's derivations also treat fitness as fundamental, whereas I derive suitable notions of fitness and turnover from demographic first principles. As a consequence, the 'extra' stochastic term corresponding to noise-induced selection that appears in my equations fundamentally emerges from the stochasticity of the underlying births and deaths of organisms and is thus of ecological/demographic origin, whereas the 'extra' stochastic term in Rice's equations emerges from the stochasticity of fitness³alone and, to the best of my knowledge, does *not* correspond to the same effect I identify in this thesis. Lastly, whereas Rice's equations are exact, my equations are derived using a system-size expansion (see part II) and thus are, strictly, speaking only an approximate statistical description of the dynamics (though these approximations are usually very good for even moderate values of K (Black and McKane, 2012; Cianci et al., 2015))

At first thought, the idea of an evolutionary force that selects individuals with lower birth and death rates over individuals with higher birth and death rates may be reminiscent of notions in life-history evolution like r vs K selection or selection on the pace of life (Stearns, 1977). However, it is not possible to conclude whether this similarity reflects some deep principle or whether it is just superficial based solely on the work conducted in this thesis. Models in life-history theory are often primarily concerned with spatiotemporally fluctuating external environments, and thus the stochasticity at the core of life-history models is extrinsic to the population. Ecological frameworks such as modern coexistence theory, which also deal with questions about similar population dynamics and would benefit from a first principles stochastic birth-death formulation, also typically work with fluctuating external environments (Chesson, 1982; Chesson, 1994; Barabás et al., 2018; E. Johnson and Hastings, 2022). I have entirely neglected such extrinsic factors in my formalism. In principle, it is possible to make the birth and death rates (3.10) in my framework also depend on a temporally varying external environment $\mathbf{E}(t)$ (whose variation may possibly depend on the population $\mathbf{n}(t)$). Incorporating such a term would ensure that the 'ecological feedback' terms in equations (5.2) and (5.9) are non-zero, but may also lead to much more complex dynamics. If the variation of the environment $\mathbf{E}(t)$ has some associated stochasticity, the complete dynamics of the system would be the result of interactions between two qualitatively different forms of noise, *extrinsic* noise from the environment, and *intrinsic* noise from the finiteness of the population. Both theoretical (Gokhale and Hauert, 2016; Wienand et al.,

³defined here as the number of offspring produced — no deaths involved (Rice, 2008; Rice and Papadopoulos, 2009; Rice, 2020)

2017) and empirical (Chavhan et al., 2020; Chavhan et al., 2021) work suggests that such dynamics can be quite complex and intricate, and may consequently be difficult to handle analytically in the sort of generality I have employed throughout this thesis. Thus, while integrating the birth-death framework I outline here with ecological ideas such as the pace-of-life syndrome (Mathot and Frankenhuis, 2018; J. Wright et al., 2019) or modern coexistence theory (Barabás et al., 2018; E. Johnson and Hastings, 2022) is biologically appealing, it is likely far from trivial and may present a promising avenue for future work.

6.4 Outlook

Actually solving the equations I formulate analytically for equilibrium/stationary state distributions of \mathbf{p} , \bar{f} , and σ_f^2 will likely quickly become impossibly difficult if the birth and death rate functions are complicated. Indeed, previous studies indicate that in high dimensions, evolutionary birth-death models can exhibit a dizzying array of complicated behavior, including limit cycles and chaotic trajectories (Doebeli and Ispolatov, 2017), and innovative future work is needed to develop tractable approximation schemes to handle these complications. However, while such complicated behavior may present impediments to formulating exact solutions to the equations I formulate, the point of these equations is not necessarily to solve them to begin with. Indeed, the most important takeaway from this thesis is not in the solutions to the equations of section 5.1 but in the formulation and the equations themselves. The equations I derive are very general, since part II makes essentially no assumptions other than density dependence, the impossibility of infinite growth starting from finite population size, and the ability to define per-capita birth and death rates, and thus, these birth and death rate function(al)s can in general be almost arbitrarily complicated. They thus apply to a very wide array of biological populations. Further, as we saw in section 5.1, the terms of these equations lend themselves to simple biological interpretation and make some general *qualitative* predictions about how evolution should operate in finite, fluctuating populations. Namely, my equations show that evolution should operate similarly to how it operates in infinite populations, with the addition of some non-directional fluctuations (this is just drift) as well as an extra *directional* force (noise-induced selection). Like the classical Price equation, the utility of the equations of section 5.1 thus lies not (necessarily) in their solutions for specific models, but instead in their generality and the fact that their terms help us clearly think about the various evolutionary phenomena at play in biological populations (Frank, 2012; Luque, 2017; Luque and Baravalle, 2021). The general spirit of this thesis is thus in line with the general idea of trying to formulate ‘model-independent’ eco-evolutionary theory that

has recently been (rapidly) gaining popularity in the literature (Grafen, 2014; Queller, 2017; Lion, 2018; Allen and McAvoy, 2019; Rice, 2020; Week et al., 2021; Wickman et al., 2022; Kuosmanen et al., 2022; Mazzolini and Grilli, 2022; Lion et al., 2023; Allen et al., 2023).

My formalism here is only a first step, and there are many biologically important factors that I have neglected in this thesis. My framework works with unstructured populations, neglecting any potential effects of groups, age, class, sex, developmental stage, or space⁴, all of which can lead to very complex and often surprising dynamics. Explicitly incorporating features such as population structure, sex, and space from first principles in an analytically tractable framework is a formidable task that may need innovative new mathematical and biological arguments, and presents a fantastic opportunity for future studies. I have also neglected any potential complications introduced by genotype-phenotype maps and genetic processes such as dominance, epistasis, and pleiotropy since my assumptions on the functional forms of the birth and death rates ((3.10) or (4.9)) should be general enough to incorporate these effects (in principle) while analyzing specific models.

Descriptions such as the classic Price equation and the equations I present in this thesis ‘abstract away’ system-specific details and almost inevitably come at the cost of precision (Levins, 1966; Potochnik, 2018). These approaches are thus intended to complement empirical studies and modelling approaches that carefully study specific systems and generate vital knowledge about how these systems behave. To quote Robert Millikan (1924), “Science walks forward on two feet, namely theory and experiment [...] Sometimes it is one foot which is put forward first, sometimes the other, but continuous progress is only made by the use of both - by theorizing and then testing, or by finding new relations in the process of experimenting and then bringing the theoretical foot up and pushing it on beyond, and so on in unending alterations.”

⁴Of course, since position in continuous space is just a special case of a quantitative trait, we have technically incorporated space in a very limited sense: The formalism of chapter 4 can equally well describe a finite population of exactly identical individuals moving through one-dimensional space, and so can technically describe phenomena like range expansion of clonal populations in one dimension. However, I am only noting this as a technicality - such populations are likely somewhat ‘evolutionarily boring’, since all individuals in the model must always be exactly identical in all aspects other than spatial location.

Appendix A

From Itô to Fokker-Planck

Here, I present a simple (informal) derivation of the Fokker-Planck equation for a one-dimensional Itô process. The result for the multi-dimensional case follows from the same logic but is more notationally cumbersome.

Consider a one-dimensional real Itô process given by $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t$ that takes values in an open set $\Omega \subseteq \mathbb{R}$ and admits a probability density function $P(x, t)$ with compact support¹ in Ω at all times $t \in [0, \infty)$. Let $f : \mathbb{R} \rightarrow \mathbb{R}$ be an arbitrary $C^2(\mathbb{R})$ function. By Itô's formula, we have:

$$df(X_t) = \left(\mu f' + \frac{\sigma^2}{2} f'' \right) dt + \sigma f' dW_t$$

Writing this in integral form and taking expectations on both sides yields:

$$\mathbb{E}[f(X_t)] = \mathbb{E} \left[\int_0^t \left(\mu f' + \frac{\sigma^2}{2} f'' \right) ds \right] + \mathbb{E} \left[\int_0^t \sigma f' dW_s \right] \quad (\text{A.1})$$

As long as X_t and $\sigma(X_t, t)$ are reasonably ‘nice’, the stochastic integral in the second term of the RHS of (A.1) will vanish upon taking an expectation (see 2.1). Using the definition of

¹For our biological context you can think of this as just meaning that very extreme values in Ω are so unlikely that we disallow them entirely

the expectation value, we are thus left with:

$$\int_{\Omega} f(X_t) P(x, t) dx = \int_{\Omega} \left(\int_0^t \mu f' + \frac{\sigma^2}{2} f'' ds \right) P(x, t) dx$$

Assuming derivatives and expectations commute², we can now differentiate with respect to time on both sides and use the fundamental theorem of calculus to write

$$\int_{\Omega} f(X_t) \frac{\partial P}{\partial t}(x, t) dx = \underbrace{\int_{\Omega} \mu f' P(x, t) dx}_{M(x, t)} + \underbrace{\int_{\Omega} \frac{\sigma^2}{2} f'' P(x, t) dx}_{N(x, t)} \quad (\text{A.2})$$

We will now use integration by parts to further evaluate the two parts $M(x, t)$ and $N(x, t)$. Since $P(\cdot, t)$ has compact support within the open set $\Omega \subset \mathbb{R}$ for any given $t > 0$, the function $P(x, t) \equiv 0$ on $\partial\Omega$ and the boundary term of the formula vanishes. Thus, we can use integration by parts once on $M(x, t)$ to obtain

$$M(x, t) = - \int_{\Omega} f(X_t) \left(\frac{\partial}{\partial x} \mu P(x, t) \right) dx \quad (\text{A.3})$$

and twice on $N(x, t)$ to obtain

$$\begin{aligned} N(x, t) &= -\frac{1}{2} \int_{\Omega} f'(X_t) \left(\frac{\partial}{\partial x} \sigma^2 P(x, t) \right) dx \\ &= \frac{1}{2} \int_{\Omega} f(X_t) \left(\frac{\partial^2}{\partial x^2} \sigma^2 P(x, t) \right) dx \end{aligned} \quad (\text{A.4})$$

Substituting (A.3) and (A.4) into (A.2) and collecting terms yields

$$\int_{\Omega} f(X_t) \frac{\partial P}{\partial t}(x, t) dx = \int_{\Omega} f(X_t) \left[-\frac{\partial}{\partial x} (\mu P(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} (\sigma^2 P(x, t)) \right] dx$$

Since this is true for an arbitrary choice of $f(x)$, we are thus led to conclude that the density

²By the Leibniz integral rule, this only requires the functions of (x, t) inside the integrals to be bounded and C^1 in an open subset of $\Omega \times [0, \infty)$

function $P(x, t)$ must satisfy:

$$\frac{\partial P}{\partial t}(x, t) = -\frac{\partial}{\partial x}(\mu(x, t)P(x, t)) + \frac{1}{2}\frac{\partial^2}{\partial x^2}((\sigma(x, t))^2 P(x, t)) \quad (\text{A.5})$$

Equation (A.5) is the Fokker-Planck equation in one dimension. Using the exact same strategy, the multidimensional Fokker-Planck equation for the m -dimensional Itô Process $d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, t)dt + \boldsymbol{\sigma}(\mathbf{X}_t, t)d\mathbf{W}_t$ can be easily found to be:

$$\frac{\partial P}{\partial t}(\mathbf{x}, t) = -\sum_{i=1}^m \frac{\partial}{\partial x_i}(\mu_i(\mathbf{x}, t)P(\mathbf{x}, t)) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2}{\partial x_i \partial x_j} (D_{ij}P(\mathbf{x}, t)) \quad (\text{A.6})$$

where $\mathbf{D} = \sigma\sigma^T$. Some authors like to define the ‘probability current’ $\mathbf{J}(\mathbf{x}, t)$, an m -dimensional function with i^{th} element

$$J_i(\mathbf{x}, t) := \mu_i(\mathbf{x}, t)P(\mathbf{x}, t) - \frac{1}{2} \sum_{j=1}^m \frac{\partial}{\partial x_j} (D_{ij}P(\mathbf{x}, t))$$

In this notation, equation (A.6) can be written in the more compact form:

$$\frac{\partial P}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (\text{A.7})$$

where $\nabla \cdot$ is the divergence operator (also denoted div in some texts). Those familiar with physics should immediately recognize that equation (A.7) is in the form of a so-called ‘continuity equation’ for a conserved quantity. Continuity equations turn up in various areas of applied mathematics, most famously in electromagnetism (conservation of charge from Maxwell’s equations), fluid dynamics (continuity equations for mass of a flowing fluid from the Euler equations), and molecular diffusion (Fick’s law). This explains the name ‘probability current’ as an analogy to currents in physics such as electrical current or fluid current. The continuity equation representation also makes it clear that the Fokker-Planck equation describes the ‘flow of probability’ in the system. In particular, equation (A.7) says that the total probability in the system is ‘conserved’, and is thus simply a mathematical formalization of the common-sense idea that whenever the probability of the system of being in a given state decreases, the probability of it being in some other state must increase (and vice versa).

Appendix B

Stochastic trait frequency dynamics using Itô's formula

We first recall the version of the multi-dimensional Itô's formula that will be relevant to us. Consider an m -dimensional real Itô process \mathbf{X}_t given by the solution to

$$d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t)dt + \boldsymbol{\sigma}(\mathbf{X}_t)d\mathbf{W}_t$$

where $\boldsymbol{\mu} : \mathbb{R}^m \rightarrow \mathbb{R}^m$ is the ‘drift vector’ and $\boldsymbol{\sigma} : \mathbb{R}^m \rightarrow \mathbb{R}^{m \times m}$ is the ‘diffusion matrix’. Let $f : \mathbb{R}^m \rightarrow \mathbb{R}$ be an arbitrary $C^2(\mathbb{R}^m)$ function. Then, Itô’s formula (Øksendal, 1998, Section 4.2) states that the stochastic process $f(\mathbf{X}_t)$ must satisfy:

$$df(\mathbf{X}_t) = \left[(\nabla_{\mathbf{X}}f)^T \boldsymbol{\mu} + \frac{1}{2} \text{Tr}[\boldsymbol{\sigma}^T (H_{\mathbf{X}}f) \boldsymbol{\sigma}] \right] dt + (\nabla_{\mathbf{X}}f)^T \boldsymbol{\sigma} d\mathbf{W}_t \quad (\text{B.1})$$

where $\text{Tr}[\cdot]$ denotes the trace of a matrix, $(\cdot)^T$ denotes the transpose, and we have suppressed the \mathbf{X}_t dependence of $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$ to reduce clutter. Here, $\nabla_{\mathbf{X}}f$ is the m -dimensional *gradient vector* of f and $H_{\mathbf{X}}f$ is the $m \times m$ *Hessian matrix* of f , respectively defined for $f([x_1, \dots, x_m]^T)$ as:

$$\begin{aligned} (\nabla_{\mathbf{x}}f)_j &= \frac{\partial f}{\partial x_j} \\ (H_{\mathbf{x}}f)_{jk} &= \frac{\partial^2 f}{\partial x_j \partial x_k} \end{aligned}$$

In our case, we have the Itô process given by (3.9), which defines how the density of each type of individual changes over time. We thus have $\boldsymbol{\mu}(\mathbf{X}_t) = \mathbf{A}^-(\mathbf{X}_t)$ and $\boldsymbol{\sigma}(\mathbf{X}_t) = \mathbf{D}(\mathbf{X}_t)/\sqrt{K}$. For each fixed $i \in \{1, 2, \dots, m\}$, let us define a scalar function $f_i : \mathbb{R}^m \rightarrow \mathbb{R}$ as

$$f_i(\mathbf{x}) = \frac{x_i}{\sum_{j=1}^m x_j}$$

Thus, $f_i(\mathbf{X}_t)$ gives us the frequency of type i individuals when the population is described by the vector \mathbf{X}_t . This function is obviously $C^2(\mathbb{R}^m)$, and we can thus use Itô's formula (B.1) to describe how it changes over time. The j^{th} element of the gradient of f_i is given by:

$$\begin{aligned} (\nabla_{\mathbf{x}} f_i)_j &= \frac{\partial}{\partial x_j} \left(\frac{x_i}{\sum_{k=1}^m x_k} \right) \\ &= \left(\left(\frac{1}{\sum_{r=1}^m x_r} \right) \frac{\partial x_i}{\partial x_j} - \left(\frac{x_i}{\left(\sum_{r=1}^m x_r \right)^2} \right) \sum_{k=1}^m \frac{\partial x_k}{\partial x_j} \right) \\ &= \frac{1}{\sum_{r=1}^m x_r} (\delta_{ij} - p_i) \end{aligned} \tag{B.2}$$

where we have defined the frequency of the i^{th} type $p_i = f_i(\mathbf{x})$ and used the fact that $\frac{\partial x_j}{\partial x_k} = \delta_{jk}$. The jk^{th} element of the Hessian is given by:

$$\begin{aligned} (H_{\mathbf{x}} f_i)_{jk} &= \frac{\partial^2}{\partial x_j \partial x_k} \left(\frac{x_i}{\sum_{l=1}^m x_l} \right) \\ &= \frac{\partial}{\partial x_j} \left(\frac{\delta_{ik}}{\sum_{r=1}^m x_r} - \frac{x_i}{\left(\sum_{r=1}^m x_r \right)^2} \right) \\ &= \frac{1}{\left(\sum_{r=1}^m x_r \right)^2} (2p_i - \delta_{ij} - \delta_{ik}) \end{aligned} \tag{B.3}$$

Thus, for the first term of (B.1), we have:

$$\begin{aligned}
(\nabla_{\mathbf{x}} f_i)^T \mathbf{A}^- &= \sum_{j=1}^m \left((\nabla_{\mathbf{x}} f_i)_j \right) A_j^- \\
&= \frac{1}{\sum_{r=1}^m x_r} \sum_{j=1}^m (\delta_{ij} - p_i) A_j^- \\
&= \frac{1}{\sum_{r=1}^m x_r} \left(A_i^- - p_i \sum_{j=1}^m A_j^- \right)
\end{aligned} \tag{B.4}$$

This term describes the effects of selection and mutation at the infinite population limit. However, the finiteness of the population adds a second directional term to these dynamics, described by the second term that multiplies dt in (B.1). To calculate it, we first calculate:

$$\begin{aligned}
\frac{1}{\sqrt{K}} (H_{\mathbf{x}} f_i \mathbf{D})_{jk} &= \frac{1}{\sqrt{K}} \sum_{l=1}^m (H_{\mathbf{x}} f_i)_{jl} (\mathbf{D})_{lk} \\
&= \frac{1}{\sqrt{K} \left(\sum_{r=1}^m x_r \right)^2} \sum_{l=1}^m (2p_i - \delta_{ij} - \delta_{il}) \delta_{lk} (A_l^+ A_k^+)^{\frac{1}{4}}
\end{aligned} \tag{B.5}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{K} \left(\sum_{r=1}^m x_r \right)^2} \left((2p_i - \delta_{ij}) (A_k^+)^{\frac{1}{2}} - \delta_{ik} (A_i^+ A_k^+)^{\frac{1}{4}} \right)
\end{aligned} \tag{B.6}$$

$$\begin{aligned}
&= \frac{1}{\sqrt{K} \left(\sum_{r=1}^m x_r \right)^2} (2p_i - \delta_{ij} - \delta_{ik}) (A_k^+)^{\frac{1}{2}}
\end{aligned} \tag{B.7}$$

and thus:

$$\begin{aligned}
\frac{1}{K} (\mathbf{D}^T H_{\mathbf{x}} f_i \mathbf{D})_{lk} &= \frac{1}{K} \sum_{j=1}^m (\mathbf{D}^T)_{lj} (H_{\mathbf{x}} f_i \mathbf{D})_{jk} \\
&= \frac{1}{K \left(\sum_{r=1}^m x_r \right)^2} \sum_{j=1}^m \delta_{lj} (A_l^+ A_j^+)^{\frac{1}{4}} (A_k^+)^{\frac{1}{2}} (2p_i - \delta_{ij} - \delta_{ik})
\end{aligned} \tag{B.8}$$

$$\begin{aligned}
&= \frac{1}{K \left(\sum_{r=1}^m x_r \right)^2} (A_k^+)^{\frac{1}{2}} \left(2p_i (A_l^+)^{\frac{1}{2}} - (A_i^+)^{\frac{1}{2}} \delta_{il} - (A_l^+)^{\frac{1}{2}} \delta_{ik} \right)
\end{aligned} \tag{B.9}$$

