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

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 study/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:

Vishwesha Guttal

Rohini Balakrishnan

Sutirth Dey

This thesis is dedicated to ?

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 Centre for Ecological Sciences, Indian Institute of Science, under the supervision of Vishwesha Guttal, with Rohini Balakrishnan as a co-supervisor, and the same has not been submitted elsewhere for any other degree.

Ananda Shikhara Bhat | Vishwesha Guttal | Rohini Balakrishnan

Acknowledgments

Not more than 250 words

Abstract

Population biology is built on a strong mathematical foundation developed during the Modern Synthesis through fields such as theoretical population genetics, evolutionary game theory, and quantitative genetics. Historically, these formalisms have often worked with infinite populations, ignoring the effects of demographic stochasticity. Finite population models in population genetics usually assume a fixed population size and are of limited applicability in the real world, where population sizes routinely fluctuate. In this thesis, I use ideas from statistical physics to analytically describe evolving populations from biological first principles. Starting from a density-dependent 'birth-death process' describing an arbitrary closed population of individuals with discrete traits, I derive a set of stochastic differential equations (SDEs) for how trait frequencies change over time. Along with recovering the effects of the standard evolutionary forces of selection, mutation, and drift, these SDEs also reveal a new directional evolutionary force, 'noise-induced selection', that is particular to finite populations and has been largely overlooked in standard formulations of evolution. Noise-induced selection can reverse the direction of evolution predicted by infinite-population frameworks, with implications for simulation studies and real world populations. Well-known results such as the replicator-mutator equation and Fisher's fundamental theorem are recovered in the infinite population limit. Finally, I extend these ideas to one-dimensional quantitative traits through a 'stochastic field theory' that yields frameworks such as Kimura's continuum-of-alleles and gradient dynamics in the infinite population limit. My work thus generalizes the formal structures of population biology to finite fluctuating populations and predicts a new evolutionary force unique to such populations.

Contents

List of Tables

List of Figures

Part I

Background

Chapter 1

Introduction

> The epistemic aim of science is not truth, but understanding
>
> Angela Potochnik

More than 150 years have passed since Ernst Haeckel first coined the term ‘ecology’ in 1866. Today, ecology and evolution are incredibly interdisciplinary, borrowing techniques and ideas from diverse fields such as computer science, statistics, economics, dynamical systems, physics, and information theory. With this development has come a cornucopia of models that use these tools and techniques to try and understand biological phenomena. Though many of these models deal with specific ‘low-level’ factors, there is also value to formulating general, idealized models in abstract terms to underscore these organizing principles and better illustrate the fundamental processes that are required to capture the ‘essence’ of a biological pattern (Frank, 2012; Vellend, 2016). Such abstractions and idealizations are useful in unifying apparently disparate modelling frameworks and are thus vital to theory-building (Luque and Baravalle, 2021).

1.1 Idealization and generality in ecology

The push and pull between the search for general patterns and the specification of minutiae has a long and torturous history in ecology (Kingsland, 1985). One of the first mathematical idealizations in ecology came in the form of the logistic equation formulated by Pearl and Verhulst, and shortly thereafter, equations for the populations of interacting predator and prey species, put forth independently by Lotka and Volterra. These models were immediately controversial, and for good reason: Many ecologists felt that they were overly idealistic and neglected many important truths about real biological populations (Kingsland, 1985). These models were nevertheless quite good at predicting the patterns of such populations, and have proved valuable to the field of population ecology. Today, these models are viewed as 'classical' and are regularly used even by hard-line empiricists, not because we believe them to be true in all their gory biological details, but because we recognize that they can be *useful* despite being blatantly false generalizations. This is a single instance of a more general philosophical idea concerning the goals and ideals of science. As suggested by the epigraph at the beginning of the chapter, philosophers of science (Potochnik, 2018) have recently argued that science does not seek truth, but instead seeks understanding. The fact that idealization is part and parcel of science is clear if one looks at the actual practice, 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). Relatively simple models of complex eco-evolutionary processes are therefore desirable as a way to shine light on these phenomena. To reiterate once more, the goal of such simple models is not truth (whatever that is), predictive power (as with models in physics), or detailed description (as with detailed individual-based simulations, very flexible statistical models, or machine learning), but *understanding*. Since the world is complicated and humans are limited, such understanding inevitably comes at the cost of other desirable qualities such as the ability to make precise quantitative predictions. It is important to remember at the outset that the models I speak about in this thesis may seem to be over-idealized and too general, and will make only qualitative predictions. This is *by design*, in pursuit of general insight over precise quantitative prediction.

Vellend has recently argued that conceptual synthesis in community ecology requires "shifting the emphasis away from an organizational structure based on the useful lines of inquiry carved out by researchers, to one based on the fundamental processes that underlie community dynamics and patterns" (Vellend, 2016). Vellend's assertion is based on the fact that population genetics has managed to come up with reasonably comprehensive theory due to its focus on the abstract 'high-level' processes of selection, mutation, drift, and gene flow instead of the myriad 'low-level' processes that may be responsible for generating them. In contrast, he believes that practitioners of community ecology often focus on specific 'low-level' processes such as predation rate, limiting resources (R^*), storage effects, priority effects, senescence, and niche partitioning, leading to a plethora of models (see Table 5.1 in (Vellend, 2016) for an in-exhaustive list of 24 such models) and the conclusion that community ecology 'is a mess'. Vellend proposes organizing ecological models according to the 'high-level' processes of selection, ecological drift (demographic stochasticity), dispersal, and speciation. Of course, no such general organization will 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). However, I believe the act of looking beyond the low-level processes present in biological systems and categorizing theories, models, and concepts in terms of a small number of (blurry-eyed) fundamental high-level processes provides a powerful unifying tool to organize concepts in biology. This is perhaps best exemplified by Darwin's theory of natural selection as first proposed in *The Origin of Species*. Darwin famously painstakingly collected a series of 'low-level' facts and observations regarding breeds, wild populations, and the geographic record to support his hypothesis. However, ultimately, these observations culminated in a synthesis whereby they were all unified under a single, abstract, 'high-level' process, namely evolution by natural selection. A similar focus on high-level processes is also present in the mathematization of evolution by natural selection carried out by evolutionary biologists during and after the Modern Synthesis.

1.2 A brief history of high-level modelling frameworks in population biology

In population genetics, the relevant high-level forces are the standard evolutionary forces of natural selection, genetic drift, dispersal, and mutation, and the description of evolution in terms of these forces was first laid out in formal mathematical terms during the Modern Synthesis (Ewens, 2004). The mathematization of the evolutionary forces proved extremely successful, unifying two major schools of thought - Mendelian genetics and Darwinian evolution - that were, at the time, thought to be incompatible cite something here.
Classical population genetics regarded forces like selection and mutation as fundamental, and the success of this approach during the Modern Synthesis illustrates the value of formulating high-level, abstract models that only provide a 'high-level' description of the fundamental processes required to capture the essence of a biological pattern. However, the drama of

evolution famously occurs in the ecological theatre (Hutchinson, 1965), and quantities like fitness are not truly fundamental, but instead emerge as the result of various ecological interactions, tradeoffs, and constraints, a fact that can have important consequences for evolution (Coulson et al., 2006; Kokko et al., 2017). Trying to model such 'eco-evolutionary dynamics' has sprouted a rich body of theoretical literature and led to the development of theoretical frameworks like evolutionary game theory and adaptive dynamics, fields which have greatly enriched our understanding of biological populations (Brown, 2016). Eco-evolutionary population dynamics can broadly be organized under a single unifying framework, the Price equation, that yields all the relevant formal structures under various limits (Page and Nowak, 2002; Lion, 2018). The Price equation partitions changes in population composition into multiple terms, each of which lends itself to a straightforward interpretation in terms of the high-level evolutionary forces of selection, mutation, and drift, thus providing a useful conceptual framework for thinking about how populations change over time (Frank, 2012). The Price equation also leads to a small number of simple yet insightful 'fundamental theorems' of population biology (Queller, 2017; Lion, 2018; Lehtonen, 2018) and unifies several various seemingly disjoint formal structures under a single theoretical banner (Lehtonen, 2020; Luque and Baravalle, 2021)

The general guiding philosophy of much of this mathematization has been 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 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) in both ecological (Schreiber et al., 2022) and evolutionary (DeLong and Cressler, 2023) models. Stochastic models often exhibit phenomena that do not occur in infinite-population models (Rogers et al., 2012a; Rogers et al., 2012b; Veller et al., 2017), prevent phenomena that occur in infinite populations from occurring (Proulx and Day, 2005; Johansson and Ripa, 2006; Claessen et al., 2007; Wakano and Iwasa, 2013; Débarre and Otto, 2016; Johnson et al., 2021), or even completely reverse the predictions of deterministic models (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Constable et al., 2016; McLeod and Day, 2019). 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. However, the classic stochastic models in population and quantitative genetics typically assume a fixed total population size (Fisher, 1930; Crow and Kimura, 1970; Lande, 1976) and their validity is therefore rather restrictive since population sizes routinely fluctuate in the real world. Importantly, the Price equation itself is usually formulated in a deterministic, dynamically-insufficient manner (but see Rice, 2008 for a stochastic formulation of the Price equation). Since real-life populations are stochastic, finite, and of non-constant population size, it is thus imperative that we develop a theoretical framework that can handle such systems directly, instead of only working with deterministic, infinite-population approximations.

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 lead to nonsensical predictions and inconsistencies (Strang et al., 2019). 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 much more natural (Black and McKane, 2012), lead to sensible predictions (Strang et al., 2019), and can fundamentally differ from the predictions made by simply adding noise terms to a deterministic model (Strang et al., 2019) (This is also true when incorporating spatial structure into non-spatial models, see for example Durrett and Levin, 1994). 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): 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 (Doebeli et al., 2017) and provide a more fundamental mechanistic description of the relevant evolutionary forces. In other words, 'all paths to fitness lead through demography' (Metcalf and Pavard, 2007b).

1.3 Outline of the rest of this thesis

In this thesis, I present a formulation of population dynamics constructed from mechanistic first principles grounded in individual-level birth and death. Part II presents the mathematical formalism in a detailed, self-contained, pedagogical manner. 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. I present a brief introduction to the relevant mathematics in section 2.1 and present a toy example of tracking population size of a population of identical individuals in section 2.2. 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. Finally, I conduct a weak noise approximation to arrive at a linear Fokker-Planck equation which can be solved exactly using some stochastic calculus to arrive at a closed-form solution given by a time-dependent Ornstein-Uhlenbeck process, thus illustrating all the major tools required. In section 2.3, I present a multivariate process to describe the evolution of discretely varying traits, and, as before, use the system size expansion to arrive at a continuous description of change in trait frequencies as an SDE. Under mild assumptions, I show that the deterministic limit of this process is the well-known replicator-mutator equation (or equivalently, Eigen's quasispecies equation), thus establishing the microscopic basis of well-known equations from stochastic first principles. While the mathematics of section 2.3 is standard and well-understood, it has, to the best of my knowledge, not been used in this context in the generality we use here, though it has been used in several specific models of specific systems. As such, chapter 2.3 can be thought of as a conceptual synthesis and unification. Chapter 3 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 a system-size approximation to arrive at a 'functional' Fokker-Planck in terms of functional derivatives. Under mild assumptions, I show that classic equations such as Kimura's infinite alleles model and the canonical equation of adaptive dynamics can be derived as the deterministic limits of this stochastic process. I also conduct a weak noise approximation to arrive at a linear functional Fokker-Planck equation. Chapter 3 generalizes the work of Tim Rogers and colleagues (Rogers et al., 2012a; Rogers et al., 2012b; Rogers and McKane, 2015), and to the best of my knowledge, is entirely original both mathematically and biologically. Mathematically, the ideas present heuristic, accessible alternatives in terms of a 'stochastic field theory' 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). This formulation in terms of stochastic field equations also opens up the model to analysis using the powerful heuristic tools of physics such as the path integral (Doi, 1976; Peliti, 1985; Dodd and Ferguson, 2009; Chow and Buice, 2015; Weber and Frey, 2017), though I do not attempt to leverage these connections here, leaving it as an opportunity for future study. Examples for one-dimensional, multi-dimensional, and quantitative trait models that can be analyzed in this formalism are presented in Appendix D.

Part III summarizes the major results of this formalism and presents some simple equations that can be argued to be 'fundamental theorems' of population biology in the sense of Queller, 2017. These theorems 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 theorems predict a new evolutionary force, 'noise-induced selection', that has still not found its way into the standard formal canon of evolutionary biology such as the Price equation, 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. If one does not care for mathematical details (not advised!), they can skip part II entirely and directly read part III for the major takeaways.

Part II

Theory

Chapter 2

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

2.1 Mathematical Background

Here, I provide a brief, informal introduction to basic notions in stochastic processes that will be helpful for some technical portions of this thesis. Interested readers looking for a more comprehensive source can refer to standard mathematics texts such as Øksendal, 1998 or Karatzas and Shreve, 1998 for a more 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 so-called 'continuous-time Markov chain' in which only transitions between local states are allowed. In other words, 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 $u < s \leq t$, then
$$\mathbb{P}(X_t|X_s, X_u) = \mathbb{P}(X_t|X_s)$$
- Direct transitions must be 'local'. Mathematicians usually reserve the phrase 'birth-death process' to processes that take values in the non-negative integers $\{0, 1, 2, 3, 4, \ldots\}$. In this case, only direct transitions from n to $n \pm 1$ are allowed to occur. 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 by using tools from stochastic calculus and partial differential equations, as we describe below.

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

$$X_t = \int_0^t F(s, X_s)ds + \int_0^t G(s, X_s)dB_s \tag{2.1}$$

where $F(t,x)$ and $G(t,x)$ are 'nice' functions[1] In the physics literature, F and G are often called the 'drift' and 'diffusion' of the process respectively. However, we will not use this terminology here due to potential confusion with genetic drift (which actually corresponds to the 'diffusion' in the physics terminology). B_t is the so-called 'standard Brownian motion'. Named after the botanist Robert Brown (who was looking at the random erratic motion of pollen grains in water under a microscope), $\{B_t\}_{t\geq 0}$ is a stochastic process that is supposed to model 'random noise' or 'undirected diffusion' of a particle. If one imagines B_t as recording the position of a small pollen grain at time t, then B_t can be formally thought of as a process that has the following properties:

- It starts at the origin, *i.e* $B_0 = 0$. This is a harmless assumption made for convenience and amounts to a choice of coordinate system.

- It moves continuously, without sudden jumps across regions of space, *i.e* the map $t \to B_t$ is continuous. This simply says that our pollen grain moves short distances in short intervals of time.

- The future movement is independent of past history. That is, given times $u < s < t$, the displacement $B_t - B_s$ is independent of the past position B_u.

- The movement is directionless and random, and displacement is normally distributed. More precisely, given two times $s < t$, the displacement $B_t - B_s$ follows a normal distribution with a mean of 0 (this is the 'directionless' part) and a variance of $t - s$ (this is the 'random' part).

[1] For the mathematically oriented reader, there are two requirements: Firstly, we require the functions to have 'linear growth', meaning that we can find a constant $C > 0$ such that $\|F(t,x)\| + \|G(t,x)\| \leq C(1 + \|x\|)$ for every $x \in \mathbb{R}^d$ and $t > 0$. We also require 'Lipschitz continuity', which means that we can find a constant $L > 0$ such that $\|F(t,x) - F(t,y)\| + \|G(t,x) - G(t,y)\| \leq L\|x - y\|$ for every pair $x, y \in \mathbb{R}^d$ and $t > 0$. Here, $\|\cdot\|$ denotes the natural norm on the space under consideration and for our cases will usually be the Euclidean norm. For biological systems, both of these conditions will usually be satisfied, and so we assume going further that all our SDEs are always well-defined and admit solutions.

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* $\mathbb{E}[B_t|B_0] = B_0 = 0$.
The second integral in equation (2.1) is Itô's 'stochastic integral', and is to be interpreted in the following sense: Fix a time $T > 0$. Partition the interval $[0, T]$ into n intervals of the form $[t_i, t_{i+1}]$ such that $0 = t_0 < t_1 < t_2 < \ldots < t_n = T$. Then, the (Itô) stochastic integral from 0 to T can be thought of as:

$$\int_0^T G(s, X_s) dB_s := \lim_{n\to\infty} \sum_{i=1}^{n} G(t_i, X_{t_i})(B_{t_{i+1}} - B_{t_i})$$

That is to say, it is obtained by making successively finer partitions of the form $[t_i, t_{i+1}]$, and then computing the 'area of the rectangle' formed with $B_{t_{i+1}} - B_{t_i}$ and $G(t_i, X_{t_i})$ as sides. 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 corresponding to the random displacement of a Brownian particle during the uniform time interval $[t_i, t_{i+1}]$.

Equation (2.1) is often represented in the compact form:

$$dX_t = F(t, X_t)dt + G(t, X_t)dB_t \tag{2.2}$$

. The physics literature also often uses the 'Langevin form':

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

where $\eta(t)$ is supposed to be 'Gaussian white noise', defined indirectly such that $\int_0^t G(s, x)\eta(s)ds$ behaves identically to $\int_0^t G(s, X_s)dB_s$. However, it is important to remember that these are both purely formal 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), which formally 'makes sense'. 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 ($i.e$ $\mathbb{E}[B_t|B_0] = B_0 = 0$), it can be shown[2] that the stochastic integral also has an expectation value of 0 for all t, $i.e$:

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

Using this, and the fact that the future path of the Brownian motion itself is independent of its history, one can derive the following 'notational algebra table' for manipulating products of formal expressions of the form (2.2):

	dt	$\mathbf{dB_t}$
dt	0	0
$\mathbf{dB_t}$	0	dt

which becomes very useful for formal manipulation. One important consequence 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 quantity $x(t)$ satisfying $\dot{x} = f(x,t) + g(x,t)$ for two 'nice'[3] real functions f and g, then, given any function $h(x)$, we have 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 (2.2) with gdt simply being replaced by GdB_t on the RHS. However, this does not

[2] We can actually prove something stronger: We can show under rather mild regularity assumptions on X_t and $G(t,x)$ that the stochastic integral is a continuous square-integrable martingale starting at the origin - This means that the map $t \to \int_0^t G(s, X_s)dB_s$ is continuous, starts at the origin, and always has an expectation value of 0.

[3] We just need the solution $x(t)$ to be a continuous function

work. The correct relation is instead given by *Itô's formula*[4]:

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

Note that there is now an extra $h''(X_t)G^2(X_t)/2$ term that does not exist in the deterministic setting. In some sense, this term is present because the random fluctuations of Brownian motion are 'too erratic' and do not follow our deterministic intuitions. 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

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

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) and pure mathematics (Øksendal, 1998) literature. If the function G is independent of x, then 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). 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. 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.

[4]Itô'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

2.1.3 Density-dependence and the intuition for system-size expansions in 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 in ecology, at an individual level, births and deaths of individuals are affected by local population density and not directly by 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 $\{0, 1, 2, \ldots, n-1, n, n+1, \ldots\}$ to the approximately continuous values $\{0, 1/K, 2/K, \ldots, x - 1/K, x, x + 1/K, \ldots\}$ by the introduction of a 'system size parameter' K. In ecology, this parameter will be some fundamental limit on resources, such as habitat size or carrying capacity. In physics and chemistry, it is usually the total volume of a container in which physical or chemical reactions take place. 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 terms. A similar approximation is well-known (ever since Fisher) in theoretical population genetics (Ewens, 2004), where it goes by the name of the 'diffusion approximation', 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 (*i.e.* is closer to a Kramers-Moyal expansion than a Van Kampen expansion).

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 trajectory (obtained by letting $K \to \infty$) 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.

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).

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. Since all individuals are identical, we can only really track the population size through time. 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', we 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 rate $b(n)$ of a population with population size n is the quantity

$$b(n) := \lim_{\epsilon \to 0} \frac{1}{\epsilon} \mathbb{P}\left(N_{t+\epsilon} = n+1 | N_t = n\right) \tag{2.5}$$

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

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

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* has just occurred, then, the probability that the event that occurred is a birth is

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

and the probability that the event was instead a death is given by

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

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 $d(n) = \left(\mu + (\lambda - \mu)\frac{n}{K}\right) n$, where μ and K are positive constants. Taking the difference between the birth and death rates, we obtain $b(n) - d(n) = (\lambda - \mu)n\left(1 - \frac{n}{K}\right)$, where, identifying $r = \lambda - \mu$, we obtain the familiar 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

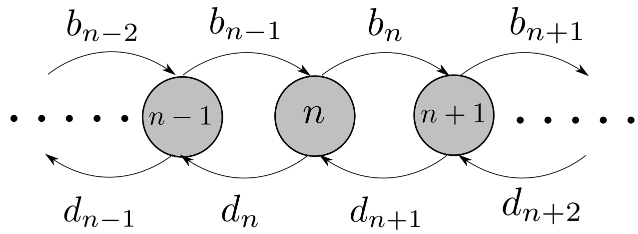

Figure 2.1: Schematic description of a one-dimensional birth-death process. Consider a population of identical individuals. The state of the system can be described by a single number, in this case, the population size. Births and deaths result in changes in the total population size, and the birth and death rates are dependent on the current population size.

probability. We can now simply measure the 'inflow' and 'outflow' of copies of the population from each state. 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

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

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

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

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

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

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

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

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

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

where we now have the new step operators

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

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

$$\boxed{\frac{\partial P}{\partial t}(x,t) = -\frac{\partial}{\partial x}\{A^-(x)P(x,t)\} + \frac{1}{2K}\frac{\partial^2}{\partial x^2}\{A^+(x)P(x,t)\}} \tag{2.13}$$

where

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

Equation (2.13) has the form of a so-called 'Fokker-Planck equation', and corresponds to the SDE:

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

interpreted in the Itô sense. 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 and scales inversely with $\sqrt{K}$.

2.2.3 Stochastic fluctuations and the weak noise approximation

If we assume the noise is *weak*, 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 $K \to \infty$, equation (2.14) describes a deterministic process, obtained as the solution to

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

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, we 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, we have introduced the variables:

$$y = \sqrt{K}\,(x - \alpha(t))$$
$$s = t$$

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

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

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

and

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

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

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

We are now ready to make a weak noise 'expansion'. 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:

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

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 rth term of the expansion satisfies:

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

If we assume we can obtain a reasonable approximation by retaining only the first term of the expansion and neglecting all higher-order terms[5], we are left with the expression:

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

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

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

[5]For example, if the deterministic trajectory is at a stable fixed point and subject to weak fluctuations

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

$$\exp\left(-\int_0^t A_1^-(s)ds\right)dY_t - Y_t A_1^-(t)\exp\left(-\int_0^t A_1^-(s)ds\right)dt = \sqrt{A_0^+(t)}\exp\left(-\int_0^t A_1^-(s)ds\right)dB_t$$

$$\Rightarrow d\left(\exp\left(-\int_0^t A_1^-(s)ds\right)Y_t\right) = \sqrt{A_0^+(t)}\exp\left(-\int_0^t A_1^-(s)ds\right)dB_t$$

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

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

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, and one can get many insights from it. 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 we will not demonstrate this here.

Importantly, higher order terms do not form FPEs, 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 method is not very useful.

2.3 Multi-dimensional processes for discrete traits

Let us now consider a slightly more complicated scenario. 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 formalism we have developed in the previous section carries out essentially unchanged in this case.

2.3.1 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 2.2A). Thus, the state of the population at a given time t is an m-dimensional *vector* of the form $\mathbf{v} = [v_1(t), v_2(t), \ldots, v_m(t)]$, where $v_i(t)$ is the number of individuals of type i.

Given a state $\mathbf{v}(t)$, 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 2.2B). In other words, for each type $i \in \{1, 2, \ldots, m\}$, we must specify a birth rate $b_i(\mathbf{v})$ and a death rate $d_i(\mathbf{v})$. By 'rates', we 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

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

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

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

As before, we can describe the rate of change of $P(\mathbf{v}, t)$, the probability of finding the population in a state $\mathbf{v}$ at time t, by measuring the inflow and outflow rates. Given a population $\mathbf{v} = [v_1, \ldots, v_m]$, the 'inflow' is from all populations of the form $[v_1, \ldots, v_i - 1, \ldots, v_m]$ through a birth of a type i individual, and from all populations of the form $[v_1, \ldots, v_i + 1, \ldots, v_m]$ through the death of a type i individual. Thus, we have the inflow

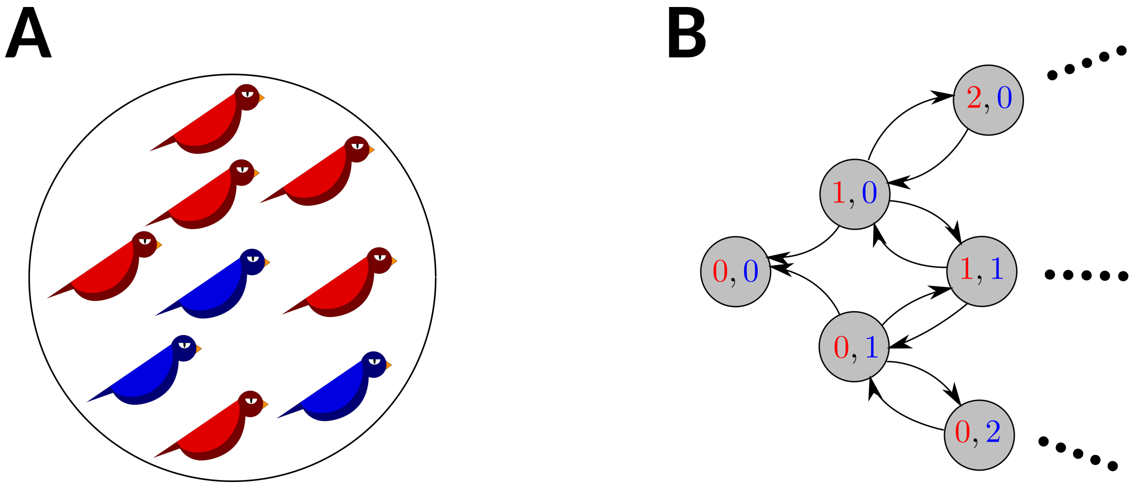

Figure 2.2: Schematic description of a one-dimensional birth-death process. **(A)** Consider a population of birds in which individuals are either red or blue. In this case, we have $m = 2$, since there are two types of individuals in the population. **(B)** The state of the system can be described by 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.

rate

$$\begin{aligned} R_{\text{in}}(\mathbf{v},t) = &\sum_{j=1}^{m} b_j([v_1,\ldots,v_j-1,\ldots,v_m])P([v_1,\ldots,v_j-1,\ldots,v_m],t) \\ &+\sum_{j=1}^{m} d_j([v_1,\ldots,v_j+1,\ldots,v_m])P([v_1,\ldots,v_j+1,\ldots,v_m],t) \end{aligned} \tag{2.23}$$

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

$$R_{\text{out}}(\mathbf{v},t) = \sum_{j=1}^{m} b_j(\mathbf{v})P(\mathbf{v},t) + \sum_{j=1}^{m} d_j(\mathbf{v})P(\mathbf{v},t) \tag{2.24}$$

As before, we now define step operators, both for notational ease and in anticipation of the system size expansion. Note that now, we need $2m$ step operators. For each $i \in \{1,\ldots,m\}$, let us define two step operators $\mathcal{E}_i^{\pm}$ by their action on any function $f([v_1,\ldots,v_m],t)$ as:

$$\mathcal{E}_i^{\pm} f([v_1,\ldots,v_m],t) = f([v_1,\ldots,v_i \pm 1,\ldots v_m],t) \tag{2.25}$$

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 $P(\mathbf{v},t)$ as

$$\frac{\partial P}{\partial t}(\mathbf{v},t) = R_{\text{in}}(\mathbf{v},t) - R_{\text{out}}(\mathbf{v},t) \tag{2.26}$$

Substituting (2.23), (2.24), and (2.25), we obtain:

$$\frac{\partial P}{\partial t}(\mathbf{v},t) = \sum_{j=1}^{m} \left[(\mathcal{E}_j^- - 1) b_j(\mathbf{v}) P(\mathbf{v},t) + (\mathcal{E}_j^+ - 1) d_j(\mathbf{v}) P(\mathbf{v},t) \right] \tag{2.27}$$

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

2.3.2 The system-size expansion

As before, we now assume we can find a system size parameter $K > 0$ such that we can make the substitutions

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

and define new step operators $\Delta_i^{\pm}$ by their action on any real-valued function $f(\mathbf{x},t)$ as

$$\Delta_i^{\pm} f([x_1, \dots, x_m], t) = f([x_1, \dots, x_i \pm \frac{1}{K}, \dots x_m], t) \tag{2.28}$$

In terms of these new variables, (2.27) becomes

$$\frac{\partial P}{\partial t}(\mathbf{x},t) = K \sum_{j=1}^{m} \left[(\Delta_j^- - 1) b_j^{(K)}(\mathbf{x}) P(\mathbf{x},t) + (\Delta_j^+ - 1) d_j^{(K)}(\mathbf{x}) P(\mathbf{x},t) \right] \tag{2.29}$$

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

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

which, after substituting into (2.29), yields the equation

$$\boxed{\frac{\partial P}{\partial t}(\mathbf{x},t) = \sum_{j=1}^{m}\left[-\frac{\partial}{\partial x_j}\{A_j^-(\mathbf{x})P(\mathbf{x},t)\} + \frac{1}{2K}\frac{\partial^2}{\partial x_j^2}\{A_j^+(\mathbf{x})P(\mathbf{x},t)\}\right]} \tag{2.30}$$

where

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

Equation (2.30) is an m-dimensional Fokker-Planck equation, and corresponds to the m-dimensional Itô process

$$d\mathbf{X}_t = \mathbf{A}^-(\mathbf{X}_t)dt + \frac{1}{\sqrt{K}}\mathbf{D}(\mathbf{X}_t)d\mathbf{B}_t \tag{2.31}$$

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

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

Finally, $\mathbf{B}_t$ is the m-dimensional Brownian motion and can be thought of as a vector of independent one-dimensional Brownian motions (which have been defined in 2.1.2). This is the 'mesoscopic' description of our process.

2.3.3 Stochastic Trait Frequency Dynamics

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

$$\begin{aligned} b_i^{(K)}(\mathbf{x}) &= x_i b_i^{(\text{ind})}(\mathbf{x}) + \mu Q_i(\mathbf{x}) \\ d_i^{(K)}(\mathbf{x}) &= x_i d_i^{(\text{ind})}(\mathbf{x}) \end{aligned} \tag{2.32}$$

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 rate[6]. Our assumptions of the functional forms (2.32) 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. We define the *Malthusian fitness* of the i^{th} type as $w_i(\mathbf{x}) \coloneqq b_i^{(\text{ind})}(\mathbf{x}) - d_i^{(\text{ind})}(\mathbf{x})$, and the *per-capita turnover rate* of the i^{th} type as $\tau_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) + b_i^{(\text{ind})}(\mathbf{x})$. 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 we will stick to w_i and 'fitness' here. 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, both the fitness and the turnover rate in our model are frequency-dependent.

Following ideas similar in spirit to (McLeod and Day, 2019), we can now use (2.31) to write down stochastic equations for $p_i(t)$, the frequency of type i individuals in the population. More precisely, given a state $\mathbf{x}(t)$, we can compute the total (scaled) population size and the frequency of each type in the population as:

$$\begin{aligned} N_K(t) &\coloneqq \sum_{i=1}^{m} x_i(t) = \frac{1}{K}\sum_{i=1}^{m} v_i(t) \\ p_i(t) &\coloneqq \frac{x_i(t)}{N_K(t)} \end{aligned} \tag{2.33}$$

We can also calculate the statistical mean value of any type level quantity f in the population as

$$\overline{f}(t) \coloneqq \sum_{i=1}^{m} f_i p_i(\mathbf{x}(t)) \tag{2.34}$$

, where f_i is the value of the quantity for the ith type. In appendix B, we show that we can use Itô's formula to write down a general stochastic equation for the frequencies of each type in the population. Unlike (McLeod and Day, 2019), we make no assumptions about the separation of ecological and evolutionary time scales or the strength of selection and are able to present an entirely general calculation. Letting $\overline{w} = \sum w_i p_i$ and $\overline{\tau} = \sum \tau_i p_i$ be

[6] When $x_i = 0$, *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.

the average population fitness and turnover respectively, we show in appendix B that the frequency of the i^{th} type in the population $\mathbf{x}(t)$ changes according to the equation:

$$\begin{aligned} dp_i(t) = &\left[(w_i(\mathbf{x}) - \overline{w})p_i + \mu\left\{Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right\}\right] dt \\ &- \frac{1}{K}\frac{1}{N_K(t)}\left[(\tau_i(\mathbf{x}) - \overline{\tau})p_i + \mu\left\{Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right\}\right] dt \\ &+ \frac{1}{\sqrt{K}}\frac{1}{N_K(t)}\left[\left(A_i^+\right)^{1/2} dB_t^{(i)} - p_i \sum_{j=1}^{m}\left(A_j^+\right)^{1/2} dB_t^{(j)}\right] \end{aligned} \tag{2.35}$$

where $B_t^{(1)}, B_t^{(2)}, \ldots, B_t^{(m)}$ are m independent one-dimensional standard Brownian motion processes and we have used the notation $Q_i(\mathbf{p}) = Q_i(\mathbf{x})/N_K(t)$ for notational clarity. We 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 Chapter 4. Finally, the last term of equation (2.35) describes non-directional stochastic effects due to fluctuations and has a 'spreading effect'.