Using this, we see that the trace of this matrix is given by:

$$\begin{aligned} \frac{1}{K} \text{Tr}[\mathbf{D}^T H_{\mathbf{x}} f_i \mathbf{D}] &= \frac{1}{K} \sum_{k=1}^m (\mathbf{D}^T H_{\mathbf{x}} f_i \mathbf{D})_{kk} \\ &= \frac{1}{K \left(\sum_{r=1}^m x_r \right)^2} \sum_{k=1}^m \left(2p_i (A_k^+ A_k^+)^{\frac{1}{2}} - (A_i^+ A_k^+)^{\frac{1}{2}} \delta_{ik} - (A_k^+ A_k^+)^{\frac{1}{2}} \delta_{ik} \right) \end{aligned} \quad (\text{B.10})$$

$$= \frac{1}{K \left(\sum_{r=1}^m x_r \right)^2} \left(2p_i \left(\sum_{k=1}^m A_k^+ \right) - 2A_i^+ \right) \quad (\text{B.11})$$

and thus, the second term multiplying dt in (B.1) is given by:

$$\frac{1}{2K} \text{Tr}[\mathbf{D}^T H_{\mathbf{x}} f_i \mathbf{D}] = \frac{-1}{K \left(\sum_{r=1}^m x_r \right)^2} \left(A_i^+ - p_i \left(\sum_{k=1}^m A_k^+ \right) \right) \quad (\text{B.12})$$

Finally, denoting $d\mathbf{W}_t = [dW_t^{(1)}, dW_t^{(2)}, \dots, dW_t^{(m)}]^T$ where each $dW_t^{(j)}$ is an independent one dimensional Wiener process, we have:

$$\begin{aligned} (\mathbf{D} d\mathbf{W}_t)_j &= \sum_{k=1}^m \mathbf{D}_{jk} dW_t^{(k)} \\ &= \sum_{k=1}^m \delta_{jk} (A_j^+ A_k^+)^{\frac{1}{4}} dW_t^{(k)} \end{aligned} \quad (\text{B.13})$$

$$= (A_j^+)^{1/2} dW_t^{(j)} \quad (\text{B.14})$$

Thus, using (B.2), we see that the last term on the RHS of (B.1) is given by:

$$\begin{aligned} \frac{1}{\sqrt{K}} (\nabla_{\mathbf{x}} f)^T \mathbf{D} d\mathbf{W}_t &= \frac{1}{\sqrt{K}} \sum_{j=1}^m (\nabla_{\mathbf{x}} f_i)_j (\mathbf{D} d\mathbf{W}_t)_j \\ &= \frac{1}{\left(\sum_{r=1}^m x_r \right) \sqrt{K}} \sum_{j=1}^m (\delta_{ij} - p_i) (A_j^+)^{1/2} dW_t^{(j)} \end{aligned} \quad (\text{B.15})$$

$$= \frac{1}{\left(\sum_{r=1}^m x_r \right) \sqrt{K}} (A_i^+)^{1/2} dW_t^{(i)} - p_i \sum_{j=1}^m (A_j^+)^{1/2} dW_t^{(j)} \quad (\text{B.16})$$

Putting equations (B.4), (B.12) and (B.16) into (B.1) and letting $N_K(t) = \sum_{r=1}^m x_r$ we see that $p_i = f_i(\mathbf{X})_t$, the frequency of the i^{th} type in the population \mathbf{X}_t , changes according to the equation:

$$\begin{aligned} dp_i &= \underbrace{\frac{1}{N_K(t)} \left(A_i^- - p_i \sum_{j=1}^m A_j^- \right) dt}_{K \rightarrow \infty \text{ prediction}} - \underbrace{\frac{1}{K} \frac{1}{N_K^2(t)} \left(A_i^+ - p_i \left(\sum_{k=1}^m A_k^+ \right) \right) dt}_{\substack{\text{Directional finite size effects} \\ \text{due to differential turnover rates}}} \\ &\quad + \underbrace{\frac{1}{\sqrt{K} N_K(t)} \left[(A_i^+)^{1/2} dW_t^{(i)} - p_i \sum_{j=1}^m (A_j^+)^{1/2} dW_t^{(j)} \right]}_{\substack{\text{Non-directional finite size effects} \\ \text{due to stochastic fluctuations}}} \end{aligned} \quad (\text{B.17})$$

Plugging the functional forms of (3.10) and the definitions of w_i and τ_i into the definitions of A_i^- and A_i^+ , we obtain the relations

$$\begin{aligned} A_i^- &= x_i w_i(\mathbf{x}) + \mu Q_i(\mathbf{x}) \\ A_i^+ &= x_i \tau_i(\mathbf{x}) + \mu Q_i(\mathbf{x}) \end{aligned} \quad (\text{B.18})$$

Thus, for the first term of (B.17), we have

$$\begin{aligned} \frac{1}{N_K(t)} \left(A_i^- - p_i \sum_{j=1}^m A_j^- \right) &= \frac{1}{N_K(t)} [w_i(\mathbf{x})x_i + \mu Q_i(\mathbf{x})] - \frac{p_i}{N_K(t)} \sum_{j=1}^m [w_j(\mathbf{x})x_j + \mu Q_j(\mathbf{x})] \\ &= w_i(\mathbf{x})p_i + \frac{\mu}{N_K(t)} Q_i(\mathbf{x}) - p_i \sum_{j=1}^m \left[w_j(\mathbf{x})p_j + \frac{\mu}{N_K(t)} Q_j(\mathbf{x}) \right] \end{aligned}$$

Where we have used the definition of p_i from (3.11). Now using the definition of mean fitness from (3.12) and rearranging terms gives us

$$\frac{1}{N_K(t)} \left(A_i^- - p_i \sum_{j=1}^m A_j^- \right) = (w_i(\mathbf{x}) - \bar{w})p_i + \mu \left[Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right] \quad (\text{B.19})$$

where we have defined $Q_j(\mathbf{p}) = Q_j(\mathbf{x})/N_K(t)$. Repeating the exact same calculations for the A_i^+ terms in the second term of (B.17) now yields equation (3.14) (which is also equation (5.1)) in the main text.

Appendix C

A Price-like equation for the variance of a type-level quantity

Let σ_f^2 denote the statistical variance of a type-level quantity, defined as:

$$\sigma_f^2 := \overline{(f^2)} - (\overline{f})^2 \quad (\text{C.1})$$

where \overline{X} is the statistical mean value defined by (3.12). By the product rule, we have

$$\frac{d\sigma_f^2}{dt} = 2\overline{f}\frac{\partial f}{\partial t} + \sum_{i=1}^m f_i^2 \frac{dp_i}{dt} - \frac{d}{dt}(\overline{f}^2) \quad (\text{C.2})$$

We will evaluate the RHS term by term. The first term is as simplified as can be without more information about f . For the second term, we can substitute dp_i from (5.1) and use the same steps used in going from (3.17) to (3.21) to write

$$\begin{aligned} \sum_{i=1}^m f_i^2 dp_i &= \text{Cov}(w, f^2)dt - \frac{1}{KN_K} \text{Cov}(\tau, f^2)dt \\ &\quad + \mu \left(1 - \frac{1}{KN_K(t)} \right) \left(\sum_{i=1}^m f_i^2 Q_i(\mathbf{p}) - \overline{f^2} \sum_{i=1}^m Q_i(\mathbf{p}) \right) dt \\ &\quad + \frac{1}{\sqrt{KN_K(t)}} \left(\sum_{i=1}^m (f_i^2 - \overline{f^2}) \sqrt{A_i^+} dW_t^{(i)} \right) \end{aligned} \quad (\text{C.3})$$

For the third term, we need to use Itô's formula. Here, the relevant version of Itô's formula is the one-dimensional version of (B.1). Given a one-dimensional process $dX_t = S(X_t)dt + \sum_j D_j(X_t)dW_t^{(j)}$ with S, D_j being suitable real functions and each $W_t^{(j)}$ being an independent Wiener process, Itô's formula says that given any $C^2(\mathbb{R})$ function $g(x)$, we have the relation:

$$dg(X_t) = \left(S(X_t)g'(X_t) + \frac{g''(X_t)}{2} \sum_j D_j^2(X_t) \right) dt + \sum_j D_j(X_t)g'(X_t)dW_t^{(j)} \quad (\text{C.4})$$

In our case, we have a one-dimensional process for the mean value $d\bar{f}$ of the type level quantity, and the $C^2(\mathbb{R})$ function $g(x) = x^2$. Itô's formula thus says that the third term of (C.2) is given by:

$$d(\bar{f}^2) = \left(2\bar{f}S(X_t) + \sum_j D_j^2(X_t) \right) dt + \sum_j 2\bar{f}D_j(X_t)dW_t^{(j)} \quad (\text{C.5})$$

where the relevant functions S and D_j can be read off from (5.2). Since the dW terms are unwieldy, we will denote the contribution of all the dW_t terms collectively by $dW_{\sigma_f^2}$ to reduce notational clutter and only explicitly calculate these terms at the end. We also note that the covariance operator is a bilinear form, *i.e.* given any three quantities X, Y and Z and any constant $a \neq 0$, we have the relations:

$$\begin{aligned} \text{Cov}(aX, Y) &= a\text{Cov}(X, Y) = \text{Cov}(X, aY) \\ \text{Cov}(X, Y + Z) &= \text{Cov}(X, Y) + \text{Cov}(X, Z) \end{aligned}$$

Substituting equations (C.3) and (C.5) into equation (C.2) and using this property of covariances, we obtain:

$$\begin{aligned} d\sigma_f^2 &= \text{Cov}(w, f^2 - 2\bar{f}f)dt - \frac{1}{KN_K} (\text{Cov}(\tau, f^2 - 2\bar{f}f)) dt + 2 \left(\overline{f \frac{\partial f}{\partial t}} - \bar{f} \overline{\left(\frac{\partial f}{\partial t} \right)} \right) dt \\ &\quad + \mu \left(1 - \frac{1}{KN_K(t)} \right) \left(\sum_{i=1}^m (f_i^2 - 2\bar{f}f_i)Q_i(\mathbf{p}) - (\bar{f}^2 - 2\bar{f}^2) \sum_{i=1}^m Q_i(\mathbf{p}) \right) dt \\ &\quad - \frac{1}{KN_K^2(t)} \left(\sum_{i=1}^m (f_i - \bar{f})^2 A_i^+ \right) dt \\ &\quad + dW_{\sigma_f^2} \end{aligned} \quad (\text{C.6})$$

Now, we note that

$$\frac{1}{N_K} A_i^+ = \frac{1}{N_K} (\tau_i x_i + \mu Q_i(\mathbf{x})) \quad (\text{C.7})$$

$$= \tau_i p_i + \mu Q_i(\mathbf{p}) \quad (\text{C.8})$$

and thus the third line of (C.6) is

$$\frac{1}{KN_K^2(t)} \left(\sum_{i=1}^m (f_i - \bar{f})^2 A_i^+ \right) dt = \frac{1}{KN_K} \sum_{i=1}^m (f_i - \bar{f})^2 (\tau_i p_i + \mu Q_i(\mathbf{p})) \quad (\text{C.9})$$

$$= \frac{1}{KN_K} \sum_{i=1}^m (f_i - \bar{f})^2 (\tau_i p_i + \mu Q_i(\mathbf{p})) \quad (\text{C.10})$$

$$= \frac{1}{KN_K} \left(\overline{\tau (f - \bar{f})^2} + \mu \sum_{i=1}^m (f_i - \bar{f})^2 Q_i(\mathbf{p}) \right) \quad (\text{C.11})$$

$$= \frac{1}{KN_K} \left(\text{Cov}(\tau, (f - \bar{f})^2) + \bar{\tau} \overline{(f - \bar{f})^2} + \mu \sum_{i=1}^m (f_i - \bar{f})^2 Q_i(\mathbf{p}) \right) \quad (\text{C.12})$$

$$= \frac{1}{KN_K} \left(\text{Cov}(\tau, (f - \bar{f})^2) + \bar{\tau} \sigma_f^2 + \mu \sum_{i=1}^m (f_i - \bar{f})^2 Q_i(\mathbf{p}) \right) \quad (\text{C.13})$$

where we have used the definition of statistical covariance in the second to last line and used the definition of statistical variance in the last line. Substituting (C.13) into (C.6) and using $M_{\sigma_f^2}(\mathbf{p}, N_K)$ to denote the contributions of all the mutational terms (*i.e.* all terms with a μ factor) for notational brevity, we obtain

$$\begin{aligned} d\sigma_f^2 &= \text{Cov}(w, f^2 - 2\bar{f}f)dt - \frac{1}{KN_K} \left(\text{Cov}(\tau, f^2 - 2\bar{f}f) + \text{Cov}(\tau, (f - \bar{f})^2) + \bar{\tau} \sigma_f^2 \right) dt \\ &\quad + 2\text{Cov} \left(\frac{\partial f}{\partial t}, f \right) dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dW_{\sigma_f^2} \end{aligned} \quad (\text{C.14})$$

We can now complete the square inside the covariance terms of the first line of the RHS by writing $f^2 - 2\bar{f}f = (f - \bar{f})^2 - \bar{f}^2$ to obtain

$$\begin{aligned} d\sigma_f^2 &= \left[\text{Cov}(w, (f - \bar{f})^2) - \text{Cov}\left(w, (\bar{f})^2\right) \right] dt \\ &\quad - \frac{1}{KN_K} \left[\text{Cov}(\tau, (f - \bar{f})^2) - \text{Cov}\left(\tau, (\bar{f})^2\right) + \text{Cov}(\tau, (f - \bar{f})^2) + \bar{\tau}\sigma_f^2 \right] dt \quad (\text{C.15}) \\ &\quad + 2\text{Cov}\left(\frac{\partial f}{\partial t}, f\right) dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dW_{\sigma_f^2} \end{aligned}$$

To simplify the covariance terms of the first line of the RHS, we observe that

$$\begin{aligned} \text{Cov}\left(w, (\bar{f})^2\right) &= \overline{(w(\bar{f})^2)} - \bar{w}\overline{(\bar{f})^2} \\ &= (\bar{f})^2 \sum_{i=1}^m w_i p_i - \bar{w}(\bar{f})^2 \sum_{i=1}^m p_i \\ &= (\bar{f})^2 \bar{w} - \bar{w}(\bar{f})^2 = 0 \end{aligned}$$

and similarly,

$$\text{Cov}\left(\tau, (\bar{f})^2\right) = 0$$

and thus, using this in (C.15), we see that the rate of change of the variance of any type-level quantity f in the population satisfies:

$$\begin{aligned} d\sigma_f^2 &= \text{Cov}(w, (f - \bar{f})^2) dt - \frac{1}{KN_K} [\bar{\tau}\sigma_f^2 + 2\text{Cov}(\tau, (f - \bar{f})^2)] dt \\ &\quad + 2\text{Cov}\left(\frac{\partial f}{\partial t}, f\right) dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dW_{\sigma_f^2} \quad (\text{C.16}) \end{aligned}$$

This is precisely equation (5.9) in the main text. To calculate the mutation term, we substitute (C.13) into (C.6) to find

$$\begin{aligned} M_{\sigma_f^2}(\mathbf{p}, N_K) &= \mu \left(\sum_{i=1}^m \left(f_i^2 - 2\bar{f}f_i - \bar{f}^2 + 2\bar{f}^2 \right) Q_i(\mathbf{p}) \right) \\ &\quad - \frac{\mu}{KN_K} \sum_{i=1}^m \left(f_i^2 - 2\bar{f}f_i - \bar{f}^2 + 2\bar{f}^2 + (f_i - \bar{f})^2 \right) Q_i(\mathbf{p}) \quad (\text{C.17}) \end{aligned}$$

We can further simplify the first term of the RHS as

$$\begin{aligned} f_i^2 - 2\bar{f}f_i - \bar{f}^2 + 2\bar{f}^2 &= (f_i^2 + \bar{f}^2 - 2\bar{f}f_i) - (\bar{f}^2 - \bar{f}^2) \\ &= (f_i - \bar{f})^2 + \sigma_f^2 \end{aligned}$$

and similarly, the second term as

$$f_i^2 - 2\bar{f}f_i - \bar{f}^2 + 2\bar{f}^2 + (f_i - \bar{f})^2 = 2(f_i - \bar{f})^2 + \sigma_f^2$$

thus, the contributions of mutations to the change in the variance of f are given by

$$\begin{aligned} M_{\sigma_f^2}(\mathbf{p}, N_K) &= \mu \left(\sum_{i=1}^m ((f_i - \bar{f})^2 + \sigma_f^2) Q_i(\mathbf{p}) \right) \\ &\quad - \frac{\mu}{KN_K} \sum_{i=1}^m (2(f_i - \bar{f})^2 + \sigma_f^2) Q_i(\mathbf{p}) \end{aligned} \tag{C.18}$$

which after slight rearrangement becomes

$$M_{\sigma_f^2}(\mathbf{p}, N_K) = \mu \left(\sum_{i=1}^m \left[\left(1 - \frac{2}{KN_K} \right) (f_i - \bar{f})^2 Q_i(\mathbf{p}) \right] + \sigma_f^2 \left(1 - \frac{1}{KN_K} \right) \sum_{i=1}^m Q_i(\mathbf{p}) \right) \tag{C.19}$$

which is equation (5.10) in the main text. For the dW terms, we can use equations (C.3) and (C.5) to calculate:

$$dW_{\sigma_f^2} = \frac{1}{\sqrt{KN_K(t)}} \left(\sum_{i=1}^m \left(f_i^2 - \bar{f}^2 - 2\bar{f}(f_i - \bar{f}) \right) \sqrt{A_i^+} dW_t^{(i)} \right) \tag{C.20}$$

$$= \frac{1}{\sqrt{KN_K(t)}} \left(\sum_{i=1}^m \left(f_i^2 - \bar{f}^2 - 2\bar{f}f_i - 2\bar{f}^2 \right) \sqrt{A_i^+} dW_t^{(i)} \right) \tag{C.21}$$

$$= \frac{1}{\sqrt{KN_K(t)}} \left(\sum_{i=1}^m (f_i - \bar{f})^2 \sqrt{A_i^+} dW_t^{(i)} \right) \tag{C.22}$$

which is equation (5.11) in the main text.

Appendix D

Some Examples

This appendix provides several examples of the birth-death processes studied in this thesis. The material covered in this thesis, including the examples in this Appendix, are currently being written up for publication. Consequently, the GitHub repository containing the scripts used to make the plots presented below is currently private. The repository will be made public after publication, and you should be able to access it via [this link](#) (which will not work until I make the repository public) if you are reading this after the material is published; Until then, scripts are available on request, just drop me an email.

D.1 An example in one dimension: The stochastic logistic equation

Here, we analyze example 1. To recap, we had a population of individuals that exhibit a constant per-capita birth rate $\lambda > 0$, and a per-capita death rate that had the linear density-dependence $\mu + (\lambda - \mu) \frac{n}{K}$, where μ and K are positive constants. Thus, we have the birth and death rates

$$\begin{aligned} b(n) &= \lambda n \\ d(n) &= \left(\mu + (\lambda - \mu) \frac{n}{K} \right) n \end{aligned} \tag{D.1}$$

Here, K is the system-size parameter. Introducing the population density $x := n/K$, we obtain

$$\begin{aligned} b_K(x) &= \frac{1}{K}b(n) = \frac{1}{K}\lambda Kx \\ d_K(x) &= \frac{1}{K}d(n) = \frac{1}{K}\left(\mu + (\lambda - \mu)\frac{Kx}{K}\right)Kx \end{aligned}$$

Thus, we have

$$A^\pm(x) = b_K(x) \pm d_K(x) = x(\lambda \pm ((\mu + (\lambda - \mu)x)))$$

Defining $r = \lambda - \mu$ and $v = \lambda + \mu$ and using equation (2.14), we see that the ‘mesoscopic view’ of the system is given by the solution of the SDE

$$dX_t = rX_t(1 - X_t)dt + \sqrt{\frac{X_t(v + rX_t)}{K}}dW_t \quad (\text{D.2})$$

From equation (2.15), we see that the deterministic dynamics are

$$\frac{dx}{dt} = A^-(x) = rx(1 - x) \quad (\text{D.3})$$

showing that in the infinite population limit, we obtain the logistic equation. This derivation also makes it clear that two systems with very different stochastic dynamics can nevertheless converge to the same infinite population limit (D.3), since equation (D.3) only depends on the difference $\lambda - \mu$. To illustrate the effects of this seemingly innocent fact, figure D.1 compares two simulations which have the same value of $\lambda - \mu$ but a ten-fold difference in $\lambda + \mu$. As is clear from the figure, though both populations have the same behavior at the infinite population limit, populations with a higher value of $\lambda + \mu$ exhibit much wilder fluctuations and are therefore more prone to stochastic extinction. This is the root cause of the noise-induced selection that occurs in higher dimensions, discussed in detail in part III.