2.3.4 The infinite population limit

Like in 2.2, we can once again take $K \to \infty$ in (2.31) to obtain a deterministic expression. Here, the expression reads

$$\frac{d\mathbf{x}}{dt} = \mathbf{A}^-(\mathbf{x}) = \mathbf{b}^{(K)}(\mathbf{x}) - \mathbf{d}^{(K)}(\mathbf{x}) \tag{2.36}$$

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

$K \to \infty$ in (2.35), we obtain a deterministic equation that reads:

$$\boxed{\frac{dp_i}{dt} = (w_i(\mathbf{x}) - \overline{w})p_i + \mu\left[Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right]} \tag{2.37}$$

In physics parlance, equations (2.36) and (2.37) constitute the 'macroscopic' description of our stochastic process. The first term of (2.37) describes changes due to faithful (non-mutational) replication, and the second describes changes due to mutation. For this reason, equation (2.37) is called the *replicator-mutator equation* in the evolutionary game theory literature, where the individual 'types' are interpreted to be pure strategies. 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 get the replicator-mutator equation for matrix games, and the constants a_{ij} form the 'payoff matrix'. 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 (2.37) 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 function[7]. 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 $f_i =$ *value of* i^{th} *phenotype* gives us the mean trait value in the population). The product rule of calculus tells us that we have the relation

$$\frac{d}{dt}\left(\sum_{i=1}^{m} f_i p_i\right) = \sum_{i=1}^{m}\left(f_i\frac{\partial p_i}{\partial t} + p_i\frac{\partial f_i}{\partial t}\right) = \sum_{i=1}^{m} f_i\frac{\partial p_i}{\partial t} + \overline{\left(\frac{\partial f}{\partial t}\right)} \tag{2.38}$$

Multiplying both sides of equation (2.37) by f_i and summing over all i, we obtain

$$\sum_{i=1}^{m} f_i\frac{\partial p_i}{\partial t} = \sum_{i=1}^{m} f_i w_i(\mathbf{x})p_i - \overline{w}\sum_{i=1}^{m} f_i p_i + \mu\left[\sum_{i=1}^{m} Q_i(\mathbf{p})f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p})\sum_{i=1}^{m} p_i f_i\right)\right]$$

[7]Mutational 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 $j \to i$ mutation (conditioned on the occurrence of a mutation). Substituting this into (2.37) yields an equation in terms of 'Q-matrices' or 'mutation matrices' that may be more familiar to some.

$$\Rightarrow \frac{d\overline{f}}{dt} = \overline{wf} - (\overline{w})(\overline{f}) + \mu \left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \overline{f} \right]$$

Using the definition of statistical covariance of two variables X and Y as $\text{Cov}(X, Y) = \overline{XY} - (\overline{X})(\overline{Y})$, we obtain

$$\sum_{i=1}^{m} f_i \frac{\partial p_i}{\partial t} = \text{Cov}(w, f) + \mu \left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \overline{f} \right] \tag{2.39}$$

Thus, substituting this into (2.38), we get

$$\boxed{\frac{d\overline{f}}{dt} = \text{Cov}(w, f) + \mu \left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \overline{f} \right] + \overline{\left(\frac{\partial f}{\partial t} \right)}} \tag{2.40}$$

This is a Price equation for quantities f_i which can vary over time. 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 $\overline{f}$ 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

$$\boxed{\frac{d\overline{f}}{dt} = \text{Cov}(w, f) + \mu \left[\sum_{i=1}^{m} Q_i(\mathbf{p}) f_i - \left(\sum_{j=1}^{m} Q_j(\mathbf{p}) \right) \overline{f} \right]} \tag{2.41}$$

, the famous Price equation in continuous time. 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'.

2.3.5 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 (2.36) be given by $\boldsymbol{\alpha}(t)$. We can once again track stochastic fluctuations from the deterministic trajectory by introducing the new variables

$$\begin{aligned}\mathbf{y} &= \sqrt{K}(\mathbf{x}-\boldsymbol{\alpha}(t))\\ s &= t\\ \tilde{P}(\mathbf{y},s) &= \frac{1}{\sqrt{K}}P(\mathbf{x},t)\end{aligned} \tag{2.42}$$

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:

$$\frac{\partial \tilde{P}_0}{\partial s}(\mathbf{y},s) = \sum_{j=1}^{m}\left(-\frac{\partial}{\partial y_j}\left\{(A_j^-)_1(s)\tilde{P}_0(\mathbf{y},s)\right\} + \frac{1}{2}{A_j}^+(\boldsymbol{\alpha}(s))\frac{\partial^2}{\partial {y_j}^2}\{\tilde{P}_0(\mathbf{y},s)\}\right) \tag{2.43}$$

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

$$A_j^-(\boldsymbol{\alpha}+\frac{\mathbf{y}}{\sqrt{K}}) = \sum_{n=1}^{\infty}(A_j^-)_n(s)\left(\frac{\mathbf{y}}{\sqrt{K}}\right)^n$$

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

$$(A_j^-)_1(s) = \sum_{i=1}^{m} y_i\left(\left.\frac{\partial A_j^-(\mathbf{x})}{\partial x_i}\right|_{\mathbf{x}=\boldsymbol{\alpha}(s)}\right) \tag{2.44}$$

In multi-variable calculus, the directional derivative[8] $D_{\mathbf{v}}(f(\mathbf{x}))$ of a multidimensional function $f:\mathbb{R}^n\to\mathbb{R}$ along a vector $\mathbf{v}$ is the function defined by:

$$D_{\mathbf{v}}(f(\mathbf{x})) \coloneqq \sum_{i=1}^{n}\left(\frac{\partial f(\mathbf{x})}{\partial x_i}\right)v_i = \lim_{h\to 0}\frac{f(\mathbf{x}+h\mathbf{v})-f(\mathbf{x})}{h} \tag{2.45}$$

Comparing with (2.44), we see that the weak-noise approximation of our process is:

$$\frac{\partial P}{\partial t}(\mathbf{y},t) = \sum_{j=1}^{m}\left(-\frac{\partial}{\partial y_j}\left\{D_{\mathbf{y}}(A_j^-(\boldsymbol{\alpha}))(t)P(\mathbf{y},t)\right\} + \frac{1}{2}{A_j}^+(\boldsymbol{\alpha}(t))\frac{\partial^2}{\partial {y_j}^2}\{P(\mathbf{y},t)\}\right) \tag{2.46}$$

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 $\boldsymbol{\alpha}(s)$ here is the multidimensional analogue of the derivative we had in (2.21). The meaning of equation (2.46) is clearer if we compute how

[8] Physicists sometimes use the notation $\partial_{\mathbf{v}} f(\mathbf{x})$ or $\mathbf{v}\cdot\nabla f(\mathbf{x})$ for this object.

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:

$$\frac{d}{dt}\mathbb{E}[y_i^n] = \frac{d}{dt}\int\limits_{\mathbb{R}^m} y_i^n P(\mathbf{y},t)d\mathbf{y} \tag{2.47}$$

$$= \int\limits_{\mathbb{R}^m} y_i^n \frac{\partial P}{\partial t}(\mathbf{y},t)d\mathbf{y} \tag{2.48}$$

where we 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 and used the shorthand $\int\limits_{\mathbb{R}^m} f(\mathbf{y})\ d\mathbf{y} = \int\limits_{\mathbb{R}}\int\limits_{\mathbb{R}}\cdots\int\limits_{\mathbb{R}} f(\mathbf{y})\ dy_1 dy_2 \dots dy_m$. 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 (2.46) into (2.48) to obtain

$$\frac{d}{dt}\mathbb{E}[y_i^n] = \int\limits_{\mathbb{R}^m} y_i^n \left(\sum_{j=1}^{m}\left(-\frac{\partial}{\partial y_j}\left\{D_j P(\mathbf{y},t)\right\} + \frac{1}{2}{A_j}^+(\boldsymbol{\alpha}(t))\frac{\partial^2}{\partial {y_j}^2}\{P(\mathbf{y},t)\}\right)\right)d\mathbf{y} \tag{2.49}$$

$$= \sum_{j=1}^{m}\left[-\int\limits_{\mathbb{R}^m} y_i^n \frac{\partial}{\partial y_j}\left\{D_j P(\mathbf{y},t)\right\}d\mathbf{y} + \frac{{A_j}^+(\boldsymbol{\alpha}(t))}{2}\int\limits_{\mathbb{R}^m} y_i^n \frac{\partial^2}{\partial {y_j}^2}\{P(\mathbf{y},t)\}d\mathbf{y}\right] \tag{2.50}$$

We will evaluate the integrals on the RHS of (2.50) 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:

$$\int\limits_{\Omega} \frac{\partial u}{\partial x_i} v d\mathbf{x} = -\int\limits_{\Omega} u\frac{\partial v}{\partial x_i}d\mathbf{x} + \int\limits_{\partial\Omega} uv\gamma_i dS(\mathbf{x}) \tag{2.51}$$

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 $\|y\| \to \infty$. We assume that the magnitude of stochastic fluctuations is bounded, and therefore impose the condition $\lim\limits_{\|y\|\to\infty} P(\mathbf{y},t) = 0$. Further, we assume that this decay is fast enough that $\lim\limits_{\|y\|\to\infty} D_j P(\mathbf{y},t) = 0\ \forall\ j$. Under these conditions, we can evaluate the two integrals in the RHS of (2.50) by using integration by parts and discarding the boundary term (The second term on the RHS of (2.51)). Note that since the y_is 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 (2.50) (once for the first term and twice for the second term), we obtain the considerably simpler expression

$$\frac{d}{dt}\mathbb{E}[y_i^n] = n \int_{\mathbb{R}^m} y_i^{n-1} D_i P(\mathbf{y}, t) d\mathbf{y} + \frac{n(n-1)}{2} A_i^+(\boldsymbol{\alpha}(t)) \int_{\mathbb{R}^m} y_i^{n-2} P(\mathbf{y}, t) d\mathbf{y} \tag{2.52}$$

$$\Rightarrow \frac{d}{dt}\mathbb{E}[y_i^n] = n\mathbb{E}[y_i^{n-1} D_i] + \frac{n(n-1)}{2} A_i^+(\boldsymbol{\alpha}(t)) \mathbb{E}[y_i^{n-2}] \tag{2.53}$$

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:

$$\frac{d}{dt}\mathbb{E}[y_i] = \mathbb{E}[D_i] \tag{2.54}$$

$$\frac{d}{dt}\mathbb{E}[y_i^2] = 2\mathbb{E}[y_i D_i] + A_i^+(\boldsymbol{\alpha}(t)) = 2\mathrm{Cov}(y_i, D_i) + 2\mathbb{E}[y_i]\mathbb{E}[D_i] + A_i^+(\boldsymbol{\alpha}(t)) \tag{2.55}$$

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 covariance between the fluctuation and D_i. Note that unlike in the Price equation (2.41), this is a true *probability* covariance (as opposed to a statistical covariance between two deterministic quantities). In the case of the functional forms given by (2.32), we have:

$$A_i^-(\mathbf{x}) = w_i(\mathbf{x}) x_i + \mu Q_i(\mathbf{x}) \tag{2.56}$$

and thus, from (2.44), we can calculate the directional derivative D_i as

$$D_i = \sum_{k=1}^{m} y_k \left(\left. \frac{\partial A_i^-(\mathbf{x})}{\partial x_k} \right|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) \tag{2.57}$$

$$= \sum_{k=1}^{m} y_k \left(\left. \frac{\partial}{\partial x_k} \left(w_i(\mathbf{x}) x_i + \mu Q_i(\mathbf{x}) \right) \right|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) \tag{2.58}$$

$$= \sum_{k=1}^{m} y_k \left(a_i \frac{\partial w_i}{\partial x_k}\bigg|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) + y_i w_i(\boldsymbol{\alpha}) + \mu \sum_{k=1}^{m} y_k \left(\frac{\partial Q_i}{\partial x_k}(\mathbf{x})\bigg|_{\mathbf{x}=\boldsymbol{\alpha}(t)} \right) \tag{2.59}$$

$$= y_i w_i(\boldsymbol{\alpha}) + a_i D_{\mathbf{y}}(w_i(\boldsymbol{\alpha})) + \mu D_{\mathbf{y}}(Q_i(\boldsymbol{\alpha})) \tag{2.60}$$

Using this in (2.54), we see that the expected change of a fluctuation in the density of type i individuals evolves as:

$$\frac{d}{dt}\mathbb{E}[y_i] = \underbrace{w_i(\boldsymbol{\alpha})\mathbb{E}[y_i]}_{\substack{\text{Current fitness of type } i\\ \text{at deterministic trajectory } \boldsymbol{\alpha}\\ \text{(scaled by expected density } \mathbb{E}[y_i])}} + \underbrace{a_i\mathbb{E}[D_{\mathbf{y}}(w_i(\boldsymbol{\alpha}))]}_{\substack{\text{Expected change in fitness}\\ \text{of type } i \text{ in going from } \boldsymbol{\alpha} \text{ to } \mathbf{y}\\ \text{(scaled by deterministic density } a_i)}} + \underbrace{\mu\mathbb{E}[D_{\mathbf{y}}(Q_i(\boldsymbol{\alpha}))]}_{\substack{\text{Expected effect of}\\ \text{mutations}}} \tag{2.61}$$

Chapter 3

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 (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 space[1]. 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 we 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 the (rather pedantic) questions of rigor, 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 $f(x,t) : \mathbb{R}^n \times [0, \infty) \to \mathbb{R}^m$, 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 we will formulate $\{\phi^{(t)}\}_{t \geq 0}$ when viewed as a sequence of functions $\{\phi^{(t)}(y)\}_{t \geq 0}$ thus describes a (scalar) 'stochastic field', and the formalism we 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

[1]The 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. All the Dirac deltas that will turn up shortly are 'properly' viewed as measures, and integrals with Dirac deltas in the integrand are to be interpreted as integration with respect to the Dirac measure. 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) 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.

(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

$$\int \frac{\delta F}{\delta\phi(x)}\rho(x)dx = \lim_{h\to 0}\frac{F[\phi+h\rho]-F[\phi]}{h} \tag{3.1}$$

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' $n \to \infty$ in (2.45) yields (3.1).

3.1 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 'nice' subset $A \subset \mathcal{T}$. Physicists often write $\delta_{x_i} = \delta(y - x_i)$ as a 'function' of a dummy variable y (which will be integrated over). I will stick to the notation δ_{x_i} because it emphasizes that the object is meant 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. Thus, if the population at any time t consists of $N(t)$ individuals with trait values $\{x_1, x_2, \ldots, x_{N(t)}\}$, then it can be completely characterized (Figure 3.1) by the 'distribution'

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

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

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

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

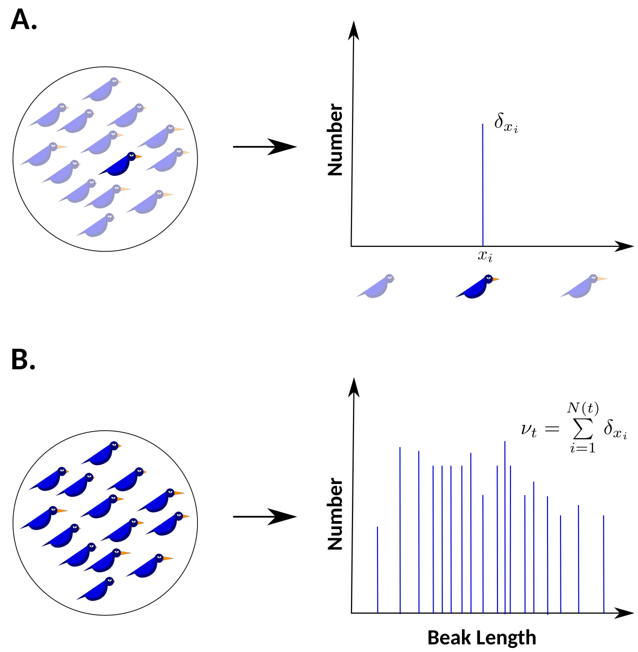

Figure 3.1: Schematic description of a function valued birth-death process. Consider a population of birds in which individuals have varying beak lengths. **(A)** Each individual in the population can be described as a Dirac delta mass centered at its beak length. This is because each individual has 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.

it changes. We assume that there exist non-negative functionals $b(x|\nu)$ and $d(x|\nu)$ which 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, we 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

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

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

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

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.