Letting $\alpha(t)$ be the solution of the logistic equation (D.3), We can Taylor expand $A^\pm(x)$ for the weak noise approximation, and we find:

$$\begin{aligned} A_1^-(x) &= \frac{d}{dx}(rx(1 - x)) \Big|_{x=\alpha} = r(1 - 2\alpha(t)) \\ A_0^+(x) &= \alpha(t)(v + r\alpha(t)) \end{aligned}$$

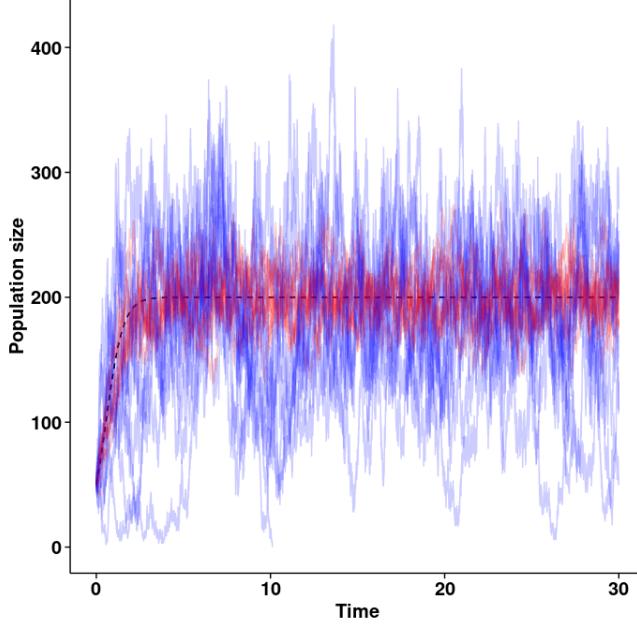


Figure D.1: Comparison of time series produced by Gillespie simulations of the stochastic logistic equation for two different parameter values. Simulations with the parameter values $\lambda = 4, \mu = 2$ are plotted in red, and simulations with the parameter values $\lambda = 31, \mu = 29$ are plotted in blue. Both simulations have $K = 200$. The infinite population limit of both simulations is plotted in black dotted lines. The graph shows 10 realizations each for the two sets of parameter values.

Thus, the weak noise approximation of D.1 is given by

$$X_t = \alpha(t) + \frac{1}{\sqrt{K}} Y_t \quad (\text{D.4})$$

where the stochastic process Y_t is an Ornstein-Uhlenbeck process given by the solution to the linear SDE

$$\begin{aligned} dY_t &= A_1^-(t)Y_t dt + \sqrt{A_0^+(t)}dW_t \\ \Rightarrow dY_t &= r(1 - 2\alpha(t))Y_t dt + \sqrt{\alpha(t)(v + r\alpha(t))}dW_t \end{aligned} \quad (\text{D.5})$$

The time series predicted by these three processes look qualitatively similar and all seem to fluctuate about the deterministic steady state (Figure D.2). The deterministic trajectory (D.3) has two fixed points, one at $x = 0$ (extinction) and one at $x = 1$ (corresponding to a population size of $n = K$). For $r > 0$, $x = 0$ is unstable and $x = 1$ is a global attractor,

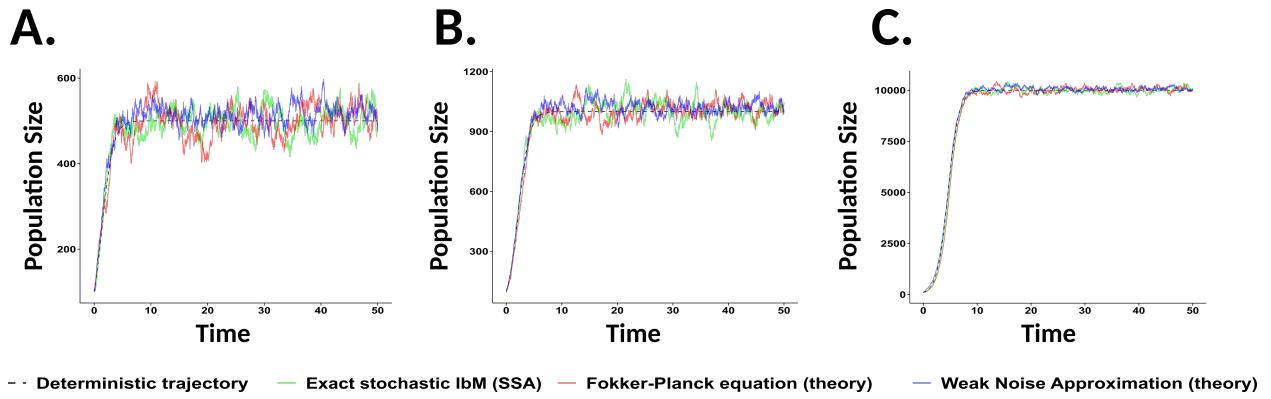


Figure D.2: Comparison of a single realization of the exact birth-death process (D.1), the deterministic trajectory (D.3), the non-linear Fokker-Planck equation (D.2), and the weak noise approximation (D.5) for (A) $K = 500$, (B) $K = 1000$, and (C) $K = 10000$. $\lambda = 2, \mu = 1$ for all thee cases.

meaning in the deterministic limit, when $r > 0$, all populations end up at $x = 1$ given enough time. The stochastic dynamics (D.2) and (D.5), however, depend not only on r , but also on v , the sum of the birth and death rates. It has been proven that $X_t = 0$ is the only recurrent state for the full stochastic dynamics (D.2), meaning that every population is guaranteed to go extinct¹ given enough time (Nåsell, 2001), thus illustrating an important difference between finite and infinite populations. $X_t = 0$ is also an ‘absorbing’ state since once a population goes extinct, it has no way of being revived in this model. However, if K is large enough, the eventual extinction of the population may take a very long time. In fact, we can make the expected time to extinction arbitrarily long by making K sufficiently large. Thus, for moderately large values of K , it is biologically meaningful only to look at a weaker version of the steady state distribution by imposing the condition that the population does not go extinct and looking at the ‘transient’ dynamics (Hastings, 2004). Conditioned on non-extinction, the solution to (D.2) has a ‘quasistationary’ distribution about the deterministic attractor $X_t = 1$, with some variance reflecting the effect of noise-induced fluctuations in population size (Nåsell, 2001) due to the finite size of the population. The weak-noise approximation (D.5) implicitly assumes non-extinction by only measuring small fluctuations from the deterministic solution to (D.3) and thus, at steady state, naturally describes a quasistationary distribution centered about $X_t = 1$. The steady-state density (probability density function as $t \rightarrow \infty$) of the exact birth-death process (D.1) is compared

¹This can be proven using tools from Markov chain theory. For those interested, the proof uses ergodicity to arrive at a contradiction if any state other than 0 exhibits a non-zero density at steady state.

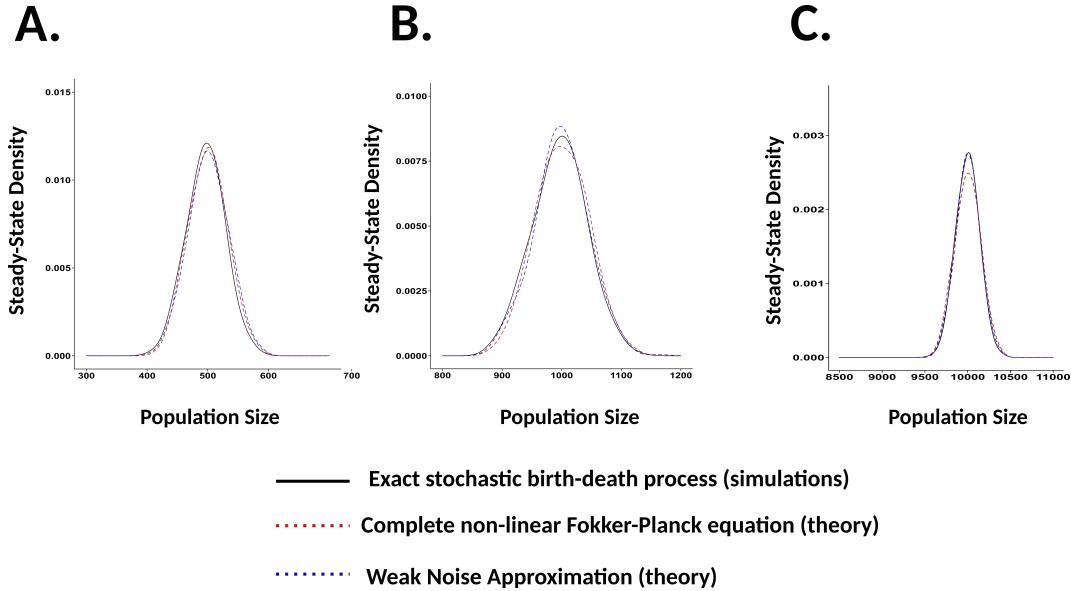


Figure D.3: Comparison of the steady-state densities given by (D.1), (D.2), and (D.5) for (A) $K = 500$, (B) $K = 1000$, and (C) $K = 10000$. $\lambda = 2, \mu = 1$ for all thee cases. Each curve was obtained using 1000 independent realizations.

with that predicted by (D.2) and (D.5) for various values of K in figure D.3.

D.2 An example for discrete traits: Lotka-Volterra and matrix games in finite populations

The methods outlined in the above section have recently been used to study the population dynamics of a finite population playing a so-called ‘matrix game’ (An evolutionary game for which you can write down a payoff matrix) with 2 pure strategies (Tao and Cressman, 2007). Based on the interpretation of what each type represents, this is mathematically equivalent to studying frequency-dependent selection on a one-locus two-allele gene (with a bijective genotype-phenotype map and no mutations) or studying two-species competitive Lotka-Volterra dynamics, as we will show below. The stochastic Lotka-Volterra competition model shown below has also been proved to be equivalent to an m -allele Moran model under certain limits (Constable and McKane, 2017).

Let us imagine a population with m types of individuals that are interacting according to some ecological rules. Let the state of the population be characterized by the vector

$\mathbf{n}(t) = [n_1(t), n_2(t), \dots, n_m(t)]^T$, where $n_i(t)$ is the number of type i individuals at time t . Let the birth and death rates of the i th type be given by:

$$\begin{aligned} b_i(\mathbf{n}) &= \left(\lambda + \frac{1}{K} \left(\sum_{j=1}^m \beta_{ij} n_j \right) \right) n_i \\ d_i(\mathbf{n}) &= \left(\mu + \frac{1}{K} \left(\sum_{j=1}^m \delta_{ij} n_j \right) \right) n_i \end{aligned} \quad (\text{D.6})$$

where $K > 0$ is our system size parameter (and represents a global carrying capacity across all types), $\lambda > 0$ and $\mu > 0$ are suitable positive constants representing the baseline natality and mortality common to all types, and β_{ij} and δ_{ij} are constants describing the effect of type j individuals on the birth and death rate of type i individuals respectively. The sign of $M_{ij} := \beta_{ij} - \delta_{ij}$ determines whether type j has a net positive or negative effect on the growth of type i . In ecological communities, this is a per-capita ecological interaction effect. In game-theoretic terms, we can interpret M_{ij} as the payoff obtained by a type j individual playing against a type i individual. I assume that $|M_{ij}| \ll K$. The values M_{ij} are often collected in an $m \times m$ matrix \mathbf{M} called the ‘payoff matrix’ (in evolutionary game theory) or ‘interaction matrix’ (in Lotka-Volterra models). Lotka-Volterra models also frequently assume that the diagonal elements M_{ii} are all equal, though I will not make that assumption here.

Going from population numbers \mathbf{n} to densities $\mathbf{x} = \mathbf{n}/K$, we obtain the birth and death rates:

$$\begin{aligned} b_i^{(K)}(\mathbf{x}) &= \left(\lambda + \sum_{j=1}^m \beta_{ij} x_j \right) x_i \\ d_i^{(K)}(\mathbf{x}) &= \left(\mu + \sum_{j=1}^m \delta_{ij} x_j \right) x_i \end{aligned} \quad (\text{D.7})$$

Thus, we have

$$A_i^\pm = x_i \left((\lambda \pm \mu) + \sum_{j=1}^m (\beta_{ij} \pm \delta_{ij}) x_j \right)$$

Defining $r := \lambda - \mu$, $\nu := \lambda + \mu$, and $T_{ij} := \beta_{ij} + \delta_{ij}$, and matching terms with equation (3.10), we can identify that we have $\mu = 0$ and

$$w_i(\mathbf{x}) = r + \sum_{j=1}^m M_{ij} x_j \quad (\text{D.8})$$

$$\tau_i(\mathbf{x}) = \nu + \sum_{j=1}^m T_{ij}x_j \quad (\text{D.9})$$

From equation (3.9), we see that the ‘mesoscopic’ description of the system is the m -dimensional SDE given by

$$d\mathbf{X}_t = \mathbf{A}^-(\mathbf{X}_t)dt + \frac{1}{\sqrt{K}}\mathbf{D}(\mathbf{X}_t)d\mathbf{W}_t \quad (\text{D.10})$$

where

$$\mathbf{A}^-_i = (\mathbf{X}_t)_i w_i(\mathbf{X}_t)$$

and

$$(\mathbf{D}\mathbf{D}^T)_i = (\mathbf{X}_t)_i \tau_i(\mathbf{X}_t)$$

From (3.16), we see that the infinite population limit of our model is a set of m coupled ODEs given by

$$\frac{dx_i}{dt} = x_i w_i(\mathbf{x}) = x_i \left(r + \sum_{j=1}^m M_{ij}x_j \right) \quad (\text{D.11})$$

These are precisely the Lotka-Volterra equations for a system of m species. If $\mathbf{p}(t) = [p_1(t), \dots, p_m(t)]^T$ is the frequency vector at time t and $N_K(t) = \sum_i x_i(t)$, then the mean fitness is given by

$$\bar{w}(t) = \sum_{i=1}^m w_i p_i \quad (\text{D.12})$$

$$= \sum_{i=1}^m \left(r + \sum_{j=1}^m M_{ij}x_j \right) p_i \quad (\text{D.13})$$

$$= r + \sum_{i=1}^m p_i \left(\sum_{j=1}^m M_{ij}x_j \right) \quad (\text{D.14})$$

where we have used the fact that $\sum_i p_i = 1$ in the last line. Using (3.17) to write down the equations for the frequencies p_i , we obtain

$$\frac{1}{N_K(t)} \frac{dp_i}{dt} = [(\mathbf{M}\mathbf{p})_i - \mathbf{p} \cdot \mathbf{M}\mathbf{p}] p_i \quad (\text{D.15})$$

which is the familiar version of the replicator equation seen in most textbooks, with an extra $N_K(t)$ factor to account for the fact that $\sum_i x_i$ is allowed to fluctuate in our model. If instead

N_K was a constant for all time, it could simply be absorbed into the definition of the payoff matrix M to obtain exactly the replicator equation as presented in most ecology/evolution textbooks. Both the stochastic dynamics (D.10) and the deterministic limit (D.11) can be simplified from an m dimensional system to an $m - 1$ dimensional system by a coordinate transformation that projects the dynamics onto an appropriate curve: If we go from the variables x_1, \dots, x_m to the variables p_1, \dots, p_{m-1}, N_K , we can exploit the fact that N_K varies much less than the p_i terms to project the system onto a ‘slow manifold’ in which N_K is approximately constant, thus obtaining an $m - 1$ dimensional system of equations and recovering the relation between the Lotka-Volterra equations for m species and the replicator equation for $m - 1$ tactics (Constable and McKane, 2017; Parsons and Rogers, 2017). However, I will not explore such dimensional reduction techniques further in this manuscript, and refer the reader to Constable et al., 2013 and Parsons and Rogers, 2017 for a review of the ideas of (stochastic) dynamics on slow manifolds.

In frequency space, the complete mesoscopic description for the stochastic dynamics can be similarly calculated in terms of the matrices \mathbf{M} and \mathbf{T} , and will yield:

$$dp_i = N_K(t) \left([(\mathbf{Mp})_i - \mathbf{p} \cdot \mathbf{Mp}] p_i - \frac{1}{KN_K} [(\mathbf{Tp})_i - \mathbf{p} \cdot \mathbf{Tp}] p_i \right) dt + \frac{1}{\sqrt{K}} \left(p_i (\mathbf{Tp})_i + p_i^2 \sum_{j=1}^m (\mathbf{Tp})_j \right)^{1/2} dW_t \quad (\text{D.16})$$

We can also carry out the weak noise approximation for this system. Let the solution to the equations (D.11) be given by $\mathbf{a}(t) = [a_1(t), \dots, a_m(t)]^\top$. For the weak noise approximation, we can Taylor expand A_i^\pm and use (3.40) to compute the directional derivative as:

$$D_i = y_i w_i(\mathbf{a}) + a_i \sum_{k=1}^m y_k \left(\frac{\partial w_i}{\partial x_k} \Big|_{\mathbf{x}=\mathbf{a}(t)} \right) \quad (\text{D.17})$$

$$= y_i w_i(\mathbf{a}) + a_i \sum_{k=1}^m y_k \left(\frac{\partial}{\partial x_k} \left(r + \sum_{j=1}^m M_{ij} x_j \right) \Big|_{\mathbf{x}=\mathbf{a}(t)} \right) \quad (\text{D.18})$$

$$= y_i w_i(\mathbf{a}) + a_i \sum_{k=1}^m y_k M_{ik} \quad (\text{D.19})$$

$$\Rightarrow D_i = y_i w_i(\mathbf{a}) + a_i w_i(\mathbf{y}) - r a_i \quad (\text{D.20})$$

where we have used the fact that $w_i(\mathbf{y}) = r + \sum_{k=1}^m y_k M_{ik}$ (from (D.8)) in the last step. Thus,

in the weak noise approximation of our process, the dynamics are given by

$$\mathbf{x}(t) = \mathbf{a}(t) + \frac{1}{\sqrt{K}} \mathbf{y}(t) \quad (\text{D.21})$$

where the stochastic fluctuations $\mathbf{y}(t)$ satisfy the linear Fokker-Planck equation

$$\frac{\partial P}{\partial t}(\mathbf{y}, t) = \sum_{i=1}^m \left(-\frac{\partial}{\partial y_i} \{ (y_i w_i(\mathbf{a}) + a_i w_i(\mathbf{y}) - r a_i) P(\mathbf{y}, t) \} + \frac{1}{2} \left(a_i \left(\nu + \sum_{j=1}^m T_{ij} a_j \right) \right) \frac{\partial^2}{\partial y_i^2} P(\mathbf{y}, t) \right) \quad (\text{D.22})$$

Using (D.20) in (3.34), we see that the fluctuations are expected to evolve as:

$$\frac{d}{dt} \mathbb{E}[y_i] = w_i(\mathbf{a}) \mathbb{E}[y_i] + a_i \sum_{k=1}^m M_{ik} \mathbb{E}[y_k] \quad (\text{D.23})$$

or, in matrix form:

$$\frac{d}{dt} \begin{bmatrix} \mathbb{E}[y_1] \\ \mathbb{E}[y_2] \\ \vdots \\ \mathbb{E}[y_i] \\ \vdots \\ \mathbb{E}[y_m] \end{bmatrix} = \begin{bmatrix} (r + \sum_{j=1}^m M_{1j} a_j + a_1 M_{11}) & a_1 M_{12} & a_1 M_{13} & \dots & \dots & \dots & a_1 M_{1m} \\ a_2 M_{21} & (r + \sum_{j=1}^m M_{2j} a_j + a_2 M_{22}) & a_2 M_{23} & \dots & \dots & \dots & a_2 M_{2m} \\ \vdots & \ddots & \ddots & & & & \vdots \\ a_i M_{i1} & a_i M_{i2} & a_i M_{i3} & \dots & (r + \sum_{j=1}^m M_{ij} a_j + a_i M_{ii}) & \dots & a_i M_{im} \\ \vdots & & & & \ddots & & \vdots \\ a_m M_{m1} & a_m M_{m2} & a_m M_{m3} & \dots & \dots & \dots & (r + \sum_{j=1}^m M_{mj} a_j + a_m M_{mm}) \end{bmatrix} \begin{bmatrix} \mathbb{E}[y_1] \\ \mathbb{E}[y_2] \\ \vdots \\ \mathbb{E}[y_i] \\ \vdots \\ \mathbb{E}[y_m] \end{bmatrix} \quad (\text{D.24})$$

The eigenvalues of the first matrix on the RHS will tell us whether the fixed point $\mathbb{E}[y_i] = 0 \forall i$ (the only fixed point of this system) is stable, or whether fluctuations are expected to grow (up to the point where the fluctuations are so large that the WNA is no longer valid). In the $m = 2$ case, Tao and Cressman, 2007 have shown that $\mathbb{E}[y_i] = 0 \forall i$ is a stable fixed point for this system iff the point \mathbf{y} is an ESS (in the usual game-theoretic sense) for the matrix game defined by the payoff matrix \mathbf{M} .

D.3 An example of systematic deviations from neutrality despite equal fitness due to noise-induced selection

To illustrate the biasing effects of noise-induced selection in otherwise neutral dynamics, I will use a simple 2 species Lotka-Volterra competition-like model where the effects of

competition are on birth rates of one species but on the death rates of the other.

To motivate this, consider a community that contains two types of birds, say type 1 and type 2. These birds compete for limited resources, but in a peculiar manner: Though the two birds feed on different food sources, the trees that type 1 birds use for nesting are the same as those that the type 2 birds rely on for food. Both types are fiercely territorial and do not tolerate other individuals of either type on either their nesting or feeding sites. Thus, competition between the two types affects the *birth rate* of type 1 birds (because they can't find good nesting sites) but the *death rate* of type 2 birds (because of starvation), whereas intratype competition affects the death rate in both cases (due to competition for food sources). Occasionally, each type can give birth to babies of the other type due to mutations. Let us construct the simplest possible model for such a system.

Let each type of bird have a constant per-capita intrinsic birth rate (rate of birth of individuals, not rate at which individuals give birth) of 1 due to reproduction. Additionally, type 1 birds face a reduction in birth rates due to competition with type 2 birds. Let us assume that the magnitude of this competition (per-capita) is equal to the per-capita competition experienced from other individuals of the same type. Both types have some additional birth rate due to rare mutations of the other type, parameterized by a mutation rate $\mu > 0$. Let n_i be the number of type i individuals (which may vary over time). Assuming trees and birds are both randomly distributed through the landscape, we arrive at the birth rates

$$\begin{aligned} b_1(n_1, n_2) &= n_1 + \mu n_2 - \frac{n_1 n_2}{K} \\ b_2(n_1, n_2) &= n_2 + \mu n_1 \end{aligned} \tag{D.25}$$

Here, the $n_1 n_2 / K$ term represents the effect of competition between types; The product $n_1 n_2$ quantifies how often a type 1 bird and a type 2 bird are expected to interact, and K is a carrying capacity for the habitat (in analogy to logistic growth or Lotka-Volterra competition), and can be thought of as a proxy for the amount of tree cover in the landscape.

For the death rates, I assume that the effect of intra-type competition on the death rate is linearly density-dependent, and thus arrive at the equations:

$$\begin{aligned} d_1(n_1, n_2) &= \frac{n_1^2}{K} \\ d_2(n_1, n_2) &= \frac{n_2^2}{K} + \frac{n_1 n_2}{K} \end{aligned} \tag{D.26}$$

Note that the effect of competition between types manifests here in an increased death rate of type 2 birds due to starvation.

Moving to density space via the change of variables $x_i = n_i/K$, letting $\mathbf{x} = [x_1, x_2]^T$, and comparing terms with equations (3.10), we can see that the per-capita fitness w_i of each type is:

$$w_1(\mathbf{x}) = w_2(\mathbf{x}) = 1 - x_1 - x_2$$

The two types of birds have the same fitness! This implies that $w_1 = w_2 = \bar{w}$ and the selection term in (3.17) vanishes. Since mutation rates are symmetric (with rate μ for both $1 \rightarrow 2$ and $2 \rightarrow 1$), we may intuitively expect that at equilibrium, both types are present in equal proportion in the population, *i.e.* $x_1 = x_2 = 1/2$ at equilibrium. Indeed, it is easy to check that this is the only fixed point in the infinite population limit.

However, if we now compute the per-capita turnover rates τ_i of each type, we see that we have

$$\begin{aligned}\tau_1(\mathbf{x}) &= 1 + x_1 - x_2 \\ \tau_2(\mathbf{x}) &= 1 + x_1 + x_2\end{aligned}$$

Thus, $\tau_1 < \tau_2$ whenever the population contains both types of individuals, and from equation (5.1), we know that this means noise-induced selection favors type 1 over type 2 in finite, fluctuating populations.

Direct simulations of the individual-based model indeed reveal that for low values of K , the fraction of individuals in the population that are of type 1 is significantly biased to be greater than 0.5, showing the effect of noise-induced selection (Figure D.4). This bias disappears for high values of K , as expected.

For this model, we can in fact quantitatively derive the effects of noise-induced selection by explicitly calculating each term of equation (5.1). Let $p = x_1/(x_1 + x_2)$ be the frequency of type 1 individuals and let $q = 1 - p$. Then, it is easy to check by direct substitution of our functional forms that we have

$$\begin{aligned}\bar{w} &= w_1(\mathbf{x}) = w_2(\mathbf{x}) \\ \bar{\tau} &= 1 + x_1 + x_2(q - p)\end{aligned}$$

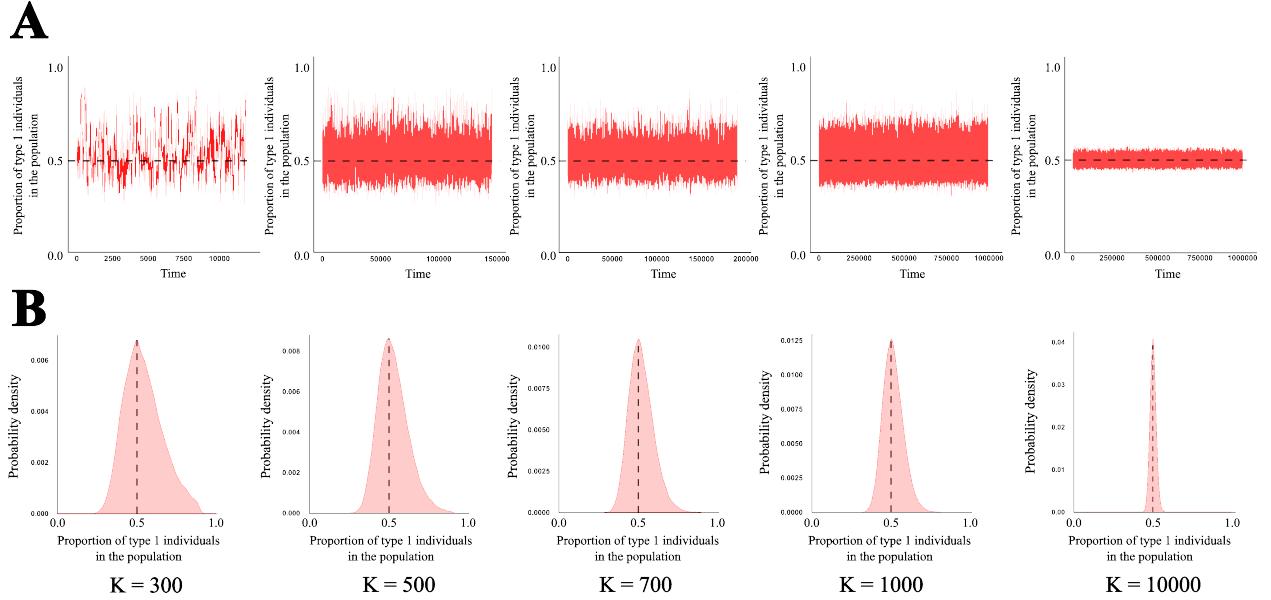


Figure D.4: (A) Time series and (B) Density estimates for p , the fraction of type 1 individuals in the population for various values of K , obtained from a direct individual-based simulation of the model defined by equations (D.25) and (D.26), simulated via the (exact) Gillespie algorithm. Dotted lines are at $p = 0.5$. At high K , the population conforms to deterministic (infinite population) predictions, but at low K , the distribution is biased towards $p > 0.5$. The time series are from single realizations. The density plots in panel (B) are estimated from 100 independent realizations, each of which were run for 10^4 timesteps. All simulations were initialized with $n_1 = n_2 = K/2$. In all simulations, $\mu = 0.05$.

$$\begin{aligned} \mu Q_i(\mathbf{p}) &= \mu p_j, \text{ where } i \neq j \\ \mu(Q_1(\mathbf{p}) - p \left(\sum_{j=1}^2 Q_j(\mathbf{p}) \right)) &= \mu(q - p) \end{aligned}$$

And thus, equation (5.1) becomes

$$dp = \left[\frac{2}{K} p^2 q + \mu \left(1 - \frac{1}{KN_K} \right) (q - p) \right] dt + \frac{1}{\sqrt{KN_K(t)}} \left[q \sqrt{A_1^+} dW_t^{(1)} - p \sqrt{A_2^+} dW_t^{(2)} \right] \quad (\text{D.27})$$

where $A_i^+ = x_i \tau_i(\mathbf{x}) + \mu Q_i(\mathbf{p})$ and each $W_t^{(i)}$ is an independent Wiener process. This equation clearly shows the biasing effect of noise-induced selection in the first component of the dt term of the RHS. Since $p^2(1 - p) > 0$ for $p \in (0, 1)$, this term always tends to increase the fraction of type 1 individuals in the population. Note that the difference in fitness between the two types remains zero if every competition term (*i.e.* every $n_i n_j / K$ term) is multiplied

by some constant $\alpha > 0$ parameterizing the strength of competition, meaning that the two types still have equal fitness. However, this constant affects the strength of noise-induced selection, and the corresponding term in equation D.27 becomes $2\alpha p^2 q/K$ instead of $2p^2 q/K$. Thus, for $\alpha > 1$, the strength of noise-induced selection (and thus the extent to which the distribution of types in the population is biased in favor of type 1 in plots like D.4) can be made arbitrarily high simply by modulating the strength of competition. The second component in the dt term captures the effects of mutations, and simply reflects the fact that we assumed that (type 1) \rightarrow (type 2) and (type 2) \rightarrow (type 1) mutations occur at the same rate μ , and thus, the net effect of mutational effects depends on the difference between the frequencies of the two types and flows towards the type with lower frequency. Finally, the two $dW_t^{(i)}$ terms are non-directional and vanish upon taking an expectation over the probability space, and therefore have no net contribution other than ‘blurring out the results’ if we look at the dynamics averaged over many realizations.

D.4 Interlude: Detecting modes in quantitative trait distributions through Fourier analysis

In Chapter 4, we used various approximations to arrive at the linear functional Fokker-Planck equation

$$\frac{\partial P}{\partial t}(\zeta, t) = \int_{\mathcal{T}} \left(-\frac{\delta}{\delta \zeta(x)} \{ \mathcal{D}_\zeta[\mathcal{A}^-](x)P(\zeta, t) \} + \frac{1}{2} \mathcal{A}^+(x|\psi) \frac{\delta^2}{\delta \zeta(x)^2} \{ P(\zeta, t) \} \right) dx \quad (\text{D.28})$$

for describing stochastic fluctuations ζ from the deterministic solution obtained by solving (4.8). Our goal is now to find a method to effectively detect and describe evolutionary branches (modes in trait space, corresponding to individual morphs) for this process. Following the methods used by Tim Rogers and colleagues for various special cases (Rogers et al., 2012a; Rogers et al., 2012b; Rogers and McKane, 2015), we will do this in a general manner by measuring the autocorrelation of the distribution of the population over trait space, a task made easier by moving to Fourier space. Specifically, a convenient theorem due to Weiner and Khinchin relates the autocorrelation of a probability distribution to its power spectral density via Fourier transformation. This has been extensively used in spatial ecology, and we too will make use of it here. We will thus restrict ourselves to cases in which we can express our focal function ϕ in terms of the Fourier basis $\{e^{ikx}\}_{k \in \mathbb{Z}}$ (Figure D.5). For example, this can be done by restricting ourselves to cases where \mathcal{T} is an interval with ‘periodic boundary

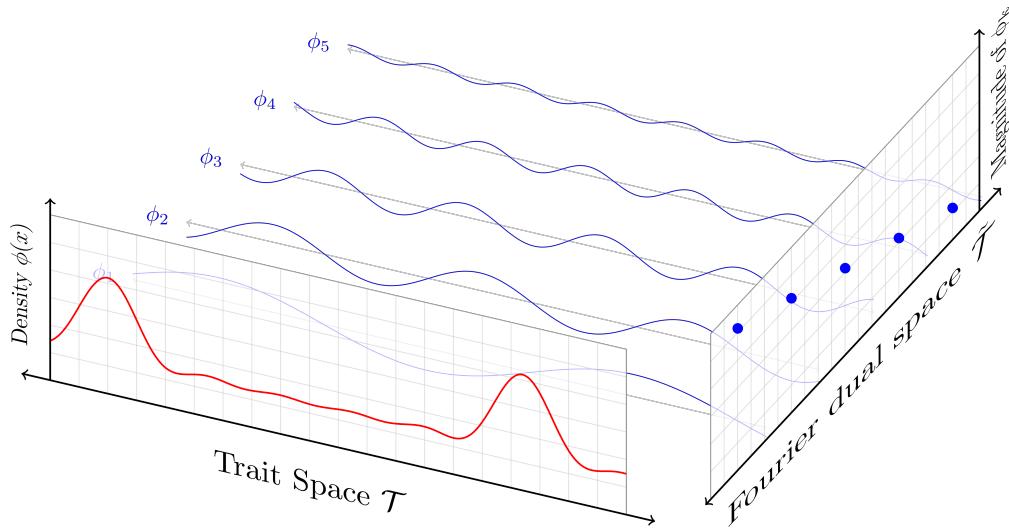


Figure D.5: Schematic description of Fourier analysis. A function $\phi(x)$ (shown in red) over the trait space can be decomposed as the sum of infinitely many Fourier modes (shown in blue) ϕ_k . In the Fourier dual space, we can look at the peaks of each of these Fourier modes: The magnitude of ϕ_k tells us how much it contributes to the actual function of interest ϕ .

conditions' (*i.e.* we will extend all our functions from \mathcal{T} to \mathbb{R} in a way that they appear periodic with period given by the length of the interval \mathcal{T}). We may also need to restrict ourselves to a ‘nice’ subspace of $\mathcal{M}(\mathcal{T})$, for example by intersecting with $L^2(\mathcal{T})$. In any case, we will assume all the prerequisites required for a Fourier basis expansion are satisfied. If $\mathcal{D}_\zeta[\mathcal{A}^-]$ is a translation-invariant² linear operator, then $\exp(ikx)$ acts as an eigenfunction, significantly simplifying the calculations. We therefore assume that $\mathcal{D}_\zeta[\mathcal{A}^-]$ takes the form:

$$\mathcal{D}_\zeta[\mathcal{A}^-](x, t) = L[\zeta(x, t)]$$

for a translation-invariant linear operator L that only depends on x and t . This is not as restrictive as it initially sounds. For example, both the Laplacian operator and the convolution operator are linear and translation invariant. The presence of phenotypic clustering and polymorphisms can be analyzed by examining the power spectrum of $\tilde{P}_0(\zeta, s)$ over the trait space, which is precisely what we will do.

²This is horrible nomenclature by the mathematicians. Though ‘invariant’ is the conventional name for this concept, the intended meaning is not really invariant but ‘equivariant’. Formally, let \mathcal{F} be a suitable function space of real valued functions. For any $c \in \mathbb{R}$, let $T_c : \mathcal{F} \rightarrow \mathcal{F}$ be the translation operator on this space, defined by $T_c[f(x)] = f(x + c)$. An operator $L : \mathcal{F} \rightarrow \mathcal{F}$ is said to be translation-invariant if it commutes with T_c for every $c \in \mathbb{R}$, *i.e.* $T_c[L[f]] = L[T_c[f]] \forall f \in \mathcal{F} \forall c \in \mathbb{R}$.