Example 2. Consider the birth and death functionals:

$$\begin{aligned} b(x|\nu) &= r \int\limits_{\mathcal{T}} m(x,y)\nu(y)dy;\ m(x,y) = \exp\left(\frac{-(x-y)^2}{\sigma_m^2}\right) \\ d(x|\nu) &= \frac{\nu(x)}{Kn(x)} \int\limits_{\mathcal{T}} \alpha(x,y)\nu(y)dy;\ \alpha(x,y) = \exp\left(\frac{-(x-y)^2}{\sigma_\alpha^2}\right) \end{aligned} \tag{3.2}$$

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. It is known (only for one-dimensional traits) that we can find a density function $P(\nu, t)$ such that the probability that the process takes value $\nu^{(t)}$ at time t is given by $\int_{\mathcal{T}} P(\nu, t) dx$.

We can now use the same trick as earlier 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 Dirac delta mass. For any state $\nu \in \mathcal{M}$, 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 outflow rate for a state ν, we see that $P(\nu, t)$ must satisfy:

$$\frac{\partial P}{\partial t}(\nu, t) = \int\limits_{\mathcal{T}} \left[(\mathcal{E}_x^- - 1) b(x|\nu) P(\nu, t) + (\mathcal{E}_x^+ - 1) d(x|\nu) P(\nu, t) \right] dx \tag{3.3}$$

This is the 'Master equation' of our process.

3.2 The functional system-size expansion

To proceed, as before, we assume that there exists a system-size parameter $K > 0$ 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, we assume that there exists a $K > 0$ such that 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

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

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

$$\frac{\partial P}{\partial t}(\phi, t) = K \int_{\mathcal{T}} \left[(\Delta_x^- - 1) b_K(x|\phi) P(\phi, t) + (\Delta_x^+ - 1) d_K(x|\phi) P(\phi, t) \right] dx \tag{3.4}$$

where we have introduced new step operators $\Delta_x^\pm$ 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:

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

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

$$\begin{aligned} \Delta_x^\pm [F[\phi]] &= F[\phi] \pm \frac{1}{K}\int_{\mathcal{T}} \frac{\delta F}{\delta\phi(y)} \delta_x dy + \frac{1}{2K^2}\int_{\mathcal{T}}\int_{\mathcal{T}} \frac{\delta^2 F}{\delta\phi(y)\delta\phi(z)} \delta_x dy \delta_x dz + \mathcal{O}(K^{-3}) \\ &= F[\phi] \pm \frac{1}{K}\frac{\delta F}{\delta\phi(x)} + \frac{1}{2K^2}\frac{\delta^2 F}{\delta\phi(x)^2} + \mathcal{O}(K^{-3}) \end{aligned} \tag{3.5}$$

Neglecting terms of $\mathcal{O}(K^{-3})$, we can now substitute (3.5) into (3.4) to obtain:

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

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

$$\boxed{\frac{\partial P}{\partial t}(\phi, t) = \int_{\mathcal{T}} \left[-\frac{\delta}{\delta \phi(x)} \{\mathcal{A}^{-}(x|\phi) P(\phi, t)\} + \frac{1}{2K} \frac{\delta^2}{\delta \phi(x)^2} \{\mathcal{A}^{+}(x|\phi) P(\phi, t)\} \right] dx} \tag{3.6}$$

where

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

This constitutes the 'mesoscopic' description. For large (but finite) K, equation (3.6) can be analyzed using a weak noise approximation as before.

3.3 The infinite population limit

We can once again appeal to the link between Fokker-Planck equations and Langevins (Lafuerza and McKane, 2016) to say that (3.6) corresponds to the Langevin equation:

$$\frac{\partial \phi}{\partial t}(x, t) = \mathcal{A}^{-}(x|\phi) + \frac{1}{\sqrt{K}} \eta(x, t) \tag{3.7}$$

where $\eta(x, t)$ is the 'Gaussian spacetime white noise' with zero mean and autocovariance function

$$\mathbb{E}[\eta(x, t)\eta(x', t')] = \sqrt{\mathcal{A}^{+}(x|\phi)\mathcal{A}^{+}(x'|\phi)}\,\delta(x - x')\delta(t - t')$$

Taking $K \to \infty$ in equation (3.7) then yields a PDE:

$$\frac{\partial \psi}{\partial t}(x, t) = \mathcal{A}^{-}(x|\psi) = b_K(x|\psi) - d_K(x|\psi) \tag{3.8}$$

where we have used a different symbol ψ simply to highlight that $\psi(x, t)$ as the solution to equation (3.8) is a deterministic function, whereas $\phi(x, t)$ as defined in equation (3.7) is really a stochastic process $\{\phi^{(t)}\}_{t \geq 0}$. Equation (3.8) 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). They are also equivalent to the 'oligomorphic dynamics' of (Sasaki and Dieckmann, 2011; Lion et al., 2022) if one assumes the population is composed of a small number of 'morphs', *i.e.* $\psi(x,t) = \sum_{k=1}^{S} n_k \psi_k(x,t)$, where $\psi_k(x,t)$ is the phenotypic distribution of the kth morph (often assumed a normal distribution with narrow variance) and S is the number of distinct morphs in the population. Models of the form (3.8) are also used to study intraspecific trait variation in community ecology (Nordbotten et al., 2020). A prominent recent example is the 'trait space equations' of (Wickman et al., 2022) in their framework for eco-evolutionary community dynamics.

If one wishes to be mathematically careful, the connection between (3.6) and (3.7) 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 function-valued processes that we are dealing with, and the interpretation of any formal 'Langevin equation' (corresponding now to a stochastic partial differential equation, or SPDE) that we write down is unclear. Nevertheless, we will pretend all is well and assume that one can do this, bolstered by the fact that we can recover some well-known deterministic equations from equation (3.8), as we show below.

3.3.1 Some familiar faces: Kimura-Crow and adaptive dynamics

The 'macroscopic' version of our function-valued process corresponds to quantitative genetics and adaptive dynamics (which is a generalization of evolutionary game theory for quantitative traits). We begin with the deterministic process given by (3.8). We assume that the birth and death functions take the form:

$$\begin{aligned} b_K(x|\psi) &= \psi(x,t) b^{(\mathrm{ind})}(x|\psi) + \mu b^{(\mathrm{mut})}(x|\psi) \\ d_K(x|\psi) &= \psi(x,t) d^{(\mathrm{ind})}(x|\psi) \end{aligned} \tag{3.9}$$

As in chapter 2, $b^{(\mathrm{mut})}(x|\psi)$ describes birth due to mutations and $\mu \geq 0$ is a constant mutation rate. The functions $b^{(\mathrm{ind})}(x|\psi)$ and $d^{(\mathrm{ind})}(x|\psi)$ 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, $b^{(\mathrm{ind})}(x|\psi)$ may incorporate the effects of mate choice in the sexual case or intrinsic duplication rates in the asexual case, and $d^{(\mathrm{ind})}(x|\psi)$ may model death due to intraspecific competition for resources. Substituting equation (3.9) into (3.8), we obtain

$$\frac{\partial \psi}{\partial t}(x,t) = w(x|\psi)\psi(x,t) + \mu b^{(\mathrm{mut})}(x|\psi) \tag{3.10}$$

where we have defined $w(x|\psi) := b^{(\text{ind})}(x|\psi) - d^{(\text{ind})}(x|\psi)$, which can be thought of as the (Malthusian) 'fitness' of the phenotype x. To track population numbers and trait frequencies, we follow the approach of (Week et al., 2021) and define

$$\begin{aligned} N_K(t) &:= \int_{\mathcal{T}} \psi(x,t)dx \\ p(x,t) &:= \frac{\psi(x,t)}{N_K(t)} \end{aligned} \tag{3.11}$$

We can also define the population mean fitness as:

$$\overline{w}(t) = \int_{\mathcal{T}} w(x|\psi)p(x,t)dx \tag{3.12}$$

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 we have used the definition of $N_K(t)$ and assumed that integrals and derivatives commute in the second line. Substituting (3.10), we now obtain

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

where we have used the definition of $p(x,t)$ in the second line. Using (3.12) and rearranging the terms gives us:

$$\boxed{\frac{\partial p}{\partial t}(x,t) = \left[w(x|\psi) - \overline{w}(t)\right]p(x,t) + \frac{\mu}{N_K(t)}\left[b^{(\text{mut})}(x|\psi) - p(x,t)\int_{\mathcal{T}} b^{(\text{mut})}(y|\psi)dy\right]} \tag{3.13}$$

This is a continuous version of the replicator-mutator equation when each x is viewed as a strategy. It also yields Kimura's continuum-of-alleles model when each x is viewed as an allele, $b^{(\mathrm{mut})}(x|\psi)$ 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 $b^{(\mathrm{mut})}(y|\psi) = \int_{\mathbb{R}} m(y-z)\psi(z,t)dz$, where $m(x)$ is a mutation kernel, which by definition is normalized such that $\int_{\mathbb{R}} m(x)dx = 1$. Let us further note that we have implicitly been assuming that the total number of individuals (scaled by K) remains finite at all times, *i.e.* $N_K(t) = \int_{\mathbb{R}} \psi(x,t)dx < \infty \ \forall\ t$. Thus, $m(x)\psi(y,t) \in \mathcal{L}^1(\mathbb{R}\times\mathbb{R}) \ \forall\, t$ and 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 (3.13).

We have:

$$\begin{aligned}
\int_{\mathbb{R}} b^{(\mathrm{mut})}(y|\psi)dy &= \int_{\mathbb{R}}\int_{\mathbb{R}} m(y-z)\psi(z,t)dzdy \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} m(y-z)\psi(z,t)dydz \\
&= \int_{\mathbb{R}} \psi(z,t)\left(\int_{\mathbb{R}} m(y-z)dy\right)dz \\
&= \int_{\mathbb{R}} \psi(z,t)\int_{\mathbb{R}} m(u)dudz \\
&= \int_{\mathbb{R}} \psi(z,t)dz\int_{\mathbb{R}} m(u)du \\
&= N_K(t)\int_{\mathbb{R}} m(u)du
\end{aligned} \tag{3.14}$$

where we 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 (3.14) therefore becomes $\int_{\mathbb{R}} b^{(\mathrm{mut})}(y|\psi)dy = N_K(t)$. Substituting this in (3.13), we have

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

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

$$\boxed{\frac{\partial p}{\partial t}(x,t) = [w(x|\psi) - \overline{w}(t)]\, p(x,t) + \mu \left[\int_{\mathbb{R}} m(x-z)p(z,t)dz - p(x,t) \right]} \tag{3.15}$$

which is Kimura's continuum of alleles model.
Note that if we define the mean trait value as

$$\overline{x}(t) = \int_{\mathcal{T}} xp(x,t)dx$$

then, by multiplying both sides of equation (3.13) by x and integrating over the trait space, we obtain

$$\begin{aligned}
\frac{d\overline{x}}{dt} &= \int_{\mathcal{T}} xw(x|\psi)p(x,t)dx - \overline{w}(t) \int_{\mathcal{T}} xp(x,t)dx + \frac{\mu}{N_K(t)} \int_{\mathcal{T}} x \left[b^{(\text{mut})}(x|\psi) - p(x,t) \int_{\mathcal{T}} b^{(\text{mut})}(y|\psi)dy \right] dx \\
&= \overline{xw} - \overline{w} \cdot \overline{x} + \frac{\mu}{N_K(t)} \int_{\mathcal{T}} x \left[b^{(\text{mut})}(x|\psi) - p(x,t) \int_{\mathcal{T}} b^{(\text{mut})}(y|\psi)dy \right] dx
\end{aligned} \tag{3.16}$$

We now observe that

$$\text{Cov}(x, w(x|\psi)) = \overline{xw} - \overline{x} \cdot \overline{w} \tag{3.17}$$

is the statistical covariance of the trait value with the Malthusian fitness function (Importantly, just like in the Price equation, this is an *analogy* - Everything here is deterministic). The second term, which we will denote by

$$M(x|\psi) \coloneqq \frac{\mu}{N_K(t)} \left[\int_{\mathcal{T}} xb^{(\text{mut})}(x|\psi)dx - \left(\overline{x} \int_{\mathcal{T}} b^{(\text{mut})}(x|\psi)dx \right) \right] \tag{3.18}$$

reflects the transmission bias of mutations. Thus, we see that equation (3.16) reads

$$\boxed{\frac{d\overline{x}}{dt} = \text{Cov}(x, w(x|\psi)) + M(x|\psi)} \tag{3.19}$$

from which it is clear that we have obtained a version of the Price equation for quantitative traits. Some more familiar dynamics are recovered under the following additional assumptions:

- Rare mutations, *i.e.* $\mu \to 0$.
- Small mutational effects with 'almost faithful' reproduction, meaning $b^{(\mathrm{mut})}(x|\psi) \to 0$, 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 $w(y|\delta_{y(t)}) = 0$.

The first two assumptions are sometimes called the 'weak mutation' limit, and the last is sometimes called the 'strong selection' limit, both for obvious reasons. 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), then it is reasonable to assume that the population remains sufficiently 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 $p(x,t) = \delta_{y(t)}$, $\overline{x}(t) = y(t)$, and $\overline{w}(t) = 0$. Thus, from equation (3.19), we have

$$\begin{aligned} \frac{d\overline{x}}{dt} &= \int_{\mathcal{T}} (x - \overline{x}(t))(w(x|\psi) - \overline{w}(t))p(x,t)dx \\ \Rightarrow \frac{dy}{dt} &= \int_{\mathcal{T}} (x - \overline{x}(t))w\left(x|N_K\delta_{y(t)}\right)\delta_{y(t)}dx \end{aligned} \tag{3.20}$$

Our 'weak mutation' assumptions 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 $w\left(x|N_K\delta_{y(t)}\right)$ about $y(t)$ as:

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

Thus, substituting in (3.20), to first order, we obtain

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

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

$$\boxed{\frac{dy}{dt} = B(y) \left(\frac{d}{dz} w \left(z | N_K \delta_{y(t)} \right) \Big|_{z=y} \right)} \tag{3.21}$$

Note that by definition, $B(y(t))$ is the variance of the trait in the population at time t. The term $w\left(z|N_K \delta_{y(t)}\right)$ 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 (3.21) represents a broad class of systems called 'gradient equations' in quantitative genetics (Lande, 1976; 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 (Dieckmann and Law, 1996; Doebeli, 2011). 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 (3.21), $B(y)$ 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.

3.4 Stochastic fluctuations and the weak noise approximation

We can now formally carry out a functional analogue of the weak noise expansion. Assume that $\psi(x,t)$ is the deterministic trajectory obtained as the solution to (3.8). 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, we introduce the new variables:

$$\begin{aligned} \zeta^{(s)}(x) &= \sqrt{K}(\phi^{(t)}(x) - \psi(x,t)) \\ s &= t \\ \tilde{P}(\zeta, s) &= \frac{1}{\sqrt{K}} P(\phi, t) \end{aligned} \tag{3.22}$$

Note that the following relations hold:

$$\frac{\delta F[\zeta]}{\delta \phi(x)} = \int_{\mathcal{T}} \frac{\delta F[\zeta]}{\delta \zeta(y)} \frac{\delta \zeta(y)}{\delta \phi(x)} dy = \sqrt{K} \frac{\delta F[\zeta]}{\delta \zeta(x)} \tag{3.23}$$

$$\frac{\partial}{\partial s} = \frac{\partial}{\partial t} \tag{3.24}$$

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

$$\begin{aligned} \frac{\partial \tilde{P}}{\partial t}(\zeta, s) &= \frac{\delta \tilde{P}}{\delta \zeta} \frac{\partial \zeta}{\partial t} + \frac{\partial \tilde{P}}{\partial s} \frac{\partial s}{\partial t} \\ &= \frac{\delta \tilde{P}}{\delta \zeta} \left(-\sqrt{K} \frac{\partial \psi}{\partial t} \right) + \frac{\partial \tilde{P}}{\partial s} \\ &= -\sqrt{K} \frac{\delta}{\delta \zeta} \{\mathcal{A}^-(x|\psi) \tilde{P}(\zeta, s)\} + \frac{\partial \tilde{P}}{\partial s} \end{aligned} \tag{3.25}$$

Reformulating equation (3.6) in terms of the new variables (3.22) and using the relations (3.23), (3.24) and (3.25), we obtain:

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

and rearranging gives us:

$$\begin{aligned}\frac{\partial\tilde{P}}{\partial s} &= -\sqrt{K}\int\limits_{\mathcal{T}}\frac{\delta}{\delta\zeta(x)}\left\{\left(\mathcal{A}^{-}\left(x\middle|\psi+\frac{\zeta}{\sqrt{K}}\right)-\mathcal{A}^{-}(x|\psi)\right)\tilde{P}(\zeta,s)\right\}dx \\ &+\frac{1}{2}\int\limits_{\mathcal{T}}\frac{\delta^2}{\delta\zeta(x)^2}\{\mathcal{A}^{+}\left(x\middle|\psi+\frac{\zeta}{\sqrt{K}}\right)\tilde{P}(\zeta,s)\}dx\end{aligned} \tag{3.26}$$

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

$$\begin{aligned}\mathcal{A}^{-}\left(x\middle|\psi+\frac{\zeta}{\sqrt{K}}\right) &= \mathcal{A}^{-}(x|\psi)+\frac{1}{\sqrt{K}}\int\limits_{\mathcal{T}}\zeta(y)\frac{\delta}{\delta\psi(y)}\{\mathcal{A}^{-}(y|\psi)\}dy+\cdots \\ \mathcal{A}^{+}\left(x\middle|\psi+\frac{\zeta}{\sqrt{K}}\right) &= \mathcal{A}^{+}(x|\psi)+\frac{1}{\sqrt{K}}\int\limits_{\mathcal{T}}\zeta(y)\frac{\delta}{\delta\psi(y)}\{\mathcal{A}^{+}(y|\psi)\}dy+\cdots\end{aligned}$$

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 (3.26) and equating coefficients of powers of K, we see that upto leading order in K (corresponding to the zeroth order terms of $\tilde{P}$ and $\mathcal{A}^{+}\left(x\middle|\psi+\frac{\zeta}{\sqrt{K}}\right)$, and the first order term of $\mathcal{A}^{-}\left(x\middle|\psi+\frac{\zeta}{\sqrt{K}}\right)$) we have:

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

We thus arrive at the functional Fokker-Planck equation:

$$\frac{\partial\tilde{P}_0}{\partial s}(\zeta,s)=\int\limits_{\mathcal{T}}\left(-\frac{\delta}{\delta\zeta(x)}\left\{\mathcal{D}_\zeta[\mathcal{A}^{-}](x)\tilde{P}_0(\zeta,s)\right\}+\frac{1}{2}\mathcal{A}^{+}(x|\psi)\frac{\delta^2}{\delta\zeta(x)^2}\{\tilde{P}_0(\zeta,s)\}\right)dx \tag{3.27}$$

where

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

can be thought of now as the functional analogue of a directional derivative of $\mathcal{A}^{-}(x|\psi)$ in the direction of the function ζ.

Part III

Summary & Discussion

Chapter 4

A unified view of population dynamics

> The grand aim of all science [is] to cover the greatest number of empirical facts by logical deduction from the smallest number of hypotheses or axioms
>
> Albert Einstein

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 (corresponding to discrete trait variants), the infinite population limit corresponds to the equations of population genetics and evolutionary game theory. In the infinite-dimensional case (corresponding to quantitative traits), we instead obtain the equations of quantitative genetics, and, in some further limits, adaptive dynamics. In both cases, the mean value of the trait in the population changes according to an equation resembling the Price equation. My derivatioon highlights the natural connections between the various equations of population dynamics - For example, the same procedures that lead to the replicator-mutator equation in the case of discretely varying traits yield Kimura's model in the quantitative case, underscoring the broad similarities between evolutionary game theory and quantitative genetics. The major formulations are summarized in Table 4.1.

Number of possible distinct trait variants (m)	**State Space**	**Model parameters**	**Mesoscopic description**	**Infinite population limit**
$m = 1$ (Identical individuals)	$[0, 1, 2, 3, \ldots]$ (Population size)	Two real-valued 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)	Single species population dynamics
$1 < m < \infty$ (Discrete traits)	$[0, 1, 2, 3, \ldots]^m$ (Number of individuals of each trait variant)	$2m$ real-valued functions, $b_i(\mathbf{v})$ and $d_i(\mathbf{v})$ (for $1 \leq i \leq m$) describing the birth and death rate of trait variant i when the population is $\mathbf{v}$	Multivariate Fokker-Planck equation (m-dimensional SDEs)	Evolutionary game theory Lotka-Volterra competition Quasispecies equation Price equation (discrete traits)
$m = \infty$ (Quantitative traits)	$\left\{\sum_{i=1}^{n} \delta_{x_i} \mid n \in \mathbb{N}\right\}$ (Each Dirac mass δ_{x_i} is an individual with trait value x_i)	Two real-valued functionals $b(x\|\nu)$ and $d(x\|\nu)$ describing the birth and death rate of trait variant x when the population is ν	Functional Fokker-Planck equation/Field theory (SPDEs)	Kimura's continuum-of-alleles model Sasaki and Dieckmann, 2011's Oligomorphic Dynamics Wickman et al., 2022's Trait Space Equations for intraspecific trait variation Gradient Dynamics Price equation (quantitative traits)

Table 4.1: Summary of the various birth-death processes studied in this thesis

4.1 The fundamental theorem for changes in type frequencies in the population

Equation (2.35), which we derived in chapter 2, 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 types of individuals, and used ecological arguments to posit the existence of a 'system-size' parameter K that leads to density-dependent growth and prevents the population from growing infinitely large. The population as a whole is characterized by a vector $\mathbf{x} = [x_1, \ldots, x_m]$ indexing the density (*i.e.* number divided by K) of each type of individual. Changes of the population are through either birth or death of individuals. 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 mutational effects. All three of these functions depend on the density (and *not* just the total number) of indidivuals of each type in the population, and may in general also be frequency-dependent. In the regime where K is not too small (corresponding to 'medium sized' populations), we identified two quantities, $w_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) - d_i^{(\text{ind})}(\mathbf{x})$ and $\tau_i(\mathbf{x}) = b_i^{(\text{ind})}(\mathbf{x}) + d_i^{(\text{ind})}(\mathbf{x})$, the Malthusian fitness and per-capita turnover rate of the i^{th} type respectively, that emerge as being important for trait frequency dynamics. In particular, we saw that p_i, the frequency of the i^{th} type in the population, changes according to the equation:

$$\begin{aligned} dp_i(t) = & \underbrace{\left[(w_i(\mathbf{x}) - \overline{w})p_i + \mu\left\{Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right\}\right] dt}_{\substack{\text{Infinite population predictions: selection-mutation balance} \\ \text{for higher fitness}}} \\ & - \frac{1}{K}\underbrace{\frac{1}{N_K(t)}\left[(\tau_i(\mathbf{x}) - \overline{\tau})p_i + \mu\left\{Q_i(\mathbf{p}) - p_i\left(\sum_{j=1}^{m} Q_j(\mathbf{p})\right)\right\}\right] dt}_{\substack{\text{Directional noise-induced effects: selection-mutation balance} \\ \text{for lower turnover rates}}} \\ & + \frac{1}{\sqrt{K}N_K(t)}\underbrace{\left[\left(A_i^+\right)^{1/2} dB_t^{(i)} - p_i\sum_{j=1}^{m}\left(A_j^+\right)^{1/2} dB_t^{(j)}\right]}_{\substack{\text{Non-directional noise-induced effects} \\ \text{due to stochastic fluctuations}}} \end{aligned} \tag{4.1}$$

where $N_K = \sum x_i$ is the total population size scaled by K (and thus KN_K is the total population size), $A_i^+ = x_i\tau_i(\mathbf{x}) + \mu Q_i(\mathbf{x})$, and each $B_t^{(i)}$ is an independent one-dimensional standard Brownian motion. Equation (4.1) is in 'replicator-mutator' form, and letting $K \to$

∞ recovers the standard replicator-mutator equation. The first term 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 (4.1), 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 (4.1)).

4.2 The fundamental theorem for the mean value of a type-level quantity in the population

We can now calculate how the statistical mean value of a type-level quantity changes over time. Let f be any type level quantity, with value $f_i(t)$ for the i^{th} type. We allow for the possibility of f_i to vary over time. By multiplying both sides of equation (4.1) by f_i and summing over all i (The same steps as going from (2.37) to (2.40)), we see that the statistical mean $\overline{f}$ of the quantity in the population varies as:

$$
\begin{aligned}
d\overline{f} = {} & \underbrace{\mathrm{Cov}(w,f)dt}_{\substack{\text{Classical}\\\text{selection}}} - \underbrace{\frac{1}{KN_K(t)}\mathrm{Cov}(\tau,f)dt}_{\substack{\text{Noise-induced}\\\text{selection}}} + \underbrace{\overline{\left(\frac{\partial f}{\partial t}\right)}dt}_{\substack{\text{Ecological timescale}\\\text{feedbacks due to}\\\text{time-dependence of } f_i}} \\
& + \underbrace{\mu\left(1-\frac{1}{KN_K(t)}\right)\left(\sum_{i=1}^{m} f_iQ_i(\mathbf{p}) - \overline{f}\sum_{i=1}^{m} Q_i(\mathbf{p})\right)dt}_{\text{Transmission bias/mutational effects}} \\
& + \underbrace{\frac{1}{\sqrt{K}N_K(t)}\left(\sum_{i=1}^{m}\left(f_i-\overline{f}\right)\sqrt{A_i^+}dB_t^{(i)}\right)}_{\text{Stochastic fluctuations}}
\end{aligned}
\tag{4.2}
$$

where all covariances are understood in the statistical sense (Note that since w_i, τ_i, $\overline{w}$, and $\overline{\tau}$ are stochastic processes depending on $\mathbf{p}$, the terms $\mathrm{Cov}(w,f)$ and $\mathrm{Cov}(\tau,f)$ are themselves stochastic processes). Taking $K \to \infty$ in equation (4.2) recovers the standard Price equation as the infinite population limit (either (2.40) or (2.41) based on whether f_i varies with time).

We saw in chapter 3 using field equations that very similar methods of attack to those used outlined in chapter (2) also hold for quantitative traits. For example, equation (3.13) and (3.19) are respectively exactly the infinite dimensional analogs of the deterministic replicator-mutator equation (2.37) and the deterministic Price equation (2.41) when f is the trait value. We may therefore expect to find equations similar to (4.1) and (4.2) for quantitative traits. Indeed, measure-theoretic tools have recently been used to rigorously show that an infinite-dimensional version of (4.2) holds for one-dimensional quantitative traits when f is the trait value and $b(x|\phi) \pm d(x|\phi)$ are Gaussian (Week et al., 2021). Equations (4.1) and (4.2) are thus fundamental theorems for the evolution of finite populations, with the replicator-mutator and Price equations as their respective infinite population limits (also see (Rice, 2020) for a stochastic Price equation in a discrete-time setting).

Each term in equation (4.2) lends itself to a simple biological interpretation. The first term, $\mathrm{Cov}(w, f)$, 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 (4.2), the effects of the second term of (4.1) decompose into a selection term $\mathrm{Cov}(\tau, f)$ for reduced turnover rates and a transmission bias term that vanishes in the weak mutation ($\mu \to 0$) limit. Following Week et al., 2021, we refer to the effect of the covariance (the second term of equation (4.2)) 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. Since this evolutionary force is unique to finite populations and has therefore been overlooked in classical population genetics, it warrants some more detailed discussion. Biologically, the $\mathrm{Cov}(\tau, f)$ 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 experiences greater turnover due to a generally higher birth and death rate and thus experience 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. 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. This stochastic analog of selection for reduced turnover rates, captured by the second term of equation (4.1), is the force responsible for the 'reversal of the direction of deterministic selection' induced by demographic noise in previous studies (Houchmandzadeh and Vallade, 2012; Houchmandzadeh, 2015; Constable et al., 2016;

McLeod and Day, 2019). 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, which could explain why this effect preferentially favors the former types in reversing the direction of deterministic selection in finite population models (Houchmandzadeh, 2015; McLeod and Day, 2019). The fact that total population size controls the strength of the second term of (4.1) also explains why cooperation is favoured in the early transient period of population growth (Melbinger 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 favouring cooperation. More generally, selection for reduced turnover rate could explain why cooperation often persists for a long time in finite population IbMs (and the real world) despite infinite population models predicting their extinction. The fact that the entire term is multiplied by $(KN_K(t))^{-1}$ suggests that the effect of this force is weak for medium to large populations, which explains why the persistence of cooperators is often only observed in restrictive sounding conditions such as quasi-neutrality, rapid attraction to a slow manifold, or a weak selection + weak mutation limit. In all three of these cases, the first term on the RHS of (2.35) becomes identically 0. It therefore no longer contributes to the trait frequency dynamics, thus allowing us to see the contributions of the second term.

The third term of (4.2) 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 is often through a changing environment or phenotypic/behavioral plasticity, but other biological phenomena may also be at play. The fourth term of (4.2) is a transmission bias term, with a correction factor due to noise-induced selection. Finally, the last term of (4.2) describes the role of stochastic fluctuations. The contributions of this last term are 'directionless' due to the dB_t factors, and this term vanishes when we take a conditional expectation value over the underlying probability space. We denote this probabilistic expectation value operation by $\mathbb{E}[\cdot]$ to distinguish it from the statistical mean (2.34). 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 (4.2) are realized upon ignoring mutations by setting $\mu = 0$ and then substituting $f = w$ and $f = \tau$. We first note that:

$$\mathrm{Cov}(w, \tau) = \mathrm{Cov}\left(b^{(\mathrm{ind})}(\mathbf{x}) - d^{(\mathrm{ind})}(\mathbf{x})\ ,\ b^{(\mathrm{ind})}(\mathbf{x}) + d^{(\mathrm{ind})}(\mathbf{x})\right) \tag{4.3}$$

$$= \sigma^2_{b^{(\text{ind})}(\mathbf{x})} - \sigma^2_{d^{(\text{ind})}(\mathbf{x})} \tag{4.4}$$

where we have defined the statistical variance of a quantity f as

$$\sigma_f^2 := \overline{(f^2)} - (\overline{f})^2 \tag{4.5}$$

Note that just like the statistical mean, the statistical variance is a random variable and is not to be confused with the probabilistic/ensemble variance. Upon substituting $f = w$ in (4.2) and taking expectations over the underlying probability space, we obtain:

$$\mathbb{E}\left[\frac{d\overline{w}}{dt}\right] = \underbrace{\mathbb{E}\left[\sigma_w^2\right]}_{\substack{\text{Fisher's} \\ \text{Fundamental theorem}}} - \underbrace{\mathbb{E}\left[\frac{\sigma^2_{b^{(\text{ind})}} - \sigma^2_{d^{(\text{ind})}}}{KN_K(t)}\right]}_{\substack{\text{Noise-induced} \\ \text{selection}}} + \underbrace{\mathbb{E}\left[\overline{\frac{\partial w}{\partial t}}\right]}_{\substack{\text{Short-term (ecological)} \\ \text{feedbacks to fitness}}} \tag{4.6}$$

Taking $K \to \infty$ in (4.6) recovers a well-known equation in population genetics upon noting that the process tends to a deterministic process as $K \to \infty$, as noted in section 2.3.4, 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 last term arises in both finite and infinite populations whenever w_i can vary over time (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. Fisher appears to have been rather vague and dismissive of this feedback (Fisher, 1930), and this 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). Finally, the second term of equation (4.6) is a manifestation of noise-induced selection and vanishes in the infinite population limit, and is thus particular to finite populations.