As mentioned before, we assume that ζ , and $\mathcal{A}^+(x|\psi)$ admit the Fourier basis representations:

$$\begin{aligned}\zeta(x, t) &= \sum_{k=-\infty}^{\infty} e^{ikx} \zeta_k(t) \quad ; \quad \zeta_k(t) = \int_{\mathcal{T}} \zeta(x, t) e^{-ikx} dx \\ \mathcal{A}^+(x|\psi) &= \sum_{k=-\infty}^{\infty} e^{ikx} A_k(t) \quad ; \quad A_k(t) = \int_{\mathcal{T}} \mathcal{A}^+(x|\psi) e^{-ikx} dx\end{aligned}\tag{D.29}$$

In this case, the functional derivative operator obeys:

$$\frac{\delta}{\delta \zeta(x)} = \sum_{k=-\infty}^{\infty} e^{-ikx} \frac{\partial}{\partial \zeta_k} \tag{D.30}$$

and since L is linear and translation-invariant, we also have the relation³:

$$L[\zeta] = \sum_{k=-\infty}^{\infty} L_k \zeta_k e^{ikx} \tag{D.31}$$

where

$$L_k = e^{-ikx} L[e^{ikx}]$$

Lastly, by definition of Fourier modes, we have, for any differentiable real function F and any fixed time $t > 0$:

$$\frac{\partial}{\partial \zeta_j(t)} F(\zeta_i(t)) = \delta_{ij} F'(\zeta_j(t)) \tag{D.32}$$

where δ_{ij} is the Kronecker delta symbol. Using (D.29), (D.30), and (D.31) in (D.28), we get, for the first term of the RHS:

$$\begin{aligned}& - \int_{\mathcal{T}} \frac{\delta}{\delta \zeta(x)} \{L[\zeta(x, t)] P(\zeta, t)\} dx \\ &= - \int_{\mathcal{T}} \sum_k e^{-ikx} \frac{\partial}{\partial \zeta_k} \left\{ \sum_n e^{inx} L_n \zeta_n P \right\} dx \\ &= - \int_{\mathcal{T}} \sum_k \sum_n e^{-i(k-n)x} \frac{\partial}{\partial \zeta_k} \{L_n \zeta_n P\} dx\end{aligned}$$

³This is because $\exp(ikx)$ acts as an eigenfunction for translation invariant linear operators, and therefore, for any function $\varphi = \sum \varphi_k \exp(ikx)$, we have the relation $L[\varphi] = L[\sum \varphi_k \exp(ikx)] = \sum \varphi_k L[\exp(ikx)] = \sum \varphi_k L_k \exp(ikx)$, where L_k is the eigenvalue of L associated with the eigenfunction $\exp(ikx)$. It is helpful to draw the analogy with eigenvectors of matrices and view $L_k \varphi_k$ as the projection of $L[\varphi]$ along the k th eigenvector $e_k = \exp(ikx)$.

$$= -2\pi \sum_k L_k \frac{\partial}{\partial \zeta_k} \{ \zeta_k P \} \quad (\text{D.33})$$

and for the second:

$$\begin{aligned} & \int_T \sum_k e^{ikx} A_k \left(\sum_m \sum_n e^{-i(m+n)x} \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} P \right) dx \\ &= \int_T \sum_k \sum_m \sum_n e^{i(k-m-n)x} A_k \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{ P \} dx \\ &= 2\pi \sum_m \sum_n A_{m+n} \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{ P \} \end{aligned} \quad (\text{D.34})$$

Substituting (D.33) and (D.34) into (D.28), we see that the Fokker-Planck equation (D.28) in Fourier space reads:

$$\frac{\partial P}{\partial t} = -2\pi \sum_k L_k \frac{\partial}{\partial \zeta_k} \{ \zeta_k P \} + \pi \sum_m \sum_n A_{m+n} \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{ P \} \quad (\text{D.35})$$

It is important to remember that since $\zeta(x, t)$ is a stochastic process, ζ_i is really a stochastic process and thus $\zeta_i(t)$ is actually shorthand for the random variable $(\zeta_i)_t(\omega)$, where ω is a sample path in the Fourier dual of our original probability space. Multiplying both sides of (D.35) by ζ_r and integrating over the probability space to obtain expectation values, we see that

$$\begin{aligned} \frac{d}{dt} \mathbb{E}[\zeta_r] &= -2\pi \sum_k \int \zeta_r L_k \frac{\partial}{\partial \zeta_k} \{ \zeta_k P \} d\omega + \pi \sum_m \sum_n A_{m+n} \int \zeta_r \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{ P \} d\omega \\ &= 2\pi \sum_k L_k \int \zeta_k \frac{\partial \zeta_r}{\partial \zeta_k} P d\omega + \pi \sum_m \sum_n A_{m+n} \int \frac{\partial^2 \zeta_r}{\partial \zeta_m \partial \zeta_n} P d\omega \\ &= 2\pi L_r \mathbb{E}[\zeta_r] \end{aligned} \quad (\text{D.36})$$

where we have used integration by parts and neglected the boundary term in the second step (assuming once again that P decays rapidly enough near the boundaries that this is doable), and then used (D.32) to arrive at the final expression. Similarly, multiplying (D.35) by $\zeta_r \zeta_s$, integrating over the probability space and using integration by parts, we get:

$$\frac{d}{dt} \mathbb{E}[\zeta_r \zeta_s] = 2\pi \sum_k L_k \int \zeta_k P \frac{\partial}{\partial \zeta_k} \{ \zeta_r \zeta_s \} d\omega + \pi \sum_m \sum_n A_{m+n} \int_{-\infty}^{\infty} P \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{ \zeta_r \zeta_s \} d\omega$$

$$= 2\pi(L_r + L_s)\mathbb{E}[\zeta_r \zeta_s] + \pi(A_{2r} + A_{2s}) \quad (\text{D.37})$$

At the stationary state, the LHS must be zero by definition, and we must therefore have, for every $r, s \in \mathbb{Z}$:

$$\mathbb{E}[\zeta_r \zeta_s] = -\frac{A_{2r} + A_{2s}}{2(L_r + L_s)} \quad (\text{D.38})$$

Recall that the Fourier modes of any real function φ must satisfy $\varphi_{-r} = \overline{\varphi}_r$. Since ζ , A and L are all real, we can substitute $s = -r$ in equation (D.38) to obtain the autocovariance relation:

$$\mathbb{E}[|\zeta_r|^2] = -\frac{\text{Re}(A_{2r})}{2\text{Re}(L_r)} \quad (\text{D.39})$$

The presence of phenotypic clustering can be detected using the ‘spatial covariance’ of our original process ϕ , defined as (Rogers et al., 2012a):

$$\Xi[x] = m(\mathcal{T}) \int_{\mathcal{T}} \mathbb{E}[\phi_{\infty}(x)\phi_{\infty}(y-x)]dy \quad (\text{D.40})$$

where ϕ_{∞} is the stationary state distribution of $\{\phi_t\}_t$ and m is the Lebesgue measure. We can use a spatial analogue of the Wiener-Khinchin theorem to calculate:

$$\Xi[x] = m(\mathcal{T}) \left[\int_{\mathcal{T}} \psi_{\infty}(x)\psi_{\infty}(y-x)dy + \frac{1}{K} \sum_{r=-\infty}^{\infty} \mathbb{E}[|\zeta_r|^2]e^{irx} \right] \quad (\text{D.41})$$

where the expectations in the second term are for the stationary state. A flat $\Xi[x]$ indicates that there are no clusters, and peaks indicate the presence of clusters.

D.5 An example for quantitative traits: The quantitative logistic equation

Recall the birth and death functionals given by (4.2). That is, the functionals

$$\begin{aligned} b(x|\nu) &= r \int_{\mathcal{T}} m(x, y)\nu(y)dy; \quad m(x, y) = \exp\left(\frac{-(x-y)^2}{\sigma_m^2}\right) \\ d(x|\nu) &= \frac{\nu(x)}{Kn(x)} \int_{\mathcal{T}} \alpha(x, y)\nu(y)dy; \quad \alpha(x, y) = \exp\left(\frac{-(x-y)^2}{\sigma_{\alpha}^2}\right) \end{aligned} \quad (\text{D.42})$$

corresponding to an asexual population having a constant (per-capita) birth rate r and mutations controlled by a Gaussian kernel $m(x, y)$. The death rate is density-dependent, mediated by a Gaussian competition kernel $\alpha(x, y)$, and also contains a phenotype-dependent carrying capacity controlled by $n(x)$, scaled by a constant K . The biological interpretation of the death rate is through ecological specialization for limiting resources - individuals have different intrinsic advantages (controlled by $n(x)$), and experience greater competition from conspecifics that are closer to them in phenotype space (controlled by $\alpha(x, y)$). In terms of the scaled variable $\phi = K\nu$, these functions read:

$$\begin{aligned} b_K(x|\phi) &= \frac{1}{K}b(x|\nu) = \frac{1}{K} \left(r \int_{\mathcal{T}} m(x, y) K\phi(y) dy \right) \\ d_K(x|\phi) &= \frac{1}{K}d(x|\nu) = \frac{1}{K} \left(\frac{K\phi(x)}{Kn(x)} \int_{\mathcal{T}} \alpha(x, y) K\phi(y) dy \right) \end{aligned} \quad (\text{D.43})$$

Thus, using equation (4.8), the deterministic trajectory becomes:

$$\frac{\partial \psi}{\partial t}(x, t) = r \int_{\mathcal{T}} m(x, y) \psi(y, t) dy - \frac{1}{n(x)} \psi(x, t) \int_{\mathcal{T}} \alpha(x, y) \psi(y, t) dy \quad (\text{D.44})$$

Note that if we employ the change of variables $\Psi = K\psi$ to go back from \mathcal{M}_K (*i.e.* $\phi^{(t)}$) to \mathcal{M} (*i.e.* $\nu^{(t)}$), we recover the familiar quantitative logistic equation as the deterministic limit:

$$\begin{aligned} \frac{\partial \Psi}{\partial t}(x, t) &= r \int_{\mathcal{T}} m(x, y) \Psi(y, t) dy - \frac{\Psi(x, t)}{Kn(x)} \int_{\mathcal{T}} \alpha(x, y) \Psi(y, t) dy \\ &\approx r\Psi(x, t) - \frac{\Psi(x, t)}{K(x)} \int_{\mathcal{T}} \alpha(x, y) \Psi(y, t) dy + D_m \nabla_x^2 \Psi(x, t) \end{aligned}$$

where $K(x) = Kn(x)$ is the carrying capacity experienced by an individual of phenotype x , and $D_m = r\sigma_m^2/2$ measures the ‘diffusion rate’ of the population in trait space.

We can also calculate $\mathcal{D}_\zeta[\mathcal{A}^-]$ as

$$\mathcal{D}_\zeta[\mathcal{A}^-] = \frac{d}{d\epsilon} \left(r \int_{\mathcal{T}} m(x, y) (\psi(y) + \epsilon\zeta(y)) dy - \frac{\psi(x) + \epsilon\zeta(x)}{n(x)} \int_{\mathcal{T}} \alpha(x, y) (\psi(y) + \epsilon\zeta(y)) dy \right) \Big|_{\epsilon=0}$$

$$= r \int_{\tau} m(x, y) \zeta(y) dy - \frac{1}{n(x)} \left(\psi(x) \int_{\tau} \alpha(x, y) \zeta(y) dy + \zeta(x) \int_{\tau} \alpha(x, y) \psi(y) dy \right)$$

Using this in equation (D.41), Rogers et al., 2012a (and later Rogers and McKane, 2015) have shown that the contribution of demographic stochasticity can lead to inhibition of branching, and thus, while the population undergoes infinitely many branching events in the infinite population prediction, this does not happen for finite populations. An alternative

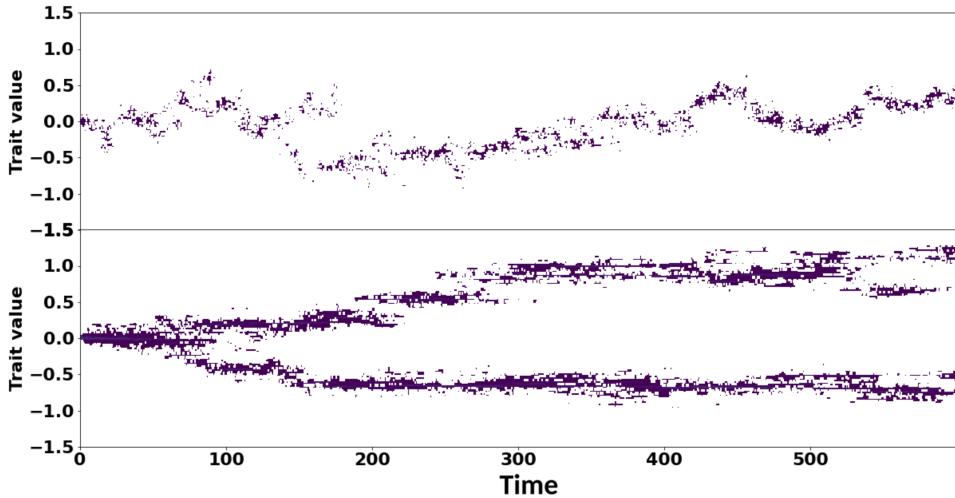


Figure D.6: Effect of population size on evolutionary branching. Two different realizations of the system (4.2) with $n(x) = \exp(-x^2/\sigma_K^2)$. Simulation parameters are $\sigma_K = 1.9$, $\sigma_\alpha = 0.7$, $\sigma_m^2 = 0.05$ for **top:** $K = 1000$ and **bottom:** $K = 10000$. Each point represents an individual. Note that the model on top remains monomorphic whereas the model on the bottom exhibits evolutionary branching, where an initially monomorphic population evolves to become dimorphic.

‘moment-based’ method that avoids moving to Fourier space has also been used to study this phenomenon of evolutionary branching and clustering in finite populations (Wakano and Iwasa, 2013; Débarre and Otto, 2016). These studies use the equation we derived in section 5.1.3 for the variance of the trait in the population and compute the conditions required for the variance to ‘explode’ (Equation A.23 in Débarre and Otto, 2016 is exactly equivalent to equation (5.9) for their choice of functional forms upon converting their change in variance to an infinitesimal rate of change *i.e.* a derivative). The method itself is relatively straightforward in principle (complications arise if the particular models are complicated) and I therefore do not explore it further in this thesis, but the broad results of such moment-based

approaches is in agreement with the predictions made from the spectral methods employed in Rogers et al., 2012a and Rogers and McKane, 2015.

It is left as an exercise for the reader to verify by the same steps that if we instead have the birth rate functional $b(x|\phi) = \lambda \int m(x,y)\phi(y)dy$ (with $m(x,y)$ as defined in (4.2)) and the death rate functional $d(x|\phi) = \phi(x)(\mu + (\lambda - \mu)\phi(x)/K)$, the infinite-population limit yields the famous Fisher-KPP equation with growth rate $r = \lambda - \mu$ and diffusion constant $D = \lambda\sigma_m^2/2$.

Appendix E

A more elegant representation of the stochastic integrals of Chapter 5

In chapter 5, we arrived at three stochastic differential equations (equations (5.1), (5.2), and (5.9)) for the frequency of a type, the population mean value of a type-level quantity, and the population variance of a type-level quantity. All three of these equations contained stochastic fluctuation terms which were of the form of a sum of stochastic integrals of several independent functions against independent Wiener processes. In this appendix, I will derive a more elegant representation of these terms via a useful lemma. I am sure this lemma is well-known, but I could not easily find a reference for it and soon realized it is quicker to prove it myself than find the appropriate reference text, and so I present a short proof below.

Lemma. Let $m \in \mathbb{N}$. Let $W_t^{(1)}, W_t^{(2)}, \dots, W_t^{(m)}$ be m independent one-dimensional Wiener processes. Let $g_1(x), g_2(x), \dots, g_m(x)$ be m ‘nice’ ($L^2(\mathbb{R})$, Lipschitz, etc.) real functions. Let

$$dX_t = \sum_{i=1}^m g_i(X_t) dW_t^{(i)}$$

Then, we can always find a *single* one-dimensional Wiener process W_t (on the same probability space) such that

$$dX_t = \left(\sum_{i=1}^m g_i^2(X_t) \right)^{1/2} dW_t$$

Proof. It suffices to prove the $m = 2$ case.

Let $dX_t = g_1(X_t)dW_t^{(1)} + g_2(X_t)dW_t^{(2)}$. Let us consider the two-dimensional process $\mathbf{W}_t = [W_t^{(1)}, W_t^{(2)}]^T$ on \mathbb{R}^2 . Let us define a function $G : \mathbb{R} \rightarrow \mathbb{R}^2$ as

$$G(x) = \frac{1}{\sqrt{g_1^2(x) + g_2^2(x)}} \begin{bmatrix} g_1(x) \\ g_2(x) \end{bmatrix} \quad (\text{E.1})$$

Now, by definition, we have

$$\int_0^t G(X_s) \cdot d\mathbf{W}_s = \int_0^t \frac{g_1(X_s)}{\sqrt{g_1^2(X_s) + g_2^2(X_s)}} dW_s^{(1)} + \int_0^t \frac{g_2(X_s)}{\sqrt{g_1^2(X_s) + g_2^2(X_s)}} dW_s^{(2)} \quad (\text{E.2})$$

By a simple corollary of the Itô isometry, we can calculate the quadratic variation of $\int G \cdot d\mathbf{W}$ as

$$\left\langle \int G(X_s) \cdot d\mathbf{W}_s \right\rangle_t = \int_0^t \|G(X_s)\|^2 d\langle \mathbf{W} \rangle_s = \int_0^t \frac{1}{g_1^2 + g_2^2} \cdot (g_1^2 + g_2^2) ds = \int_0^t ds = t \quad (\text{E.3})$$

Further, since $\int G \cdot d\mathbf{W}$ is a stochastic integral, the process $M_t = \int_0^t G(X_s) \cdot d\mathbf{W}_s$ is guaranteed to be a continuous square-integrable martingale. But, by Lévy's characterization of Brownian motion, the only continuous martingale M_t that satisfies $\langle M \rangle_t = t$ is the Brownian motion. Thus, from equation (E.3), we are led to conclude that there is a one-dimensional Wiener process W_t on the same probability space such that we can write

$$G(X_t) \cdot d\mathbf{W}_t = dW_t \quad (\text{E.4})$$

We can now use equation (E.2) on the LHS of equation (E.4) to write

$$\frac{g_1(X_t)}{\sqrt{g_1^2(X_t) + g_2^2(X_t)}} dW_t^{(1)} + \frac{g_2(X_t)}{\sqrt{g_1^2(X_t) + g_2^2(X_t)}} dW_t^{(2)} = dW_t \quad (\text{E.5})$$

By definition of our original process X_t , we can now conclude that

$$dX_t = \sqrt{g_1^2(X_t) + g_2^2(X_t)} dW_t \quad (\text{E.6})$$

thus completing the proof. \square

Using this lemma, we can now calculate the stochastic integral terms of our equations. For equation (5.1), the stochastic analog of the replicator-mutator equation, we can use this lemma and the functional form $A_i^+ = x_i \tau_i(\mathbf{x}) + \mu Q_i(\mathbf{x})$ to find that the noise term can be written as a stochastic integral against a single Wiener process W_t is given by

$$\frac{1}{\sqrt{KN_K(t)}} \left[p_i(1-p_i)\tau_i + p_i^2 \left(\sum_{j \neq i} \tau_j p_j \right) + \mu \left\{ (1-p_i)^2 Q_i(\mathbf{p}) + p_i^2 \sum_{j \neq i} Q_j(\mathbf{p}) \right\} \right]^{1/2} dW_t \quad (\text{E.7})$$

For equation (5.2), the stochastic analog of the Price equation, we have:

$$dW_{\bar{f}} = \sum_{i=1}^m (f_i - \bar{f}) \sqrt{A_i^+} dW_t^{(i)} = \left(\sum_{i=1}^m (f_i - \bar{f})^2 A_i^+ \right)^{1/2} dW_t \quad (\text{E.8})$$

where dW_t is now a single one-dimensional Wiener process. This is precisely the term calculated in equation (C.13) (barring the $1/KN_K^2$ pre-factor), and thus the stochastic term for the mean value is given by:

$$dW_{\bar{f}} = \sqrt{N_K(t) \left(\text{Cov}(\tau, (f - \bar{f})^2) + \bar{\tau} \sigma_f^2 + \mu \sum_{i=1}^m (f_i - \bar{f})^2 Q_i(\mathbf{p}) \right)} dW_t \quad (\text{E.9})$$

Similarly, for the variance equation (5.9), we can use our lemma to write

$$dW_{\sigma_f^2} = \sum_{i=1}^m (f_i - \bar{f})^2 \sqrt{A_i^+} dW_t^{(i)} = \left(\sum_{i=1}^m (f_i - \bar{f})^4 A_i^+ \right)^{1/2} dW_t \quad (\text{E.10})$$

where dW_t is now a single one-dimensional Wiener process. A calculation exactly analogous to that done in obtaining (C.13) reveals that this term can be written as

$$dW_{\sigma_f^2} = \sqrt{N_K(t) \left(\text{Cov}(\tau, (f - \bar{f})^4) + \bar{\tau} (\sigma_f^2)^2 + \mu \sum_{i=1}^m (f_i - \bar{f})^4 Q_i(\mathbf{p}) \right)} dW_t \quad (\text{E.11})$$

Appendix F

Behavior of our fundamental equation for trait frequencies at boundaries

We begin with the stochastic generalization of the replicator-mutator equation, equation (5.1). Since the p_i s describe frequencies, if the system is well-behaved, then if it begins in $[0, 1]^m$, it should remain inside $[0, 1]^m$ for all time. We are thus interested in the behavior of this equation at the boundaries of $[0, 1]^m$. Using the representation of the noise-term presented in equation (E.7), we can rewrite equation (5.1) as

$$\begin{aligned}
 dp_i(t) = & \left[(w_i(\mathbf{x}) - \bar{w})p_i + \mu \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right\} \right] dt \\
 & - \frac{1}{K} \frac{1}{N_K(t)} \left[(\tau_i(\mathbf{x}) - \bar{\tau})p_i + \mu \left\{ Q_i(\mathbf{p}) - p_i \left(\sum_{j=1}^m Q_j(\mathbf{p}) \right) \right\} \right] dt \\
 & + \frac{1}{\sqrt{KN_K(t)}} \left[p_i(1 - p_i)\tau_i + p_i^2 \left(\sum_{j \neq i} \tau_j p_j \right) + \mu \left\{ (1 - p_i)^2 Q_i(\mathbf{p}) + p_i^2 \sum_{j \neq i} Q_j(\mathbf{p}) \right\} \right]^{1/2} dW_t
 \end{aligned} \tag{F.1}$$

Now, if we let $p_i \rightarrow 0$ to look at the behavior at the 0 boundary, we are left with

$$\lim_{p_i \rightarrow 0} dp_i(t) = \lim_{p_i \rightarrow 0} \left(\mu \left[\left(1 - \frac{1}{KN_K} \right) Q_i(\mathbf{p}) \right] dt + \frac{1}{\sqrt{KN_K}} \sqrt{\mu Q_i(\mathbf{p})} dW_t \right)$$

Since $\mu \geq 0$, $Q_i \geq 0$, and $KN_K \geq 1$ by definition, all terms on the RHS are non-negative. The strongest effect is from the dt term due to the $1/\sqrt{K}$ pre-factor in the stochastic term, meaning that dp_i will almost certainly be ≥ 0 . Thus, we can conclude that

$$\lim_{p_i \rightarrow 0} \frac{dp_i}{dt} \geq 0 \text{ a.s.}$$

where the inequality is due to mutational effects. Further, now letting $\mu \rightarrow 0$ (no mutation in the system) or $Q_i(\mathbf{p}) \rightarrow 0$ (No mutations of individuals of other types into type i individuals), both terms entirely vanish, and we get

$$\lim_{\substack{p_i \rightarrow 0 \\ Q_i \rightarrow 0}} \frac{dp_i}{dt} = 0$$

which is what one would expect if things are working correctly.

We can also look at the scenario $p_i \rightarrow 1$. Note that as $p_i \rightarrow 1$, we must obviously have $p_j \rightarrow 0 \forall j \neq i$ (*i.e.* $\mathbf{p} \rightarrow e_i$, where $e_i = [0, \dots, 0, 1, 0, \dots, 0]$ is the i^{th} basis vector, with 1 in the i^{th} entry and 0 everywhere else). This means that $\bar{w} \rightarrow w_i$ and $\bar{\tau} \rightarrow \tau_i$, and thus both the selection terms in equation (F.1) vanish. We are left with

$$\lim_{p_i \rightarrow 1} dp_i(t) = -\mu \left[\left(1 - \frac{1}{KN_K} \right) \left(\lim_{\mathbf{p} \rightarrow e_i} \sum_{j \neq i}^m Q_j(\mathbf{p}) \right) \right] dt + \frac{1}{\sqrt{KN_K(t)}} \left[\mu \left(\lim_{\mathbf{p} \rightarrow e_i} \sum_{j \neq i}^m Q_j(\mathbf{p}) \right) \right]^{1/2} dW_t$$

Since by definition, $\mu \geq 0$, $Q_j \geq 0$, and $KN_K \geq 1$, we can conclude that we must have

$$\lim_{p_i \rightarrow 1} \frac{dp_i}{dt} \leq 0 \text{ a.s.}$$

which again is as expected. Note that just like before, the inequality is purely due to mutational effects. If we now impose $\mu \rightarrow 0$ (no mutation in the system) or $Q_j(\mathbf{p}) \rightarrow 0 \forall j \neq i$ (No mutations of type i individuals into individuals of other types), we will once again get

$$\lim_{\substack{p_i \rightarrow 1 \\ Q_j \rightarrow 0 \forall j \neq i}} \frac{dp_i}{dt} = 0$$

showing that our equations are always well-behaved.

References

- Abrams, P.A., Harada, Y., and Matsuda, H., 1993. On the Relationship between Quantitative Genetic and ESS Models. *Evolution* 47.3, 982–985. ISSN: 0014-3820. DOI: [10.2307/2410204](https://doi.org/10.2307/2410204).
- Abu Awad, D. and Coron, C., 2018. Effects of demographic stochasticity and life-history strategies on times and probabilities to fixation. *Heredity* 121.4, 374–386. ISSN: 1365-2540. DOI: [10.1038/s41437-018-0118-6](https://doi.org/10.1038/s41437-018-0118-6).
- Alexander, H. and Wahl, L., 2008. Fixation Probabilities Depend on Life History: Fecundity, Generation time and Survival in a burst-death model. *Evolution* 62.7, 1600–1609. ISSN: 0014-3820. DOI: [10.1111/j.1558-5646.2008.00396.x](https://doi.org/10.1111/j.1558-5646.2008.00396.x).
- Allen, B., Khwaja, A.-R., Donahue, J.L., Lattanzio, C., Dementieva, Y.A., and Sample, C., 2023. Natural selection for collective action. DOI: [10.48550/arXiv.2302.14700](https://doi.org/10.48550/arXiv.2302.14700). arXiv: [2302.14700\[q-bio\]](https://arxiv.org/abs/2302.14700).
- Allen, B. and McAvoy, A., 2019. A mathematical formalism for natural selection with arbitrary spatial and genetic structure. *Journal of Mathematical Biology* 78.4, 1147–1210. ISSN: 1432-1416. DOI: [10.1007/s00285-018-1305-z](https://doi.org/10.1007/s00285-018-1305-z).
- Assaf, M. and Meerson, B., 2017. WKB theory of large deviations in stochastic populations. *Journal of Physics A: Mathematical and Theoretical* 50.26, 263001. ISSN: 1751-8121. DOI: [10.1088/1751-8121/aa669a](https://doi.org/10.1088/1751-8121/aa669a).
- Baez, J.C., 2021. The Fundamental Theorem of Natural Selection. *Entropy* 23.11, 1436. ISSN: 1099-4300. DOI: [10.3390/e23111436](https://doi.org/10.3390/e23111436).
- Baez, J.C. and Biamonte, J.D., 2018. *Quantum techniques in stochastic mechanics* (New Jersey: World Scientific). 263 pp. ISBN: 978-981-322-693-7.
- Balan, R.M., 2018. A gentle introduction to SPDEs: the random field approach. DOI: [10.48550/arXiv.1812.02812](https://doi.org/10.48550/arXiv.1812.02812). arXiv: [1812.02812\[math\]](https://arxiv.org/abs/1812.02812).
- Balasekaran, M., Johanis, M., Rychtář, J., Taylor, D., and Zhu, J., 2022. Quasi-neutral evolution in populations under small demographic fluctuations. *Journal of Theoretical Biology* 538, 111040. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2022.111040](https://doi.org/10.1016/j.jtbi.2022.111040).
- Barabás, G., D’Andrea, R., and Stump, S.M., 2018. Chesson’s coexistence theory. *Ecological Monographs* 88.3, 277–303. ISSN: 1557-7015. DOI: [10.1002/ecm.1302](https://doi.org/10.1002/ecm.1302).

- Barton, N. and Etheridge, A., 2019. “Mathematical Models in Population Genetics”. *Handbook of Statistical Genomics*, (John Wiley & Sons, Ltd), pp. 115–20. ISBN: 978-1-119-48784-5. DOI: [10.1002/9781119487845.ch4](https://doi.org/10.1002/9781119487845.ch4).
- Behar, H., Brenner, N., Ariel, G., and Louzoun, Y., 2016. Fluctuations-induced coexistence in public goods dynamics. *Physical Biology* 13.5, 056006. ISSN: 1478-3975. DOI: [10.1088/1478-3975/13/5/056006](https://doi.org/10.1088/1478-3975/13/5/056006).
- Black, A.J. and McKane, A.J., 2012. Stochastic formulation of ecological models and their applications. *Trends in Ecology & Evolution* 27.6, 337–345. ISSN: 01695347. DOI: [10.1016/j.tree.2012.01.014](https://doi.org/10.1016/j.tree.2012.01.014).
- Boettiger, C., 2018. From noise to knowledge: how randomness generates novel phenomena and reveals information. *Ecology Letters* 21.8, 1255–1267. ISSN: 1461-0248. DOI: [10.1111/ele.13085](https://doi.org/10.1111/ele.13085).
- Bogachev, V.I., Krylov, N.V., Röckner, M., and Shaposhnikov, S.V., 2015. *Fokker-Planck-Kolmogorov equations* Mathematical surveys and monographs. (Providence, Rhode Island: American Mathematical Society). 479 pp. ISBN: 978-1-4704-2558-6.
- Boulding, K.E., 1956. General Systems Theory-The Skeleton of Science. *Management Science* 2.3, 197–208. ISSN: 0025-1909.
- Bourrat, P., 2019. Evolution is About Populations, But Its Causes are About Individuals. *Biological Theory* 14.4, 254–266. ISSN: 1555-5550. DOI: [10.1007/s13752-019-00329-3](https://doi.org/10.1007/s13752-019-00329-3).
- Brenner, S., 1999. Theoretical biology in the third millennium. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 354.1392, 1963–1965. DOI: [10.1098/rstb.1999.0535](https://doi.org/10.1098/rstb.1999.0535).
- Bressloff, P.C., 2010. Stochastic Neural Field Theory and the System-Size Expansion. *SIAM Journal on Applied Mathematics* 70.5, 1488–1521. ISSN: 0036-1399. DOI: [10.1137/090756971](https://doi.org/10.1137/090756971).
- Brown, J.S., 2016. Why Darwin would have loved evolutionary game theory. *Proceedings of the Royal Society B: Biological Sciences* 283.1838, 20160847. DOI: [10.1098/rspb.2016.0847](https://doi.org/10.1098/rspb.2016.0847).
- Buice, M.A. and Cowan, J.D., 2007. Field-theoretic approach to fluctuation effects in neural networks. *Physical Review E* 75.5, 051919. DOI: [10.1103/PhysRevE.75.051919](https://doi.org/10.1103/PhysRevE.75.051919).
- Carmona, R. and Rozovskii, B.L., 1999. *Stochastic partial differential equations: six perspectives* Mathematical surveys and monographs. (Providence, R.I.: American Mathematical Society). 334 pp. ISBN: 978-0-8218-0806-1.
- Champagnat, N., Ferrière, R., and Méléard, S., 2006. Unifying evolutionary dynamics: From individual stochastic processes to macroscopic models. *Theoretical Population Biology*. ESS Theory Now 69.3, 297–321. ISSN: 0040-5809. DOI: [10.1016/j.tpb.2005.10.004](https://doi.org/10.1016/j.tpb.2005.10.004).
- 2008. From Individual Stochastic Processes to Macroscopic Models in Adaptive Evolution. *Stochastic Models* 24 (sup1), 2–44. ISSN: 1532-6349. DOI: [10.1080/15326340802437710](https://doi.org/10.1080/15326340802437710).

- Chavhan, Y., Ali, S.I., and Dey, S., 2019. Larger Numbers Can Impede Adaptation in Asexual Populations despite Entailing Greater Genetic Variation. *Evolutionary Biology* 46.1, 1–13. ISSN: 1934-2845. DOI: [10.1007/s11692-018-9467-6](https://doi.org/10.1007/s11692-018-9467-6).
- Chavhan, Y., Malusare, S., and Dey, S., 2020. Larger bacterial populations evolve heavier fitness trade-offs and undergo greater ecological specialization. *Heredity* 124.6, 726–736. ISSN: 1365-2540. DOI: [10.1038/s41437-020-0308-x](https://doi.org/10.1038/s41437-020-0308-x).
- 2021. Interplay of population size and environmental fluctuations: A new explanation for fitness cost rarity in asexuals. *Ecology Letters* 24.9, 1943–1954. ISSN: 1461-0248. DOI: [10.1111/ele.13831](https://doi.org/10.1111/ele.13831).
- Chesson, P.L., 1982. The stabilizing effect of a random environment. *Journal of Mathematical Biology* 15.1, 1–36. ISSN: 1432-1416. DOI: [10.1007/BF00275786](https://doi.org/10.1007/BF00275786).
- 1994. Multispecies Competition in Variable Environments. *Theoretical Population Biology* 45.3, 227–276. ISSN: 0040-5809. DOI: [10.1006/tpbi.1994.1013](https://doi.org/10.1006/tpbi.1994.1013).
- Chotibut, T. and Nelson, D.R., 2015. Evolutionary dynamics with fluctuating population sizes and strong mutualism. *Physical Review E* 92.2, 022718. DOI: [10.1103/PhysRevE.92.022718](https://doi.org/10.1103/PhysRevE.92.022718).
- Chow, C.C. and Buice, M.A., 2015. Path Integral Methods for Stochastic Differential Equations. *The Journal of Mathematical Neuroscience (JMN)* 5.1, 1–35. ISSN: 2190-8567. DOI: [10.1186/s13408-015-0018-5](https://doi.org/10.1186/s13408-015-0018-5).
- Cianci, C., Fanelli, D., and McKane, A.J., 2015. WKB versus generalized van Kampen system-size expansion: the stochastic logistic equation. *arXiv*. DOI: [10.48550/arXiv.1508.00490](https://doi.org/10.48550/arXiv.1508.00490).
- Claessen, D., Andersson, J., Persson, L., and Roos, A.M.d., 2007. Delayed evolutionary branching in small populations. *Evolutionary Ecology Research* 9.1, 51–69. ISSN: 1522-0613.
- Cohen, J.E., 2004. Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better. *PLOS Biology* 2.12, e439. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.0020439](https://doi.org/10.1371/journal.pbio.0020439).
- Constable, G.W.A. and McKane, A.J., 2017. Mapping of the stochastic Lotka-Volterra model to models of population genetics and game theory. *Physical Review E* 96.2, 022416. DOI: [10.1103/PhysRevE.96.022416](https://doi.org/10.1103/PhysRevE.96.022416).
- Constable, G.W.A., McKane, A.J., and Rogers, T., 2013. Stochastic dynamics on slow manifolds. *Journal of Physics A: Mathematical and Theoretical* 46.29, 295002. ISSN: 1751-8121. DOI: [10.1088/1751-8113/46/29/295002](https://doi.org/10.1088/1751-8113/46/29/295002).
- Constable, G.W.A., Rogers, T., McKane, A.J., and Tarnita, C.E., 2016. Demographic noise can reverse the direction of deterministic selection. *Proceedings of the National Academy of Sciences* 113.32, E4745–E4754. DOI: [10.1073/pnas.1603693113](https://doi.org/10.1073/pnas.1603693113).

- Coombes, S., beim Graben, P., Potthast, R., and Wright, J., eds., 2014. *Neural Fields: Theory and Applications*. (Berlin, Heidelberg: Springer Berlin Heidelberg). ISBN: 978-3-642-54592-4. DOI: [10.1007/978-3-642-54593-1](https://doi.org/10.1007/978-3-642-54593-1).
- Coomer, M.A., Ham, L., and Stumpf, M.P.H., 2022. Noise distorts the epigenetic landscape and shapes cell-fate decisions. *Cell Systems* 13.1, 83–102.e6. ISSN: 2405-4712. DOI: [10.1016/j.cels.2021.09.002](https://doi.org/10.1016/j.cels.2021.09.002).
- Coulson, T., Benton, T.G., Lundberg, P., Dall, S.R.X., and Kendall, B.E., 2006. Putting evolutionary biology back in the ecological theatre: a demographic framework mapping genes to communities. *Evolutionary Ecology Research* 8.7, 1155–1171. ISSN: 1522-0613.
- Coulson, T., Rohani, P., and Pascual, M., 2004. Skeletons, noise and population growth: the end of an old debate? *Trends in Ecology & Evolution* 19.7, 359–364. ISSN: 0169-5347. DOI: [10.1016/j.tree.2004.05.008](https://doi.org/10.1016/j.tree.2004.05.008).
- Cressman, R. and Tao, Y., 2014. The replicator equation and other game dynamics. *Proceedings of the National Academy of Sciences* 111 (supplement _3), 10810–10817. DOI: [10.1073/pnas.1400823111](https://doi.org/10.1073/pnas.1400823111).
- Crow, J.F. and Kimura, M., 1970. *An introduction to population genetics theory* Harper international editions. (New York: Harper & Row). 591 pp.
- Czuppon, P. and Traulsen, A., 2018. Fixation probabilities in populations under demographic fluctuations. *Journal of Mathematical Biology* 77.4, 1233–1277. ISSN: 1432-1416. DOI: [10.1007/s00285-018-1251-9](https://doi.org/10.1007/s00285-018-1251-9).
- 2021. Understanding evolutionary and ecological dynamics using a continuum limit. *Ecology and Evolution* 11.11, 5857–5873. ISSN: 2045-7758. DOI: [10.1002/ece3.7205](https://doi.org/10.1002/ece3.7205).
- Da Prato, G. and Zabczyk, J., 2014. *Stochastic Equations in Infinite Dimensions* 2nd ed. Encyclopedia of Mathematics and its Applications. (Cambridge: Cambridge University Press). ISBN: 978-1-107-05584-1. DOI: [10.1017/CBO9781107295513](https://doi.org/10.1017/CBO9781107295513).
- Dawson, D.A., 1975. Stochastic evolution equations and related measure processes. *Journal of Multivariate Analysis* 5.1, 1–52. ISSN: 0047-259X. DOI: [10.1016/0047-259X\(75\)90054-8](https://doi.org/10.1016/0047-259X(75)90054-8).
- Dawson, D.A., Vaillancourt, J., and Wang, H., 2000. Stochastic partial differential equations for a class of interacting measure-valued diffusions. *Annales de l'Institut Henri Poincaré (B) Probability and Statistics* 36.2, 167–180. ISSN: 0246-0203. DOI: [10.1016/S0246-0203\(00\)00121-7](https://doi.org/10.1016/S0246-0203(00)00121-7).
- Day, T., Parsons, T.L., Lambert, A., and Gandon, S., 2020. The Price equation and evolutionary epidemiology. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, 20190357. DOI: [10.1098/rstb.2019.0357](https://doi.org/10.1098/rstb.2019.0357).
- Débarre, F. and Otto, S.P., 2016. Evolutionary dynamics of a quantitative trait in a finite asexual population. *Theoretical Population Biology* 108, 75–88. ISSN: 0040-5809. DOI: [10.1016/j.tpb.2015.12.002](https://doi.org/10.1016/j.tpb.2015.12.002).
- Del Razo, M.J., Frömberg, D., Straube, A.V., Schütte, C., Höfling, F., and Winkelmann, S., 2022. A probabilistic framework for particle-based reaction-diffusion dynamics using classical Fock space representa-

- tions. *Letters in Mathematical Physics* 112.3, 49. ISSN: 1573-0530. DOI: [10.1007/s11005-022-01539-w](https://doi.org/10.1007/s11005-022-01539-w).
- DeLong, J.P. and Cressler, C.E., 2023. Stochasticity directs adaptive evolution toward nonequilibrium evolutionary attractors. *Ecology* 104.1, e3873. ISSN: 1939-9170. DOI: [10.1002/ecy.3873](https://doi.org/10.1002/ecy.3873).
- Dembo, A. and Zeitouni, O., 1998. *Large deviations techniques and applications* 2. ed. Applications of mathematics / Springer. (New York, N.Y.: Springer). 396 pp. ISBN: 978-0-387-98406-3.
- Dieckmann, U. and Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. *Journal of Mathematical Biology* 34.5, 579–612. ISSN: 1432-1416. DOI: [10.1007/BF02409751](https://doi.org/10.1007/BF02409751).
- Dodd, P.J. and Ferguson, N.M., 2009. A Many-Body Field Theory Approach to Stochastic Models in Population Biology. *PLOS ONE* 4.9, e6855. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0006855](https://doi.org/10.1371/journal.pone.0006855).
- Doebeli, M., 2011. *Adaptive diversification* (Princeton, N.J.: Princeton University Press). ISBN: 978-0-691-12894-8.
- Doebeli, M. and Ispolatov, I., 2017. Diversity and Coevolutionary Dynamics in High-Dimensional Phenotype Spaces. *The American Naturalist* 189.2, 105–120. ISSN: 0003-0147. DOI: [10.1086/689891](https://doi.org/10.1086/689891).
- Doebeli, M., Ispolatov, Y., and Simon, B., 2017. Towards a mechanistic foundation of evolutionary theory. *eLife* 6. Wenying Shou, ed., e23804. ISSN: 2050-084X. DOI: [10.7554/eLife.23804](https://doi.org/10.7554/eLife.23804).
- Doi, M., 1976. Second quantization representation for classical many-particle system. *Journal of Physics A: Mathematical and General* 9.9, 1465. ISSN: 0305-4470. DOI: [10.1088/0305-4470/9/9/008](https://doi.org/10.1088/0305-4470/9/9/008).
- Elton, C., 1935. “eppur Si Muove.” *Journal of Animal Ecology* 4.1, 148–150. ISSN: 0021-8790. DOI: [10.2307/1225](https://doi.org/10.2307/1225).
- Engen, S., Lande, R., and Sæther, B.-E., 2009. Fixation probability of beneficial mutations in a fluctuating population. *Genetics Research* 91.1, 73–82. ISSN: 1469-5073, 0016-6723. DOI: [10.1017/S0016672308000013](https://doi.org/10.1017/S0016672308000013).
- Etheridge, A., 2000. *An introduction to superprocesses* University lecture series (Providence, R.I.) (Providence, RI: American Mathematical Society). 187 pp. ISBN: 978-0-8218-2706-2.
- , 2011. *Some mathematical models from population genetics: École d’été de Probabilités de Saint-Flour XXXIX-2009* Lecture Notes in Mathematics. (Berlin: Springer-Verlag). 119 pp. ISBN: 978-3-642-16631-0.
- Ethier, S.N. and Kurtz, T.G., 1986. *Markov processes: characterization and convergence* Wiley series in probability and mathematical statistics. (New York: Wiley). 534 pp. ISBN: 978-0-471-08186-9.
- Evans, S.N. and Perkins, E.A., 1994. Measure-Valued Branching Diffusions with Singular Interactions. *Canadian Journal of Mathematics* 46.1, 120–168. ISSN: 0008-414X, 1496-4279. DOI: [10.4153/CJM-1994-004-6](https://doi.org/10.4153/CJM-1994-004-6).