Carrying out the same steps with $f = \tau$ in (4.2) yields a new equation/theorem due to Kuosmanen et al., 2022 that has only recently been recognized as important. This theorem is an analog of Fisher's fundamental theorem for the turnover rates, and reads:

$$\mathbb{E}\left[\frac{d\overline{\tau}}{dt}\right] = \underbrace{\mathbb{E}\left[\sigma^2_{b^{(\text{ind})}} - \sigma^2_{d^{(\text{ind})}}\right]}_{\substack{\text{Classical selection} \\ \text{effects}}} - \underbrace{\mathbb{E}\left[\frac{\sigma_\tau^2}{KN_K(t)}\right]}_{\substack{\text{Noise-induced selection} \\ \text{effects}}} + \underbrace{\mathbb{E}\left[\overline{\frac{\partial \tau}{\partial t}}\right]}_{\substack{\text{Short-term (ecological)} \\ \text{feedbacks to } \tau_i}} \tag{4.7}$$

The implications of this theorem have been extensively discussed in (Kuosmanen et al., 2022), which is where we refer the interested reader. Not sure if there's much more to write?

4.3 The fundamental theorem for the variance of a type-level quantity in the population

Equation (4.2) 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 the statistical variance of any type level quantity f obeys

$$\begin{aligned} d\sigma_f^2 &= \mathrm{Cov}\left(w, (f-\overline{f})^2\right)dt - \frac{1}{KN_K}\left[\,\overline{\tau}\sigma_f^2 + 2\mathrm{Cov}\left(\tau, (f-\overline{f})^2\right)\,\right]dt \\ &+ 2\mathrm{Cov}\left(\frac{\partial f}{\partial t}, f\right)dt + M_{\sigma_f^2}(\mathbf{p}, N_K)dt + dB_{\sigma_f^2} \end{aligned} \tag{4.8}$$

where

$$M_{\sigma_f^2}(\mathbf{p}, N_K) = \mu\left[\left(1-\frac{2}{KN_K}\right)\left(\sum_{i=1}^{m}(f_i-\overline{f})^2 Q_i(\mathbf{p})\right) + \sigma_f^2\left(1-\frac{1}{KN_K}\right)\sum_{i=1}^{m}Q_i(\mathbf{p})\right] \tag{4.9}$$

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

$$dB_{\sigma_f^2} = \frac{1}{\sqrt{K}N_K(t)}\left(\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2\sqrt{A_i^+}dB_t^{(i)}\right) \tag{4.10}$$

is a stochastic integral term measuring the (non-directional) effect of stochastic fluctuations that vanishes upon taking an expectation over the probability space. In the case of one-dimensional quantitative traits, an infinite-dimensional version of (4.8) 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 (4.8) 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 (Equation A.23 in Débarre and Otto, 2016 is equivalent to equation (4.8) for their choice of functional forms upon converting their change in variance to an infinitesimal rate of change *i.e.* a derivative).

Once again, terms of equation (4.8) lend themselves to straightforward biological interpretation. The quantity $(f_i - \overline{f})^2$ is a measure of the distance of f_i from the population mean value $\overline{f}$, and thus covariance with $(f - \overline{f})^2$ 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 $f_i \not\equiv \overline{f}$ (*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 $\mathrm{Cov}(w, (f - \overline{f})^2)$ is strongly positive and the population experiences a sudden explosion in variance. If $\mathrm{Cov}(w, (f - \overline{f})^2)$ is still positive after the initial emergence of multiple morphs, evolution eventually leads to the emergence of stable coexisting polymorphisms in the population. The $\mathrm{Cov}\,(\partial f/\partial t, f)$ 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 (2.32) 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 (4.9), 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.

The $\overline{\tau}\sigma_f^2$ term represents a loss of diversity due to stochastic extinctions (i.e. demographic stochasticity). 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, we have $\mathrm{Cov}\left(w, (f - \overline{f})^2\right) \equiv \mathrm{Cov}\left(\tau, (f - \overline{f})^2\right) \equiv 0$. Since the labels do not change over time, we also have $\mathrm{Cov}\,(\partial f/\partial t, f) = 0$. From (4.8), we see that in this case, the variance changes as

$$d\sigma_f^2 = -\frac{\overline{\tau}\sigma_f^2}{KN_K}dt + dB_{\sigma_f^2} \tag{4.11}$$

On taking expectations, we see that the expected variance in the population obeys

$$\frac{d\mathbb{E}[\sigma_f^2]}{dt} = -\left(\mathbb{E}\left[\frac{\overline{\tau}}{KN_K}\right]\right)\mathbb{E}[\sigma_f^2] \tag{4.12}$$

where we 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 $\overline{\tau}$ and $N_K(t)$. Equation (4.12) is a simple linear ODE and has the solution

$$\mathbb{E}[\sigma_f^2](t) = \sigma_f^2(0)e^{-\mathbb{E}\left[\frac{\overline{\tau}}{KN_K}\right]t} \tag{4.13}$$

which tells us that the expected diversity (variance) of labels in the population decreases exponentially over time. The rate of loss is $\mathbb{E}\left[\overline{\tau}(KN_K)^{-1}\right]$, and thus, populations with higher mean turnover $\overline{\tau}$ and/or lower population size KN_K lose diversity faster. This is because populations with higher $\overline{\tau}$ 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.

4.4 Weak-Noise approximation

If the population is large (but finite) and the stochasticity is sufficiently weak, stochastic dynamics can be studied analytically using the weak noise approximation. Usually, this approximation is valid if we are studying a stochastic trajectory that is fluctuating about a point that is stable in the deterministic limit (Van Kampen, 1981). Such situations occur often, since many systems of the forms studied here quickly relax to stable equilibria.

4.5 Towards a stochastic evolutionary theory

One striking feature that repeatedly shows up in our derivations is that finite populations exhibit phenomena that are not visible in infinite population models. For example, in both the stochastic logistic equation D.1 and in two-strategy games with finite population sizes (Tao and Cressman, 2007), demographic noise ensures that all populations are guaranteed to go extinct given enough time, even if the deterministic limit predicts a stable state far from extinction. In the case of quantitative traits, demographic noise can hinder adaptive diversification by increasing the time before evolutionary branching occurs (Claessen et al., 2007; Wakano and Iwasa, 2013; Débarre and Otto, 2016), causing stochastic extinction of existing evolutionary branches (Rogers et al., 2012a; Johansson and Ripa, 2006), or preventing branching altogether if the population is too small (Rogers and McKane, 2015; Johnson et al., 2021). Stochastic systems also routinely exhibit evolution towards attractors that cannot be attained in the deterministic limit (DeLong and Cressler, 2023), sometimes even completely reversing the direction of evolution predicted by deterministic dynamics (Constable et al., 2016; McLeod and Day, 2019). Since real-life populations are stochastic and finite, it is thus imperative that modellers work with stochastic first-principles models instead of their deterministic limits, lest they risk missing important phenomena that are unique to stochastic systems (Black and McKane, 2012; Schreiber et al., 2022; Hastings, 2004; Shoemaker et al., 2020). In the context of our models, we have seen that if we observe the change in trait frequencies instead of the change in densities, finite populations are subject to an additional evolutionary force that vanishes in infinite population models.

Appendix A

Deriving the Fokker-Planck equations for Itô SDEs

Here, I present a simple (informal) derivation of the Fokker-Planck equation (FPE) 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 $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dB_t$ on a filtered probability space $\Omega \subseteq \mathbb{R}$ with probability measure $\mathbb{P}$ such that $\mathbb{P}(\cdot) \equiv 0$ on $\partial\Omega$ and $\mathbb{P} \ll m$, where m is the Lebesgue measure. The latter requirement allows us to use the Radon-Nikodym theorem to write $\int \cdot \, d\mathbb{P} = \int \cdot \, P(x,t)dx$, where $P(x,t)$ is a 'probability density function' defined at every point in $\Omega \times [0, \infty)$. Now, Let $f : \mathbb{R} \to \mathbb{R}$ be an arbitrary $C^2(\mathbb{R})$ function. By Itô's lemma, we have:

$$df(X_t) = f' dX_t + \frac{1}{2} f'' d\langle X \rangle_t$$

where $\langle \cdot \rangle$ denotes the quadratic variation. For $dX_t = \mu dt + \sigma dB_t$, it is clear that $d\langle X \rangle_t = \sigma^2 d\langle B \rangle_t = \sigma^2 dt$, and thus, we obtain:

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

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

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

Since the Brownian motion is a martingale, as long as X_t and $\sigma(X_t, t)$ are reasonably 'nice'[1], the stochastic integral in the second term of the RHS of (A.1) will be a continuous $L^2(\mathbb{P})$ martingale starting at the origin, and its expectation will therefore be 0. Using the definition of 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 commute, we can now differentiate with respect to time on both sides and use the fundamental theorem of calculus to write

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

We will now use integration by parts to further evaluate $I(x,t)$ and $J(x,t)$. Recall that the general formula for integration by parts is given by:

$$\int_\Omega u_{x_i} v dx = -\int_\Omega u v_{x_i} dx + \int_{\partial\Omega} uv\gamma_i dS(x)$$

where subscript indicates differentiation and γ is the unit outward normal. In our case, assuming that $P(x,t) \equiv 0$ on $\partial\Omega$, the boundary term (second term of the RHS) vanishes and we can use integration by parts once on $I(x,t)$ to obtain

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

[1]We require $\sigma(X_t,t)f'(X_t) \in \mathcal{L}^*(B)$, which is a highly technical condition. Since existence/uniqueness of solutions for the SDE already requires Lipschitz continuity of $\sigma(X_t,t)$, this seems like a reasonable assumption to make.

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

$$J(x,t) = -\frac{1}{2}\int\limits_{\Omega} f'(X_t)\left(\frac{\partial}{\partial x}\sigma^2 P(x,t)\right)dx$$
$$= \frac{1}{2}\int\limits_{\Omega} f(X_t)\left(\frac{\partial^2}{\partial x^2}\sigma^2 P(x,t)\right)dx \tag{A.4}$$

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

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

Since this is true for an arbitrary choice of $f(x)$ (as long as f is C^2), we are thus led to conclude that the density function $P(x,t)$ must satisfy:

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

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

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

where $\mathbf{D} = \sigma\sigma^T$.

Appendix B

Deriving 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

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

where $\boldsymbol{\mu} : \mathbb{R}^m \to \mathbb{R}^m$ is the 'drift vector' and $\boldsymbol{\sigma} : \mathbb{R}^m \to \mathbb{R}^{m\times m}$ is the 'diffusion matrix'. Let $f : \mathbb{R}^m \to \mathbb{R}$ 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:

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

where $\mathrm{Tr}[\cdot]$ denotes the trace of a matrix, $(\cdot)^{\mathrm{T}}$ 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 $f(x_1, \ldots, x_m)$ 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 (2.31), 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 $i \in \{1, 2, \ldots, m\}$, let us define a scalar function $f_i : \mathbb{R}^m \to \mathbb{R}$ as

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

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:

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

where we have defined the total (scaled) population size[1] $N = \sum_i x_i$ and the frequency of the i^{th} type $p_i = f_i(x)$ and used the fact that $\frac{\partial x_j}{\partial x_k} = \delta_{jk}$. The jk^{th} element of the Hessian is given by:

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

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

$$(\nabla_{\mathbf{X}} f_i)^{\mathrm{T}} \mathbf{A}^{-} = \sum_{j=1}^{m} \left((\nabla_{\mathbf{x}} f_i)_j \right) A_j^{-}$$

[1] This is $N_K(t)$ in the main text, but we omit the subscript here to reduce notational clutter

$$= \frac{1}{N} \sum_{j=1}^{m} (\delta_{ij} - p_i) A_j^-$$

$$= \frac{1}{N} \left(A_i^- - p_i \sum_{j=1}^{m} A_j^- \right) \tag{B.4}$$

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:

$$\frac{1}{\sqrt{K}} (H_{\mathbf{x}} f_i \mathbf{D})_{jk} = \frac{1}{\sqrt{K}} \sum_{l=1}^{m} (H_{\mathbf{x}} f_i)_{jl} (\mathbf{D})_{lk}$$

$$= \frac{1}{\sqrt{K} N^2} \sum_{l=1}^{m} (2p_i - \delta_{ij} - \delta_{il}) \delta_{lk} \left(A_l^+ A_k^+ \right)^{\frac{1}{4}} \tag{B.5}$$

$$= \frac{1}{\sqrt{K} N^2} \left((2p_i - \delta_{ij}) (A_k^+)^{\frac{1}{2}} - \delta_{ik} \left(A_i^+ A_k^+ \right)^{\frac{1}{4}} \right) \tag{B.6}$$

$$= \frac{1}{\sqrt{K} N^2} (2p_i - \delta_{ij} - \delta_{ik}) (A_k^+)^{\frac{1}{2}} \tag{B.7}$$

and thus:

$$\frac{1}{K} \left(\mathbf{D}^{\mathrm{T}} H_{\mathbf{x}} f_i \mathbf{D} \right)_{lk} = \frac{1}{K} \sum_{j=1}^{m} \left(\mathbf{D}^{\mathrm{T}} \right)_{lj} (H_{\mathbf{x}} f_i \mathbf{D})_{jk}$$

$$= \frac{1}{K N^2} \sum_{j=1}^{m} \delta_{lj} \left(A_l^+ A_j^+ \right)^{\frac{1}{4}} (A_k^+)^{\frac{1}{2}} (2p_i - \delta_{ij} - \delta_{ik}) \tag{B.8}$$

$$= \frac{1}{K N^2} (A_k^+)^{\frac{1}{2}} \left(2p_i (A_l^+)^{\frac{1}{2}} - (A_i^+)^{\frac{1}{2}} \delta_{il} - (A_l^+)^{\frac{1}{2}} \delta_{ik} \right) \tag{B.9}$$

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

$$\frac{1}{K} \mathrm{Tr}[\mathbf{D}^{\mathrm{T}} H_{\mathbf{x}} f_i \mathbf{D}] = \frac{1}{K} \sum_{k=1}^{m} \left(\mathbf{D}^{\mathrm{T}} H_{\mathbf{x}} f_i \mathbf{D} \right)_{kk}$$

$$= \frac{1}{K N^2} \sum_{k=1}^{m} \left(2p_i (A_k^+ A_k^+)^{\frac{1}{2}} - (A_i^+ A_k^+)^{\frac{1}{2}} \delta_{ik} - (A_k^+ A_k^+)^{\frac{1}{2}} \delta_{ik} \right) \tag{B.10}$$

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

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

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

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

$$\begin{aligned}
(\mathbf{D}d\mathbf{B}_t)_j &= \sum_{k=1}^{m}\mathbf{D}_{jk}dB_t^{(k)} \\
&= \sum_{k=1}^{m}\delta_{jk}\left(A_j^+A_k^+\right)^{\frac{1}{4}}dB_t^{(k)} && \text{(B.13)} \\
&= \left(A_j^+\right)^{1/2}dB_t^{(j)} && \text{(B.14)}
\end{aligned}$$

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

$$\begin{aligned}
\frac{1}{\sqrt{K}}\left(\nabla_{\mathbf{X}}f\right)^{\mathrm{T}}\mathbf{D}d\mathbf{B}_t &= \frac{1}{\sqrt{K}}\sum_{j=1}^{m}\left(\nabla_{\mathbf{X}}f_i\right)_j\left(\mathbf{D}d\mathbf{B}_t\right)_j \\
&= \frac{1}{N\sqrt{K}}\sum_{j=1}^{m}\left(\delta_{ij}-p_i\right)\left(A_j^+\right)^{1/2}dB_t^{(j)} && \text{(B.15)} \\
&= \frac{1}{N\sqrt{K}}\left(A_i^+\right)^{1/2}dB_t^{(i)} - p_i\sum_{j=1}^{m}\left(A_j^+\right)^{1/2}dB_t^{(j)} && \text{(B.16)}
\end{aligned}$$

Putting equations (B.4), (B.12) and (B.16) into (B.1), 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:

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

Plugging the functional forms of (2.32) and the definitions of w_i and τ_i into the definitions of A_i^- and A_i^+, we obtain the relations

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

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

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

Where we have used the definition of p_i from (2.33). Now using the definition of mean fitness from (2.34) and rearranging terms gives us

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

where we have defined $Q_j(\mathbf{p}) = Q_j(\mathbf{x})/N(t)$. Repeating the exact same calculations for the A_i^+ terms in the second term of (B.17) now yields equation (2.35) 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:

$$\sigma_f^2 \coloneqq \overline{(f^2)} - (\overline{f})^2 \tag{C.1}$$

where $\overline{X}$ is the statistical mean value defined by (2.34). By the product rule, we have