- Ewens, W.J., 1967. The probability of survival of a new mutant in a fluctuating environment. *Heredity* 22.3, 438–443. ISSN: 1365-2540. DOI: [10.1038/hdy.1967.53](https://doi.org/10.1038/hdy.1967.53).
- 2004. *Mathematical Population Genetics 1: Theoretical Introduction* (Springer Science & Business Media). 448 pp. ISBN: 978-0-387-20191-7.
- Finkelshtein, D., Kondratiev, Y., and Kutoviy, O., 2012. Semigroup approach to birth-and-death stochastic dynamics in continuum. *Journal of Functional Analysis* 262.3, 1274–1308. ISSN: 0022-1236. DOI: [10.1016/j.jfa.2011.11.005](https://doi.org/10.1016/j.jfa.2011.11.005).
- Fisher, R.A., 1930. *The genetical theory of natural selection* (Oxford: The Clarendon Press). 272 pp.
- 1931. The Distribution of Gene Ratios for Rare Mutations. *Proceedings of the Royal Society of Edinburgh* 50, 204–219. ISSN: 0370-1646. DOI: [10.1017/S0370164600044886](https://doi.org/10.1017/S0370164600044886).
- Fleming, W.H. and Viot, M., 1979. Some Measure-Valued Markov Processes in Population Genetics Theory. *Indiana University Mathematics Journal* 28.5, 817–843. ISSN: 0022-2518.
- Frank, S.A., 1997. The Price Equation, Fisher's Fundamental Theorem, Kin Selection, and Causal Analysis. *Evolution* 51.6, 1712–1729. ISSN: 1558-5646. DOI: [10.1111/j.1558-5646.1997.tb05096.x](https://doi.org/10.1111/j.1558-5646.1997.tb05096.x).
- 2012. Natural selection. IV. The Price equation*. *Journal of Evolutionary Biology* 25.6, 1002–1019. ISSN: 1420-9101. DOI: [10.1111/j.1420-9101.2012.02498.x](https://doi.org/10.1111/j.1420-9101.2012.02498.x).
- 2017. Universal expressions of population change by the Price equation: Natural selection, information, and maximum entropy production. *Ecology and Evolution* 7.10, 3381–3396. ISSN: 2045-7758. DOI: [10.1002/ece3.2922](https://doi.org/10.1002/ece3.2922).
- Frank, S.A. and Slatkin, M., 1992. Fisher's fundamental theorem of natural selection. *Trends in Ecology & Evolution* 7.3, 92–95. ISSN: 0169-5347. DOI: [10.1016/0169-5347\(92\)90248-A](https://doi.org/10.1016/0169-5347(92)90248-A).
- Gardiner, C.W., 2009. *Stochastic methods: a handbook for the natural and social sciences* (Berlin: Springer). ISBN: 978-3-540-70712-7.
- Geritz, S.A.H. and Kisdi, É., 2012. Mathematical ecology: why mechanistic models? *Journal of Mathematical Biology* 65.6, 1411–1415. ISSN: 1432-1416. DOI: [10.1007/s00285-011-0496-3](https://doi.org/10.1007/s00285-011-0496-3).
- Geritz, S.A.H., Kisdi, É., Meszéna, G., and Metz, J.A.J., 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. *Evolutionary Ecology* 12.1, 35–57. ISSN: 1573-8477. DOI: [10.1023/A:1006554906681](https://doi.org/10.1023/A:1006554906681).
- Gillespie, J.H., 1974. Natural selection for within-generation variance in offspring number. *Genetics* 76.3, 601–606. ISSN: 1943-2631. DOI: [10.1093/genetics/76.3.601](https://doi.org/10.1093/genetics/76.3.601).
- Gokhale, C.S. and Hauert, C., 2016. Eco-evolutionary dynamics of social dilemmas. *Theoretical Population Biology* 111, 28–42. ISSN: 0040-5809. DOI: [10.1016/j.tpb.2016.05.005](https://doi.org/10.1016/j.tpb.2016.05.005).
- Govaert, L., Fronhofer, E.A., Lion, S., Eizaguirre, C., Bonte, D., Egas, M., Hendry, A.P., De Brito Martins, A., Melián, C.J., Raeymaekers, J.A.M., Ratikainen, I.I., Saether, B.-E., Schweitzer, J.A., and Matthews,

- B., 2019. Eco-evolutionary feedbacks—Theoretical models and perspectives. *Functional Ecology* 33.1, 13–30. ISSN: 1365-2435. DOI: [10.1111/1365-2435.13241](https://doi.org/10.1111/1365-2435.13241).
- Grafen, A., 2014. The formal darwinism project in outline. *Biology & Philosophy* 29.2, 155–174. ISSN: 1572-8404. DOI: [10.1007/s10539-013-9414-y](https://doi.org/10.1007/s10539-013-9414-y).
- Grainger, T.N., Senthilnathan, A., Ke, P.-J., Barbour, M.A., Jones, N.T., DeLong, J.P., Otto, S.P., O'Connor, M.I., Coblenz, K.E., Goel, N., Sakarchi, J., Szojka, M.C., Levine, J.M., and Germain, R.M., 2022. An Empiricist's Guide to Using Ecological Theory. *The American Naturalist* 199.1, 1–20. ISSN: 0003-0147. DOI: [10.1086/717206](https://doi.org/10.1086/717206).
- Greenman, C.D., 2020. Duality relations between spatial birth–death processes and diffusions in Hilbert space. *Journal of Physics A: Mathematical and Theoretical* 53.44, 445002. ISSN: 1751-8121. DOI: [10.1088/1751-8121/abb1e0](https://doi.org/10.1088/1751-8121/abb1e0).
- Grilli, J., Azaele, S., Banavar, J.R., and Maritan, A., 2012. Absence of detailed balance in ecology. *Europhysics Letters* 100.3, 38002. ISSN: 0295-5075. DOI: [10.1209/0295-5075/100/38002](https://doi.org/10.1209/0295-5075/100/38002).
- Grodwohl, J.-B., 2017. “The Theory was Beautiful Indeed”: Rise, Fall and Circulation of Maximizing Methods in Population Genetics (1930–1980). *Journal of the History of Biology* 50.3, 571–608. ISSN: 1573-0387. DOI: [10.1007/s10739-016-9449-4](https://doi.org/10.1007/s10739-016-9449-4).
- Haldane, J.B.S., 1927. A Mathematical Theory of Natural and Artificial Selection, Part V: Selection and Mutation. *Mathematical Proceedings of the Cambridge Philosophical Society* 23.7, 838–844. ISSN: 1469-8064, 0305-0041. DOI: [10.1017/S0305004100015644](https://doi.org/10.1017/S0305004100015644).
- 1964. A Defense of Beanbag Genetics. *Perspectives in Biology and Medicine* 7, 343–359. ISSN: 0031-5982. DOI: [10.1353/pbm.1964.0042](https://doi.org/10.1353/pbm.1964.0042).
- Harper, J.L., 1967. A Darwinian Approach to Plant Ecology. *Journal of Ecology* 55.2, 247–270. ISSN: 0022-0477. DOI: [10.2307/2257876](https://doi.org/10.2307/2257876).
- Hastings, A., 2004. Transients: the key to long-term ecological understanding? *Trends in Ecology & Evolution* 19.1, 39–45. ISSN: 0169-5347. DOI: [10.1016/j.tree.2003.09.007](https://doi.org/10.1016/j.tree.2003.09.007).
- Hendry, A.P., 2019. A critique for eco-evolutionary dynamics. *Functional Ecology* 33.1, 84–94. ISSN: 1365-2435. DOI: [10.1111/1365-2435.13244](https://doi.org/10.1111/1365-2435.13244).
- Hochberg, D., Molina-París, C., Pérez-Mercader, J., and Visser, M., 1999. Effective action for stochastic partial differential equations. *Physical Review E* 60.6, 6343–6360. DOI: [10.1103/PhysRevE.60.6343](https://doi.org/10.1103/PhysRevE.60.6343).
- Hofbauer, J. and Sigmund, K., 1998. *Evolutionary games and population dynamics* (Cambridge: Cambridge University Press). 323 pp. ISBN: 978-0-521-62570-8.
- Houchmandzadeh, B., 2015. Fluctuation driven fixation of cooperative behavior. *Biosystems* 127, 60–66. ISSN: 0303-2647. DOI: [10.1016/j.biosystems.2014.11.006](https://doi.org/10.1016/j.biosystems.2014.11.006).

- Houchmandzadeh, B. and Vallade, M., 2012. Selection for altruism through random drift in variable size populations. *BMC Evolutionary Biology* 12.1, 61. ISSN: 1471-2148. DOI: [10.1186/1471-2148-12-61](https://doi.org/10.1186/1471-2148-12-61).
- Huang, W., Hauert, C., and Traulsen, A., 2015. Stochastic game dynamics under demographic fluctuations. *Proceedings of the National Academy of Sciences* 112.29, 9064–9069. DOI: [10.1073/pnas.1418745112](https://doi.org/10.1073/pnas.1418745112).
- Humplik, J., Hill, A.L., and Nowak, M.A., 2014. Evolutionary dynamics of infectious diseases in finite populations. *Journal of Theoretical Biology* 360, 149–162. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2014.06.039](https://doi.org/10.1016/j.jtbi.2014.06.039).
- Hutchinson, G.E., 1965. *The ecological theater and the evolutionary play* (New Haven: Yale University Press). 139 pp.
- Huynh, L., Scott, J.G., and Thomas, P.J., 2023. Inferring density-dependent population dynamics mechanisms through rate disambiguation for logistic birth-death processes. *Journal of Mathematical Biology* 86.4, 50. ISSN: 1432-1416. DOI: [10.1007/s00285-023-01877-w](https://doi.org/10.1007/s00285-023-01877-w).
- Iizuka, M., 2010. Effective population size of a population with stochastically varying size. *Journal of Mathematical Biology* 61.3, 359–375. ISSN: 1432-1416. DOI: [10.1007/s00285-009-0304-5](https://doi.org/10.1007/s00285-009-0304-5).
- Jafarpour, F., Biancalani, T., and Goldenfeld, N., 2017. Noise-induced symmetry breaking far from equilibrium and the emergence of biological homochirality. *Physical Review E* 95.3, 032407. DOI: [10.1103/PhysRevE.95.032407](https://doi.org/10.1103/PhysRevE.95.032407).
- Jhawar, J., Morris, R.G., Amith-Kumar, U.R., Danny Raj, M., Rogers, T., Rajendran, H., and Guttal, V., 2020. Noise-induced schooling of fish. *Nature Physics* 16.4, 488–493. ISSN: 1745-2481. DOI: [10.1038/s41567-020-0787-y](https://doi.org/10.1038/s41567-020-0787-y).
- Johansson, J. and Ripa, J., 2006. Will Sympatric Speciation Fail due to Stochastic Competitive Exclusion? *The American Naturalist* 168.4, 572–578. ISSN: 0003-0147. DOI: [10.1086/507996](https://doi.org/10.1086/507996).
- Johnson, B., Altrock, P.M., and Kimmel, G.J., 2021. Two-dimensional adaptive dynamics of evolutionary public goods games: finite-size effects on fixation probability and branching time. *Royal Society Open Science* 8.5, 210182. DOI: [10.1098/rsos.210182](https://doi.org/10.1098/rsos.210182).
- Johnson, E. and Hastings, A., 2022. Resolving conceptual issues in Modern Coexistence Theory. DOI: [10.48550/arXiv.2201.07926](https://doi.org/10.48550/arXiv.2201.07926). arXiv: [2201.07926\[q-bio\]](https://arxiv.org/abs/2201.07926).
- Karatzas, I. and Shreve, S.E., 1998. *Brownian motion and stochastic calculus* 2nd ed. Graduate texts in mathematics. (New York: Springer-Verlag). 470 pp. ISBN: 978-0-387-97655-6.
- Kimura, M., 1957. Some Problems of Stochastic Processes in Genetics. *The Annals of Mathematical Statistics* 28.4, 882–901. ISSN: 0003-4851.
- 1964. Diffusion models in population genetics. *Journal of Applied Probability* 1.2, 177–232. ISSN: 0021-9002, 1475-6072. DOI: [10.2307/3211856](https://doi.org/10.2307/3211856).

- 1965. A stochastic model concerning the maintenance of genetic variability in quantitative characters. *Proceedings of the National Academy of Sciences* 54.3, 731–736. DOI: [10.1073/pnas.54.3.731](https://doi.org/10.1073/pnas.54.3.731).
- Kimura, M. and Crow, J.F., 1964. The Number of Alleles that Can Be Maintained in a Finite Population. *Genetics* 49.4, 725–738. ISSN: 1943-2631. DOI: [10.1093/genetics/49.4.725](https://doi.org/10.1093/genetics/49.4.725).
- Kimura, M. and Ohta, T., 1974. Probability of Gene Fixation in an Expanding Finite Population. *Proceedings of the National Academy of Sciences* 71.9, 3377–3379. DOI: [10.1073/pnas.71.9.3377](https://doi.org/10.1073/pnas.71.9.3377).
- Kingsland, S.E., 1985. *Modeling nature: episodes in the history of population ecology* Science and its conceptual foundations. (Chicago: University of Chicago Press). 267 pp. ISBN: 978-0-226-43726-2.
- Kogan, O., Khasin, M., Meerson, B., Schneider, D., and Myers, C.R., 2014. Two-strain competition in quasineutral stochastic disease dynamics. *Physical Review E* 90.4, 042149. DOI: [10.1103/PhysRevE.90.042149](https://doi.org/10.1103/PhysRevE.90.042149).
- Kokko, H., 2021. The stagnation paradox: the ever-improving but (more or less) stationary population fitness. *Proceedings of the Royal Society B: Biological Sciences* 288.1963, 20212145. DOI: [10.1098/rspb.2021.2145](https://doi.org/10.1098/rspb.2021.2145).
- Kokko, H., Chaturvedi, A., Croll, D., Fischer, M.C., Guillaume, F., Karrenberg, S., Kerr, B., Rolshausen, G., and Stapley, J., 2017. Can Evolution Supply What Ecology Demands? *Trends in Ecology & Evolution* 32.3, 187–197. ISSN: 0169-5347. DOI: [10.1016/j.tree.2016.12.005](https://doi.org/10.1016/j.tree.2016.12.005).
- Konno, N. and Shiga, T., 1988. Stochastic partial differential equations for some measure-valued diffusions. *Probability Theory and Related Fields* 79.2, 201–225. ISSN: 1432-2064. DOI: [10.1007/BF00320919](https://doi.org/10.1007/BF00320919).
- Kuosmanen, T., Särkkä, S., and Mustonen, V., 2022. Turnover shapes evolution of birth and death rates. DOI: [10.1101/2022.07.11.499527](https://doi.org/10.1101/2022.07.11.499527).
- Lambert, A., 2006. Probability of fixation under weak selection: A branching process unifying approach. *Theoretical Population Biology* 69.4, 419–441. ISSN: 0040-5809. DOI: [10.1016/j.tpb.2006.01.002](https://doi.org/10.1016/j.tpb.2006.01.002).
- 2010. Population genetics, ecology and the size of populations. *Journal of Mathematical Biology* 60.3, 469–472. ISSN: 1432-1416. DOI: [10.1007/s00285-009-0286-3](https://doi.org/10.1007/s00285-009-0286-3).
- Lande, R., 1976. Natural Selection and Random Genetic Drift in Phenotypic Evolution. *Evolution* 30.2, 314–334. ISSN: 0014-3820. DOI: [10.2307/2407703](https://doi.org/10.2307/2407703).
- 1982. A Quantitative Genetic Theory of Life History Evolution. *Ecology* 63.3, 607–615. ISSN: 1939-9170. DOI: [10.2307/1936778](https://doi.org/10.2307/1936778).
- Lehtonen, J., 2018. The Price Equation, Gradient Dynamics, and Continuous Trait Game Theory. *The American Naturalist* 191.1, 146–153. ISSN: 0003-0147. DOI: [10.1086/694891](https://doi.org/10.1086/694891).
- 2020. The Price equation and the unity of social evolution theory. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, 20190362. DOI: [10.1098/rstb.2019.0362](https://doi.org/10.1098/rstb.2019.0362).

- Levins, R., 1966. The Strategy of Model Building in Population Biology. *American Scientist* 54.4, 421–431. ISSN: 0003-0996.
- Lin, Y.T., Kim, H., and Doering, C.R., 2012. Features of Fast Living: On the Weak Selection for Longevity in Degenerate Birth-Death Processes. *Journal of Statistical Physics* 148.4, 647–663. ISSN: 1572-9613. DOI: [10.1007/s10955-012-0479-9](https://doi.org/10.1007/s10955-012-0479-9).
- Lion, S., 2018. Theoretical Approaches in Evolutionary Ecology: Environmental Feedback as a Unifying Perspective. *The American Naturalist* 191.1, 21–44. ISSN: 0003-0147. DOI: [10.1086/694865](https://doi.org/10.1086/694865).
- Lion, S., Boots, M., and Sasaki, A., 2022. Multimorph Eco-Evolutionary Dynamics in Structured Populations. *The American Naturalist* 200.3, 345–372. ISSN: 0003-0147. DOI: [10.1086/720439](https://doi.org/10.1086/720439).
- Lion, S., Sasaki, A., and Boots, M., 2023. Extending eco-evolutionary theory with oligomorphic dynamics. *Ecology Letters* n/a (n/a). ISSN: 1461-0248. DOI: [10.1111/ele.14183](https://doi.org/10.1111/ele.14183).
- Liu, W. and Röckner, M., 2015. *Stochastic Partial Differential Equations: An Introduction* Universitext. (Cham: Springer International Publishing). ISBN: 978-3-319-22353-7. DOI: [10.1007/978-3-319-22354-4](https://doi.org/10.1007/978-3-319-22354-4).
- Luque, V.J., 2017. One equation to rule them all: a philosophical analysis of the Price equation. *Biology & Philosophy* 32.1, 97–125. ISSN: 1572-8404. DOI: [10.1007/s10539-016-9538-y](https://doi.org/10.1007/s10539-016-9538-y).
- Luque, V.J. and Baravalle, L., 2021. The mirror of physics: on how the Price equation can unify evolutionary biology. *Synthese* 199.5, 12439–12462. ISSN: 1573-0964. DOI: [10.1007/s11229-021-03339-6](https://doi.org/10.1007/s11229-021-03339-6).
- Mathot, K.J. and Frankenhus, W.E., 2018. Models of pace-of-life syndromes (POLS): a systematic review. *Behavioral Ecology and Sociobiology* 72.3, 41. ISSN: 1432-0762. DOI: [10.1007/s00265-018-2459-9](https://doi.org/10.1007/s00265-018-2459-9).
- Mazzolini, A. and Grilli, J., 2022. Universality of evolutionary trajectories under arbitrary competition dynamics. DOI: [10.1101/2021.06.17.448795](https://doi.org/10.1101/2021.06.17.448795).
- McAvoy, A., Fraiman, N., Hauert, C., Wakeley, J., and Nowak, M.A., 2018. Public goods games in populations with fluctuating size. *Theoretical Population Biology* 121, 72–84. ISSN: 0040-5809. DOI: [10.1016/j.tpb.2018.01.004](https://doi.org/10.1016/j.tpb.2018.01.004).
- McLeod, D.V. and Day, T., 2019. Social evolution under demographic stochasticity. *PLOS Computational Biology* 15.2, e1006739. ISSN: 1553-7358. DOI: [10.1371/journal.pcbi.1006739](https://doi.org/10.1371/journal.pcbi.1006739).
- Melbinger, A., Cremer, J., and Frey, E., 2010. Evolutionary Game Theory in Growing Populations. *Physical Review Letters* 105.17, 178101. DOI: [10.1103/PhysRevLett.105.178101](https://doi.org/10.1103/PhysRevLett.105.178101).
- Metcalf, C.J.E. and Pavard, S., 2007a. Why evolutionary biologists should be demographers. *Trends in Ecology & Evolution* 22.4, 205–212. ISSN: 0169-5347. DOI: [10.1016/j.tree.2006.12.001](https://doi.org/10.1016/j.tree.2006.12.001).
- 2007b. All paths to fitness lead through demography. *Trends in Ecology & Evolution* 22.11, 563–564. ISSN: 0169-5347. DOI: [10.1016/j.tree.2007.07.003](https://doi.org/10.1016/j.tree.2007.07.003).