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

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 (4.1) and use the same steps used in going from (2.37) to (2.41) to write

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

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 $dX_t = S(X_t)dt + \sum D_j(X_t)dB_t^{(j)}$ with S, D_j being suitable real functions and each $B_t^{(j)}$ being an independent Wiener process, Itô's formula says that given any $C^2(\mathbb{R})$ function $g(x)$, we have the relation:

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

In our case, we have a one-dimensional process for the mean value $d\overline{f}$ 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:

$$d(\overline{f}^2) = \left(2\overline{f}S(X_t) + \sum_j D_j^2(X_t)\right)dt + \sum_j 2\overline{f}D_j(X_t)dB_t^{(j)} \tag{C.5}$$

where the relevant functions S and D_j can be read off from (4.2). Since the dB terms are unwieldy, we will denote the contribution of all the dB_t terms collectively by $dB_{\sigma_f^2}$ 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:

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

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

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

Now, we note that

$$\frac{1}{N_K}A_i^+ = \frac{1}{N_K}\left(\tau_i x_i + \mu Q_i(\mathbf{x})\right) \tag{C.7}$$
$$= \tau_i p_i + \mu Q_i(\mathbf{p}) \tag{C.8}$$

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

$$\frac{1}{KN_K^2(t)}\left(\sum_{i=1}^{m}(f_i-\overline{f})^2 A_i^+\right)dt = \frac{1}{KN_K}\sum_{i=1}^{m}(f_i-\overline{f})^2\left(\tau_i p_i + \mu Q_i(\mathbf{p})\right) \tag{C.9}$$
$$= \frac{1}{KN_K}\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2\left(\tau_i p_i + \mu Q_i(\mathbf{p})\right) \tag{C.10}$$
$$= \frac{1}{KN_K}\left(\overline{\tau\left(f-\overline{f}\right)^2} + \mu\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2 Q_i(\mathbf{p})\right) \tag{C.11}$$
$$= \frac{1}{KN_K}\left(\mathrm{Cov}(\tau,\left(f-\overline{f}\right)^2) + \overline{\tau}\,\overline{\left(f-\overline{f}\right)^2} + \mu\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2 Q_i(\mathbf{p})\right) \tag{C.12}$$
$$= \frac{1}{KN_K}\left(\mathrm{Cov}(\tau,\left(f-\overline{f}\right)^2) + \overline{\tau}\sigma_f^2 + \mu\sum_{i=1}^{m}\left(f_i-\overline{f}\right)^2 Q_i(\mathbf{p})\right) \tag{C.13}$$

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

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

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

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

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

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

and similarly,

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

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:

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

This is precisely equation (4.8) in the main text. To calculate the mutation term, we substitute (C.13) into (C.6) to find

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

We can further simplify the first term of the RHS as

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

and similarly, the second term as

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

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

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

which after slight rearrangement becomes

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

which is equation (4.9) in the main text. For the dB terms, we can use equations (C.3) and (C.5) to calculate:

$$dB_{\sigma_f^2} = \frac{1}{\sqrt{K} N_K(t)} \left(\sum_{i=1}^{m} \left(f_i^2 - \overline{f^2} - 2\overline{f}(f_i - \overline{f}) \right) \sqrt{A_i^+} dB_t^{(i)} \right) \tag{C.20}$$

$$= \frac{1}{\sqrt{K} N_K(t)} \left(\sum_{i=1}^{m} \left(f_i^2 - \overline{f^2} - 2\overline{f} f_i - 2\overline{f}^2 \right) \sqrt{A_i^+} dB_t^{(i)} \right) \tag{C.21}$$

$$= \frac{1}{\sqrt{K} N_K(t)} \left(\sum_{i=1}^{m} \left(f_i - \overline{f} \right)^2 \sqrt{A_i^+} dB_t^{(i)} \right) \tag{C.22}$$

which is equation (4.10) in the main text.

Appendix D

Some Examples

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 $d(n) = \left(\mu + (\lambda - \mu)\frac{n}{K}\right) n$, where μ and K are positive constants. Thus, we have the equations

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

Here, K is the system-size parameter. Introducing the new variable $x = n/K$, we obtain

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

Thus, we have

$$A^{\pm}(x) = b_K(x) \pm d_K(x) = x\left(\lambda \pm ((\mu + (\lambda - \mu)x))\right)$$

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

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

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

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

showing that in the infinite population limit, we obtain the logistic equation. 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:

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

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

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

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

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

The time series predicted by these three processes look qualitatively similar and all seem to fluctuate about the deterministic steady state (Figure D.1).

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, 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) depend not only on r, but also on v, the sum of the birth and death rates. It has been proven that

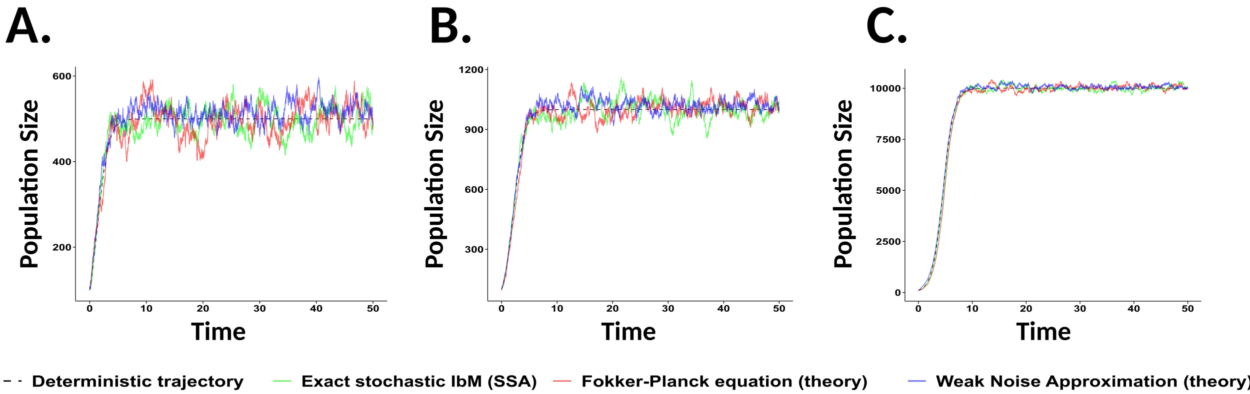

Figure D.1: Comparison of a single realization of the exact birth-death process (D.1), the deterministic trajectory (D.3), the non-linear Fokker-Planck equation (D.2), and the weak noise approximation (D.5) for **(A)** $K = 500$, **(B)** $K = 1000$, and **(C)** $K = 10000$. $\lambda = 2, \mu = 1$ for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

$X_t = 0$ is the only recurrent state for the full stochastic dynamics (D.2), meaning that every population is guaranteed to go extinct[1] 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 $t \to \infty$) of the exact birth-death process (D.1) is compared with that predicted by (D.2) and (D.5) for various values of K in figure D.2.

[1]This 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.

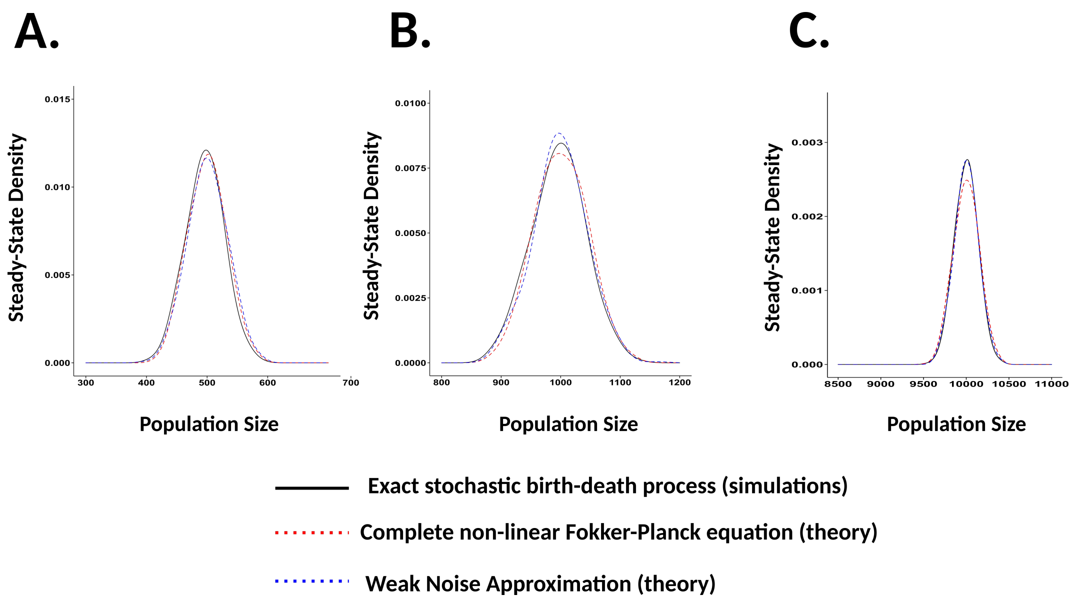

Figure D.2: Comparison of the steady-state densities given by (D.1), (D.2), and (D.5) for **(A)** $K = 500$, **(B)** $K = 1000$, and **(C)** $K = 10000$. $\lambda = 2, \mu = 1$ for all thee cases. Each curve was obtained using 1000 independent realizations.

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 $\mathbf{v}(t) = [v_1(t), v_2(t), \ldots, v_m(t)]$, where $v_i(t)$ is the number of type i individuals at time t. Let the birth and death rates of the ith type be given by:

$$\begin{aligned} b_i(\mathbf{v}) &= \left(\lambda + \frac{1}{K}\left(\sum_{j=1}^{m} \beta_{ij} v_j\right)\right) v_i \\ d_i(\mathbf{v}) &= \left(\mu + \frac{1}{K}\left(\sum_{j=1}^{m} \delta_{ij} v_j\right)\right) v_i \end{aligned} \tag{D.6}$$

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 $M_{ij} \coloneqq \beta_{ij} - \delta_{ij}$ 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 $\mathbf{v}$ to densities $\mathbf{x} = \mathbf{v}/K$, we obtain the birth and death rates:

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

Thus, we have

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

Defining $r \coloneqq \lambda - \mu$, $\nu \coloneqq \lambda + \mu$, and $T_{ij} \coloneqq \beta_{ij} + \delta_{ij}$, we see from equation (2.31) that the mesoscopic view is the m dimensional SDE given by

$$d\mathbf{X}_t = \mathbf{A}^{-}(\mathbf{X}_t)dt + \frac{1}{\sqrt{K}}\mathbf{D}(\mathbf{X}_t)d\mathbf{B}_t \tag{D.8}$$

where

$$\mathbf{A}^{-}{}_{i} = (\mathbf{X}_t)_i(r + \sum_{j=1}^{m} M_{ij}(\mathbf{X}_t)_j)$$

and

$$(\mathbf{D}\mathbf{D}^{\mathrm{T}})_i = (\mathbf{X}_t)_i(\nu + \sum_{j=1}^{m} T_{ij}(\mathbf{X}_t)_j)$$

From (2.36), we see that the deterministic limit is a set of m coupled ODEs given by

$$\frac{dx_i}{dt} = x_i \left(r + \sum_{j=1}^{m} M_{ij}x_j \right) \tag{D.9}$$

These are precisely the Lotka-Volterra equations for a system of m species. By matching the terms of (D.7) with those of (2.32), we can identify that we have $\mu = 0$ and

$$w_i(\mathbf{x}) = r + \sum_{j=1}^{m} M_{ij}x_j \tag{D.10}$$

If $p_i(t)$ is the frequency of type i individuals in the population at time t and $N_K(t) = \sum_i x_i(t)$, then the mean fitness is given by

$$\begin{aligned}
\overline{w}(t) &= \sum_{i=1}^{m} w_i p_i && \text{(D.11)}\\
&= \sum_{i=1}^{m} \left(r + \sum_{j=1}^{m} M_{ij}x_j \right) p_i && \text{(D.12)}\\
&= r + \sum_{i=1}^{m} p_i \left(\sum_{j=1}^{m} M_{ij}x_j \right) && \text{(D.13)}
\end{aligned}$$

where we have used the fact that $\sum_i p_i = 1$ in the last line. Using (2.37) to write down the equations for the frequencies p_i, we obtain

$$\frac{1}{N_K(t)} \frac{dp_i}{dt} = \left[(\mathbf{M}\mathbf{p})_i - \mathbf{p} \cdot \mathbf{M}\mathbf{p} \right] p_i \tag{D.14}$$

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.8) and the deterministic limit (D.9) 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 $x_1, \ldots, x_m$ to the variables $p_1, \ldots, p_{m-1}, N_K$, 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.

Let the solution to the equations (D.9) be given by $\mathbf{a}(t) = [a_1(t), \ldots, a_m(t)]$. For the weak noise approximation, we can Taylor expand $A_i^{\pm}$ and use (2.60) to compute the directional derivative as:

$$D_i = y_i w_i(\mathbf{a}) + a_i \sum_{k=1}^{m} y_k \left(\left.\frac{\partial w_i}{\partial x_k}\right|_{\mathbf{x}=\mathbf{a}(t)} \right) \tag{D.15}$$

$$= y_i w_i(\mathbf{a}) + a_i \sum_{k=1}^{m} y_k \left(\left.\frac{\partial}{\partial x_k}\left(r + \sum_{j=1}^{m} M_{ij} x_j\right)\right|_{\mathbf{x}=\mathbf{a}(t)} \right) \tag{D.16}$$

$$= y_i w_i(\mathbf{a}) + a_i \sum_{k=1}^{m} y_k M_{ik} \tag{D.17}$$

$$\Rightarrow D_i = y_i w_i(\mathbf{a}) + a_i w_i(\mathbf{y}) - r a_i \tag{D.18}$$

where we have used the fact that $w_i(\mathbf{y}) = r + \sum_{k=1}^{m} y_k M_{ik}$ (from (D.10)) in the last step. Thus, in the weak noise approximation of our process, the dynamics are given by

$$\mathbf{x}(t) = \mathbf{a}(t) + \frac{1}{\sqrt{K}} \mathbf{y}(t) \tag{D.19}$$

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

$$\frac{\partial P}{\partial t}(\mathbf{y}, t) = \sum_{i=1}^{m} \left(-\frac{\partial}{\partial y_i} \left\{ \left(y_i w_i(\mathbf{a}) + a_i w_i(\mathbf{y}) - r a_i \right) P(\mathbf{y}, t) \right\} + \frac{1}{2} \left(a_i \left(\nu + \sum_{j=1}^{m} T_{ij} a_j \right) \right) \frac{\partial^2}{\partial {y_i}^2} P(\mathbf{y}, t) \right) \tag{D.20}$$

Using (D.18) in (2.54), we see that the fluctuations are expected to evolve as:

$$\frac{d}{dt}\mathbb{E}[y_i] = w_i(\mathbf{a})\mathbb{E}[y_i] + a_i \sum_{k=1}^{m} M_{ik}\mathbb{E}[y_k] \tag{D.21}$$

or, in matrix form:

$$\frac{d}{dt}\begin{bmatrix}\mathbb{E}[y_1]\\ \mathbb{E}[y_2]\\ \vdots\\ \mathbb{E}[y_i]\\ \vdots\\ \mathbb{E}[y_m]\end{bmatrix} = \begin{bmatrix} (r+\sum_{j=1}^{m} M_{1j}a_j + a_1M_{11}) & a_1M_{12} & a_1M_{13} & \dots & \dots & \dots & a_1M_{1m}\\ a_2M_{21} & (r+\sum_{j=1}^{m} M_{2j}a_j + a_2M_{22}) & a_2M_{23} & \dots & \dots & \dots & a_2M_{2m}\\ \vdots & & \ddots & & & & \vdots\\ a_iM_{i1} & a_iM_{i2} & a_iM_{i3} & \dots & (r+\sum_{j=1}^{m} M_{ij}a_j + a_iM_{ii}) & \dots & a_iM_{im}\\ \vdots & & & & & \ddots & \vdots\\ a_mM_{m1} & a_mM_{m2} & a_mM_{m3} & \dots & \dots & \dots & (r+\sum_{j=1}^{m} M_{mj}a_j + a_mM_{mm}) \end{bmatrix}\begin{bmatrix}\mathbb{E}[y_1]\\ \mathbb{E}[y_2]\\ \vdots\\ \mathbb{E}[y_i]\\ \vdots\\ \mathbb{E}[y_m]\end{bmatrix} \tag{D.22}$$