- Metz, J.A.J., Nisbet, R.M., and Geritz, S.A.H., 1992. How should we define ‘fitness’ for general ecological scenarios? *Trends in Ecology & Evolution* 7.6, 198–202. ISSN: 0169-5347. DOI: [10.1016/0169-5347\(92\)90073-K](https://doi.org/10.1016/0169-5347(92)90073-K).
- Millikan, R.A., 1924. The electron and the light-quant from the experimental point of view. In *Nobel Lectures*, 1923 Nobel Prize in Physics. Nobel Lectures. (Amsterdam, Netherlands: Elsevier Publishing Company, Amsterdam, 1965). DOI: <https://www.nobelprize.org/uploads/2018/06/millikan-lecture.pdf>.
- Moran, P.a.P., 1958. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* 54.1, 60–71. ISSN: 1469-8064, 0305-0041. DOI: [10.1017/S0305004100033193](https://doi.org/10.1017/S0305004100033193).
- Nåsell, I., 2001. Extinction and Quasi-stationarity in the Verhulst Logistic Model. *Journal of Theoretical Biology* 211.1, 11–27. ISSN: 0022-5193. DOI: [10.1006/jtbi.2001.2328](https://doi.org/10.1006/jtbi.2001.2328).
- Nordbotten, J.M., Bokma, F., Hermansen, J.S., and Stenseth, N.C., 2020. The dynamics of trait variance in multi-species communities. *Royal Society Open Science* 7.8, 200321. DOI: [10.1098/rsos.200321](https://doi.org/10.1098/rsos.200321).
- Ó Laighléis, E., Evans, M.R., and Blythe, R.A., 2018. Minimal stochastic field equations for one-dimensional flocking. *Physical Review E* 98.6, 062127. DOI: [10.1103/PhysRevE.98.062127](https://doi.org/10.1103/PhysRevE.98.062127).
- O'Dwyer, J.P. and Green, J.L., 2010. Field theory for biogeography: a spatially explicit model for predicting patterns of biodiversity. *Ecology Letters* 13.1, 87–95. ISSN: 1461-0248. DOI: [10.1111/j.1461-0248.2009.01404.x](https://doi.org/10.1111/j.1461-0248.2009.01404.x).
- O'Dwyer, J.P., Lake, J.K., Ostling, A., Savage, V.M., and Green, J.L., 2009. An integrative framework for stochastic, size-structured community assembly. *Proceedings of the National Academy of Sciences* 106.15, 6170–6175. DOI: [10.1073/pnas.0813041106](https://doi.org/10.1073/pnas.0813041106).
- Okasha, S., 2006. *Evolution and the levels of selection* (Oxford, Oxford: Clarendon Press ; Oxford University Press). 263 pp. ISBN: 978-0-19-926797-2.
- Øksendal, B.K., 1998. *Stochastic differential equations: an introduction with applications* (Berlin; New York: Springer). ISBN: 978-3-662-03620-4.
- Oliveira, M.M.d. and Dickman, R., 2017. The advantage of being slow: The quasi-neutral contact process. *PLOS ONE* 12.8, e0182672. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0182672](https://doi.org/10.1371/journal.pone.0182672).
- Otto, S.P. and Whitlock, M.C., 1997. The Probability of Fixation in Populations of Changing Size. *Genetics* 146.2, 723–733. ISSN: 1943-2631. DOI: [10.1093/genetics/146.2.723](https://doi.org/10.1093/genetics/146.2.723).
- Page, K.M. and Nowak, M.A., 2002. Unifying Evolutionary Dynamics. *Journal of Theoretical Biology* 219.1, 93–98. ISSN: 0022-5193. DOI: [10.1006/jtbi.2002.3112](https://doi.org/10.1006/jtbi.2002.3112).
- Parsons, T.L., Lambert, A., Day, T., and Gandon, S., 2018. Pathogen evolution in finite populations: slow and steady spreads the best. *Journal of The Royal Society Interface* 15.147, 20180135. DOI: [10.1098/rsif.2018.0135](https://doi.org/10.1098/rsif.2018.0135).

- Parsons, T.L., Quince, C., and Plotkin, J.B., 2010. Some Consequences of Demographic Stochasticity in Population Genetics. *Genetics* 185.4, 1345–1354. ISSN: 1943-2631. DOI: [10.1534/genetics.110.115030](https://doi.org/10.1534/genetics.110.115030).
- Parsons, T.L. and Rogers, T., 2017. Dimension reduction for stochastic dynamical systems forced onto a manifold by large drift: a constructive approach with examples from theoretical biology. *Journal of Physics A: Mathematical and Theoretical* 50.41, 415601. ISSN: 1751-8121. DOI: [10.1088/1751-8121/aa86c7](https://doi.org/10.1088/1751-8121/aa86c7).
- Peliti, L., 1985. Path integral approach to birth-death processes on a lattice. *Journal de Physique* 46.9, 1469–1483. ISSN: 0302-0738, 2777-3396. DOI: [10.1051/jphys:019850046090146900](https://doi.org/10.1051/jphys:019850046090146900).
- Potochnik, A., 2018. *Idealization and the aims of science* (). ISBN: 978-0-226-50705-7.
- Prévôt, C. and Röckner, M., 2007. *A Concise Course on Stochastic Partial Differential Equations* (Springer Science & Business Media). 149 pp. ISBN: 978-3-540-70780-6.
- Price, G.R., 1972. Fisher's 'fundamental theorem' made clear. *Annals of Human Genetics* 36.2, 129–140. ISSN: 1469-1809. DOI: [10.1111/j.1469-1809.1972.tb00764.x](https://doi.org/10.1111/j.1469-1809.1972.tb00764.x).
- Proulx, S.R. and Day, T., 2005. What can Invasion Analyses Tell us about Evolution under Stochasticity in Finite Populations? *Selection* 2.1, 2–15. ISSN: 1585-1931, 1588-287X. DOI: [10.1556/select.2.2001.1-2.2](https://doi.org/10.1556/select.2.2001.1-2.2).
- Provine, W.B., 2001. *The Origins of Theoretical Population Genetics: With a New Afterword* (Chicago, IL: University of Chicago Press). 240 pp. ISBN: 978-0-226-68464-2.
- Queller, D.C., 2017. Fundamental Theorems of Evolution. *The American Naturalist* 189.4, 345–353. ISSN: 0003-0147. DOI: [10.1086/690937](https://doi.org/10.1086/690937).
- Raatz, M. and Traulsen, A., 2023. Promoting extinction or minimizing growth? The impact of treatment on trait trajectories in evolving populations. DOI: [10.1101/2022.06.17.496570](https://doi.org/10.1101/2022.06.17.496570).
- Reimers, M., 1989. One dimensional stochastic partial differential equations and the branching measure diffusion. *Probability Theory and Related Fields* 81.3, 319–340. ISSN: 1432-2064. DOI: [10.1007/BF00340057](https://doi.org/10.1007/BF00340057).
- Rice, S.H., 2004. *Evolutionary theory: mathematical and conceptual foundations* (Sunderland, Mass., USA: Sinauer Associates). ISBN: 978-0-87893-702-8.
- 2008. A stochastic version of the Price equation reveals the interplay of deterministic and stochastic processes in evolution. *BMC Evolutionary Biology* 8.1, 262. ISSN: 1471-2148. DOI: [10.1186/1471-2148-8-262](https://doi.org/10.1186/1471-2148-8-262).
 - 2020. Universal rules for the interaction of selection and transmission in evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, 20190353. DOI: [10.1098/rstb.2019.0353](https://doi.org/10.1098/rstb.2019.0353).
- Rice, S.H. and Papadopoulos, A., 2009. Evolution with Stochastic Fitness and Stochastic Migration. *PLOS ONE* 4.10, e7130. ISSN: 1932-6203. DOI: [10.1371/journal.pone.0007130](https://doi.org/10.1371/journal.pone.0007130).

- Rogers, T., McKane, A.J., and Rossberg, A.G., 2012a. Demographic noise can lead to the spontaneous formation of species. *Europhysics Letters* 97.4, 40008. ISSN: 0295-5075. DOI: [10.1209/0295-5075/97/40008](https://doi.org/10.1209/0295-5075/97/40008).
- Rogers, T. and McKane, A.J., 2015. Modes of competition and the fitness of evolved populations. *Physical Review E* 92.3, 032708. DOI: [10.1103/PhysRevE.92.032708](https://doi.org/10.1103/PhysRevE.92.032708).
- Rogers, T., McKane, A.J., and Rossberg, A.G., 2012b. Spontaneous genetic clustering in populations of competing organisms. *Physical Biology* 9.6, 066002. ISSN: 1478-3975. DOI: [10.1088/1478-3975/9/6/066002](https://doi.org/10.1088/1478-3975/9/6/066002).
- Sanchez, A. and Gore, J., 2013. Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations. *PLOS Biology* 11.4, e1001547. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.1001547](https://doi.org/10.1371/journal.pbio.1001547).
- Sasaki, A. and Dieckmann, U., 2011. Oligomorphic dynamics for analyzing the quantitative genetics of adaptive speciation. *Journal of Mathematical Biology* 63.4, 601–635. ISSN: 1432-1416. DOI: [10.1007/s00285-010-0380-6](https://doi.org/10.1007/s00285-010-0380-6).
- Saunders, P.A., Neuenschwander, S., and Perrin, N., 2018. Sex chromosome turnovers and genetic drift: a simulation study. *Journal of Evolutionary Biology* 31.9, 1413–1419. ISSN: 1420-9101. DOI: [10.1111/jeb.13336](https://doi.org/10.1111/jeb.13336).
- Schoener, T.W., 2011. The Newest Synthesis: Understanding the Interplay of Evolutionary and Ecological Dynamics. *Science* 331.6016, 426–429. DOI: [10.1126/science.1193954](https://doi.org/10.1126/science.1193954).
- Schreiber, S.J., Levine, J.M., Godoy, O., Kraft, N.J.B., and Hart, S.P., 2022. Does deterministic coexistence theory matter in a finite world? *Ecology*, e3838. ISSN: 1939-9170. DOI: [10.1002/ecy.3838](https://doi.org/10.1002/ecy.3838).
- Servedio, M.R., Brandvain, Y., Dhole, S., Fitzpatrick, C.L., Goldberg, E.E., Stern, C.A., Cleve, J.V., and Yeh, D.J., 2014. Not Just a Theory—The Utility of Mathematical Models in Evolutionary Biology. *PLOS Biology* 12.12, e1002017. ISSN: 1545-7885. DOI: [10.1371/journal.pbio.1002017](https://doi.org/10.1371/journal.pbio.1002017).
- Shoemaker, L.G., Sullivan, L.L., Donohue, I., Cabral, J.S., Williams, R.J., Mayfield, M.M., Chase, J.M., Chu, C., Harpole, W.S., Huth, A., HilleRisLambers, J., James, A.R.M., Kraft, N.J.B., May, F., Muthukrishnan, R., Satterlee, S., Taubert, F., Wang, X., Wiegand, T., Yang, Q., and Abbott, K.C., 2020. Integrating the underlying structure of stochasticity into community ecology. *Ecology* 101.2, e02922. ISSN: 1939-9170. DOI: [10.1002/ecy.2922](https://doi.org/10.1002/ecy.2922).
- Simon, B., 2014. Continuous-time models of group selection, and the dynamical insufficiency of kin selection models. *Journal of Theoretical Biology* 349, 22–31. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2014.01.030](https://doi.org/10.1016/j.jtbi.2014.01.030).
- Sjödin, P., Kaj, I., Krone, S., Lascoux, M., and Nordborg, M., 2005. On the Meaning and Existence of an Effective Population Size. *Genetics* 169.2, 1061–1070. ISSN: 1943-2631. DOI: [10.1534/genetics.104.026799](https://doi.org/10.1534/genetics.104.026799).

- Stearns, S.C., 1977. The Evolution of Life History Traits: A Critique of the Theory and a Review of the Data. *Annual Review of Ecology and Systematics* 8.1, 145–171. DOI: [10.1146/annurev.es.08.110177.001045](https://doi.org/10.1146/annurev.es.08.110177.001045).
- Strang, A.G., Abbott, K.C., and Thomas, P.J., 2019. How to avoid an extinction time paradox. *Theoretical Ecology* 12.4, 467–487. ISSN: 1874-1746. DOI: [10.1007/s12080-019-0416-5](https://doi.org/10.1007/s12080-019-0416-5).
- Svensson, E.I., 2019. Eco-evolutionary dynamics of sexual selection and sexual conflict. *Functional Ecology* 33.1, 60–72. ISSN: 1365-2435. DOI: [10.1111/1365-2435.13245](https://doi.org/10.1111/1365-2435.13245).
- Tao, Y. and Cressman, R., 2007. Stochastic Fluctuations Through Intrinsic Noise in Evolutionary Game Dynamics. *Bulletin of Mathematical Biology* 69.4, 1377–1399. ISSN: 1522-9602. DOI: [10.1007/s11538-006-9170-0](https://doi.org/10.1007/s11538-006-9170-0).
- Thompson, R.P., 2014. “Darwin’s theory and the value of mathematical formalization”. *Evolutionary Biology: Conceptual, Ethical, and Religious Issues*, Denis Walsh and R. Paul Thompson, ed. (Cambridge: Cambridge University Press), pp. 109–136. ISBN: 978-1-107-02701-5. DOI: [10.1017/CBO9781139208796.008](https://doi.org/10.1017/CBO9781139208796.008).
- Van Kampen, N.G., 1981. *Stochastic processes in physics and chemistry* (Amsterdam, New York, New York: North-Holland). xiv, 419. ISBN: 978-0-444-86200-6.
- Van Veelen, M., 2005. On the use of the Price equation. *Journal of Theoretical Biology* 237.4, 412–426. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2005.04.026](https://doi.org/10.1016/j.jtbi.2005.04.026).
- , 2020. The problem with the Price equation. *Philosophical Transactions of the Royal Society B: Biological Sciences* 375.1797, 20190355. DOI: [10.1098/rstb.2019.0355](https://doi.org/10.1098/rstb.2019.0355).
- Vellend, M., 2016. *The Theory of Ecological Communities* (Princeton University Press). ISBN: 978-1-4008-8379-0.
- Veller, C., Muralidhar, P., Constable, G.W.A., and Nowak, M.A., 2017. Drift-Induced Selection Between Male and Female Heterogamety. *Genetics* 207.2, 711–727. ISSN: 1943-2631. DOI: [10.1534/genetics.117.300151](https://doi.org/10.1534/genetics.117.300151).
- Vittadello, S.T. and Stumpf, M.P.H., 2022. Open problems in mathematical biology. *Mathematical Biosciences* 354, 108926. ISSN: 0025-5564. DOI: [10.1016/j.mbs.2022.108926](https://doi.org/10.1016/j.mbs.2022.108926).
- Wakano, J.Y. and Iwasa, Y., 2013. Evolutionary Branching in a Finite Population: Deterministic Branching vs. Stochastic Branching. *Genetics* 193.1, 229–241. ISSN: 1943-2631. DOI: [10.1534/genetics.112.144980](https://doi.org/10.1534/genetics.112.144980).
- Walsh, J.B., 1986. An introduction to stochastic partial differential equations. In *École d’Été de Probabilités de Saint Flour XIV - 1984*, René Carmona, Harry Kesten, John B. Walsh, and P. L. Hennequin, ed. Lecture Notes in Mathematics. (Berlin, Heidelberg: Springer), pp. 265–439. ISBN: 978-3-540-39781-6. DOI: [10.1007/BFb0074920](https://doi.org/10.1007/BFb0074920).

- Wang, G., Su, Q., Wang, L., and Plotkin, J.B., 2023. Reproductive variance can drive behavioral dynamics. *Proceedings of the National Academy of Sciences* 120.12, e2216218120. DOI: [10.1073/pnas.2216218120](https://doi.org/10.1073/pnas.2216218120).
- Waxman, D., 2011. A Unified Treatment of the Probability of Fixation when Population Size and the Strength of Selection Change Over Time. *Genetics* 188.4, 907–913. ISSN: 1943-2631. DOI: [10.1534/genetics.111.129288](https://doi.org/10.1534/genetics.111.129288).
- Weber, M.F. and Frey, E., 2017. Master equations and the theory of stochastic path integrals. *Reports on Progress in Physics* 80.4, 046601. ISSN: 0034-4885. DOI: [10.1088/1361-6633/aa5ae2](https://doi.org/10.1088/1361-6633/aa5ae2).
- Week, B., Nuismer, S.L., Harmon, L.J., and Krone, S.M., 2021. A white noise approach to evolutionary ecology. *Journal of Theoretical Biology* 521, 110660. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2021.110660](https://doi.org/10.1016/j.jtbi.2021.110660).
- Wickman, J., Koffel, T., and Klausmeier, C.A., 2022. A theoretical framework for trait-based eco-evolutionary dynamics: population structure, intraspecific variation, and community assembly. *The American Naturalist*. ISSN: 0003-0147. DOI: [10.1086/723406](https://doi.org/10.1086/723406).
- Wienand, K., Frey, E., and Mobilia, M., 2017. Evolution of a Fluctuating Population in a Randomly Switching Environment. *Physical Review Letters* 119.15, 158301. DOI: [10.1103/PhysRevLett.119.158301](https://doi.org/10.1103/PhysRevLett.119.158301).
- Wodarz, D., Goel, A., and Komarova, N.L., 2017. Effect of cell cycle duration on somatic evolutionary dynamics. *Evolutionary Applications* 10.10, 1121–1129. ISSN: 1752-4571. DOI: [10.1111/eva.12518](https://doi.org/10.1111/eva.12518).
- Wright, J., Bolstad, G.H., Araya-Ajoy, Y.G., and Dingemanse, N.J., 2019. Life-history evolution under fluctuating density-dependent selection and the adaptive alignment of pace-of-life syndromes. *Biological Reviews* 94.1, 230–247. ISSN: 1469-185X. DOI: [10.1111/brv.12451](https://doi.org/10.1111/brv.12451).
- Wright, S., 1931. Evolution in Mendelian Populations. *Genetics* 16.2, 97–159. ISSN: 1943-2631. DOI: [10.1093/genetics/16.2.97](https://doi.org/10.1093/genetics/16.2.97).
- Zuk, M. and Travisano, M., 2018. Models on the Runway: How Do We Make Replicas of the World? *The American Naturalist* 192.1, 1–9. ISSN: 0003-0147. DOI: [10.1086/697508](https://doi.org/10.1086/697508).