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 Interlude: Detecting modes in quantitative trait distribution through Fourier analysis

In Chapter 3, we used various approximations to arrive at the linear functional Fokker-Planck equation

$$\frac{\partial P}{\partial t}(\zeta, t) = \int_{\mathcal{T}} \left(-\frac{\delta}{\delta\zeta(x)}\left\{\mathcal{D}_\zeta[\mathcal{A}^-](x)P(\zeta,t)\right\} + \frac{1}{2}\mathcal{A}^+(x|\psi)\frac{\delta^2}{\delta\zeta(x)^2}\{P(\zeta,t)\}\right)dx \tag{D.23}$$

for describing stochastic fluctuations ζ from the deterministic solution obtained by solving (3.8). 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

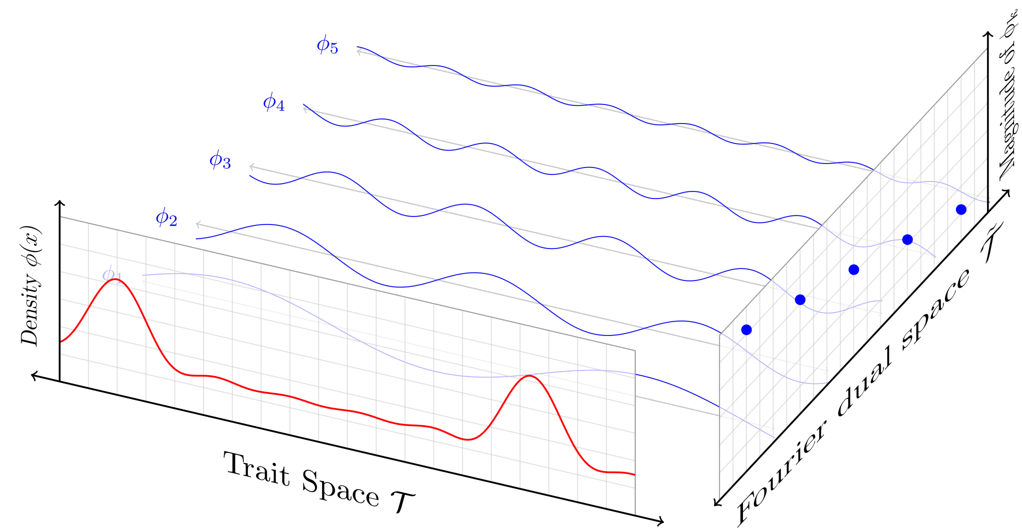

Figure D.3: Schematic description of Fourier analysis. A function $\phi(x)$ (shown in red) over the trait space can be decomposed as the sum of infinitely many Fourier modes (shown in blue) ϕ_k. In the Fourier dual space, we can look at the peaks of each of these Fourier modes: The magnitude of ϕ_k tells us how much it contributes to the actual function of interest ϕ.

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 by carrying out a basis expansion of our functions in the Fourier basis $\{e^{ikx}\}_{k\in\mathbb{Z}}$. If $\mathcal{D}_\zeta[\mathcal{A}^-]$ is a translation-invariant[2] linear operator, then $\exp(ikx)$ acts as an eigenfunction, significantly simplifying the calculations. We therefore assume this below.

Assume that $\mathcal{D}_\zeta[\mathcal{A}^-]$ takes the form:

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

[2] This 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 $T_c : \mathcal{F} \to \mathcal{F}$ be the translation operator on this space, defined by $T_c[f(x)] = f(x+c)$. An operator $L : \mathcal{F} \to \mathcal{F}$ is said to be translation-invariant if it commutes with T_c for every $c \in \mathbb{R}$, *i.e.* $T_c[L[f]] = L[T_c[f]]\ \forall\ f \in \mathcal{F}\ \forall\ c \in \mathbb{R}$

for a translation-invariant linear operator L that only depends on x and t. 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.

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

$$\begin{aligned}\zeta(x,t) &= \sum_{k=-\infty}^{\infty} e^{ikx}\zeta_k(t) \quad ; \quad \zeta_k(t) = \int_{\mathcal{T}} \zeta(x,t)e^{-ikx}dx \\ \mathcal{A}^+(x|\psi) &= \sum_{k=-\infty}^{\infty} e^{ikx}A_k(t) \quad ; \quad A_k(t) = \int_{\mathcal{T}} \mathcal{A}^+(x|\psi)e^{-ikx}dx\end{aligned} \tag{D.24}$$

In this case, the functional derivative operator obeys:

$$\frac{\delta}{\delta\zeta(x)} = \sum_{k=-\infty}^{\infty} e^{-ikx}\frac{\partial}{\partial\zeta_k} \tag{D.25}$$

and since L is linear and translation-invariant, we also have the relation[3]:

$$L[\zeta] = \sum_{k=-\infty}^{\infty} L_k\zeta_k e^{ikx} \tag{D.26}$$

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 $t > 0$:

$$\frac{\partial}{\partial\zeta_j(t)}F(\zeta_i(t)) = \delta_{ij}F'(\zeta_j(t)) \tag{D.27}$$

where δ_{ij} is the Kronecker delta symbol. Using (D.24), (D.25), and (D.26) in (3.27), we get, for the first term of the RHS:

$$-\int_{\mathcal{T}} \frac{\delta}{\delta\zeta(x)}\left\{L[\zeta(x,t)]P(\zeta,t)\right\}dx$$

[3]This 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 kth eigenvector $e_k = \exp(ikx)$.

$$
\begin{aligned}
&= -\int_{\mathcal{T}} \sum_k e^{-ikx} \frac{\partial}{\partial \zeta_k} \{\sum_n e^{inx} L_n \zeta_n P\} dx \\
&= -\int_{\mathcal{T}} \sum_k \sum_n e^{-i(k-n)x} \frac{\partial}{\partial \zeta_k} \{L_n \zeta_n P\} dx \\
&= -2\pi \sum_k L_k \frac{\partial}{\partial \zeta_k} \{\zeta_k P\}
\end{aligned}
\tag{D.28}
$$

and for the second:

$$
\begin{aligned}
&\int_{\mathcal{T}} \sum_k e^{ikx} A_k \left(\sum_m \sum_n e^{-i(m+n)x} \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} P \right) dx \\
&= \int_{\mathcal{T}} \sum_k \sum_m \sum_n e^{i(k-m-n)x} A_k \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{P\} dx \\
&= 2\pi \sum_m \sum_n A_{m+n} \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{P\}
\end{aligned}
\tag{D.29}
$$

Substituting (D.28) and (D.29) into (3.6), we see that the Fokker-Planck equation in Fourier space reads:

$$
\frac{\partial P}{\partial t} = -2\pi \sum_k L_k \frac{\partial}{\partial \zeta_k} \{\zeta_k P\} + \pi \sum_m \sum_n A_{m+n} \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} \{P\} \tag{D.30}
$$

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.30) by ζ_r and integrating over the probability space to obtain expectation values, we see that

$$
\begin{aligned}
\frac{d}{dt}\mathbb{E}[\zeta_r] &= -2\pi \sum_k \int \zeta_r L_k \frac{\partial}{\partial \zeta_k} \{\zeta_k P\} d\omega + \pi \sum_m \sum_n A_{m+n} \int \zeta_r \frac{\partial}{\partial \zeta_m} \frac{\partial}{\partial \zeta_n} (P) d\omega \\
&= 2\pi \sum_k L_k \int \zeta_k \frac{\partial \zeta_r}{\partial \zeta_k} P d\omega + \pi \sum_m \sum_n A_{m+n} \int \frac{\partial^2 \zeta_r}{\partial \zeta_m \partial \zeta_n} P d\omega \\
&= 2\pi L_r \mathbb{E}[\zeta_r]
\end{aligned}
\tag{D.31}
$$

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.27) to arrive at the final expression. Similarly, multiplying (D.30) by $\zeta_r \zeta_s$, integrating over the probability space and using integration by parts, we get:

$$\frac{d}{dt}\mathbb{E}[\zeta_r\zeta_s] = 2\pi\sum_k L_k \int \zeta_k P \frac{\partial}{\partial\zeta_k}\{\zeta_r\zeta_s\}d\omega + \pi\sum_m\sum_n A_{m+n}\int_{-\infty}^{\infty} P\frac{\partial}{\partial\zeta_m}\frac{\partial}{\partial\zeta_n}\{\zeta_r\zeta_s\}d\omega$$
$$= 2\pi(L_r + L_s)\mathbb{E}[\zeta_r\zeta_s] + \pi(A_{2r} + A_{2s}) \tag{D.32}$$

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

$$\mathbb{E}[\zeta_r\zeta_s] = -\frac{A_{2r} + A_{2s}}{2(L_r + L_s)} \tag{D.33}$$

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.33) to obtain the autocovariance relation:

$$\mathbb{E}[|\zeta_r|^2] = -\frac{\mathrm{Re}(A_{2r})}{2\mathrm{Re}(L_r)} \tag{D.34}$$

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

$$\Xi[x] = m(\mathcal{T})\int_{\mathcal{T}} \mathbb{E}[\phi_\infty(x)\phi_\infty(y-x)]dy \tag{D.35}$$

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:

$$\Xi[x] = m(\mathcal{T})\left[\int_{\mathcal{T}} \psi_\infty(x)\psi_\infty(y-x)dy + \frac{1}{K}\sum_{r=-\infty}^{\infty}\mathbb{E}[|\zeta_r|^2]e^{irx}\right] \tag{D.36}$$

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.4 An example for quantitative traits: The quantitative logistic equation

Recall the birth and death functionals given by (3.2). That is, the functionals

$$
\begin{aligned}
b(x|\nu) &= r \int\limits_{\mathcal{T}} m(x,y)\nu(y)dy;\ m(x,y) = \exp\left(\frac{-(x-y)^2}{\sigma_m^2}\right) \\
d(x|\nu) &= \frac{\nu(x)}{Kn(x)} \int\limits_{\mathcal{T}} \alpha(x,y)\nu(y)dy;\ \alpha(x,y) = \exp\left(\frac{-(x-y)^2}{\sigma_\alpha^2}\right)
\end{aligned}
\tag{D.37}
$$

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:

$$
\begin{aligned}
b_K(x|\phi) &= \frac{1}{K} b(x|\nu) = \frac{1}{K}\left(r \int\limits_{\mathcal{T}} m(x,y)K\phi(y)dy\right) \\
d_K(x|\phi) &= \frac{1}{K} d(x|\nu) = \frac{1}{K}\left(\frac{K\phi(x)}{Kn(x)} \int\limits_{\mathcal{T}} \alpha(x,y)K\phi(y)dy\right)
\end{aligned}
\tag{D.38}
$$

Thus, using equation (3.8), the deterministic trajectory becomes:

$$
\frac{\partial \psi}{\partial t}(x,t) = r \int\limits_{\mathcal{T}} m(x,y)\psi(y,t)dy - \frac{1}{n(x)}\psi(x,t) \int\limits_{\mathcal{T}} \alpha(x,y)\psi(y,t)dy \tag{D.39}
$$

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:

$$
\frac{\partial \Psi}{\partial t}(x,t) = r \int\limits_{\mathcal{T}} m(x,y)\Psi(y,t)dy - \frac{\Psi(x,t)}{Kn(x)} \int\limits_{\mathcal{T}} \alpha(x,y)\Psi(y,t)dy
$$

$$\approx r\Psi(x,t) - \frac{\Psi(x,t)}{K(x)} \int_{\mathcal{T}} \alpha(x,y)\Psi(y,t)dy + D_m \nabla_x^2 \Psi(x,t)$$

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. 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 (3.2)) and the death rate functional $d(x|\phi) = \phi(x)\left(\mu + (\lambda - \mu)\phi(x)/K\right)$, the macroscopic limit yields the famous Fisher-KPP equation with growth rate $r = \lambda - \mu$ and diffusion constant $D = \lambda\sigma_m^2/2$.

In any case, for the system defined by (D.38), we can also calculate $\mathcal{D}_\zeta[\mathcal{A}^-]$ as

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

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

References

Abrams, Peter A. et al. (1993).
On the Relationship between Quantitative Genetic and Ess Models.
Evolution *47*.3. Publisher: [Society for the Study of Evolution, Wiley], 982–985.
ISSN: 0014-3820. DOI: 10.2307/2410204.
URL: https://www.jstor.org/stable/2410204 (visited on 02/27/2023).

Baez, John C. (Nov. 2021). The Fundamental Theorem of Natural Selection. Entropy *23*.11. Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, 1436.
ISSN: 1099-4300. DOI: 10.3390/e23111436.
URL: https://www.mdpi.com/1099-4300/23/11/1436 (visited on 02/09/2023).

Black, Andrew J. and McKane, Alan J. (June 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.
URL: https://linkinghub.elsevier.com/retrieve/pii/S016953471200033X
(visited on 08/26/2022).

Boettiger, Carl (2018). From noise to knowledge: how randomness generates novel phenomena and reveals information. Ecology Letters *21*.8. _eprint:
https://onlinelibrary.wiley.com/doi/pdf/10.1111/ele.13085, 1255–1267. ISSN: 1461-0248.
DOI: 10.1111/ele.13085.
URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/ele.13085 (visited on 01/01/2023).

Bressloff, Paul C. (Jan. 2010).
Stochastic Neural Field Theory and the System-Size Expansion.
SIAM Journal on Applied Mathematics *70*.5. Publisher: Society for Industrial and Applied Mathematics, 1488–1521. ISSN: 0036-1399. DOI: 10.1137/090756971.
URL: https://epubs.siam.org/doi/abs/10.1137/090756971 (visited on 11/05/2022).

Brown, Joel S. (Sept. 14, 2016). Why Darwin would have loved evolutionary game theory.
Proceedings of the Royal Society B: Biological Sciences *283*.1838. Publisher: Royal Society, 20160847. DOI: 10.1098/rspb.2016.0847.
URL: https://royalsocietypublishing.org/doi/10.1098/rspb.2016.0847 (visited on 02/28/2023).

Champagnat, Nicolas et al. (May 1, 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.
URL: https://www.sciencedirect.com/science/article/pii/S0040580905001632 (visited on 12/07/2022).

— (Nov. 5, 2008).
From Individual Stochastic Processes to Macroscopic Models in Adaptive Evolution.
Stochastic Models *24* (sup1). Publisher: Taylor & Francis _eprint:
https://doi.org/10.1080/15326340802437710, 2–44. ISSN: 1532-6349.
DOI: 10.1080/15326340802437710.
URL: https://doi.org/10.1080/15326340802437710 (visited on 12/25/2022).

Chow, Carson C. and Buice, Michael A. (Dec. 2015).
Path Integral Methods for Stochastic Differential Equations. The Journal of Mathematical Neuroscience (JMN) *5*.1. Number: 1 Publisher: SpringerOpen, 1–35.
ISSN: 2190-8567. DOI: 10.1186/s13408-015-0018-5. URL: https://mathematical-neuroscience.springeropen.com/articles/10.1186/s13408-015-0018-5 (visited on 03/01/2023).

Claessen, David et al. (2007). Delayed evolutionary branching in small populations.
Evolutionary Ecology Research *9*.1. Publisher: Evolutionary Ecology, Ltd., 51–69.
ISSN: 1522-0613.
URL: https://www.evolutionary-ecology.com/abstracts/v09/2073.html (visited on 07/14/2022).

Constable, George W. A. and McKane, Alan J. (Aug. 29, 2017). Mapping of the stochastic Lotka-Volterra model to models of population genetics and game theory. Physical Review E *96*.2. Publisher: American Physical Society, 022416. DOI: 10.1103/PhysRevE.96.022416. URL: https://link.aps.org/doi/10.1103/PhysRevE.96.022416 (visited on 01/24/2023).

Constable, George W. A. et al. (June 2013). Stochastic dynamics on slow manifolds. Journal of Physics A: Mathematical and Theoretical *46*.29. Publisher: IOP Publishing, 295002. ISSN: 1751-8121. DOI: 10.1088/1751-8113/46/29/295002. URL: https://dx.doi.org/10.1088/1751-8113/46/29/295002 (visited on 01/22/2023).

Constable, George W. A. et al. (Aug. 9, 2016). Demographic noise can reverse the direction of deterministic selection. Proceedings of the National Academy of Sciences *113*.32. Publisher: Proceedings of the National Academy of Sciences, E4745–E4754. DOI: 10.1073/pnas.1603693113. URL: https://www.pnas.org/doi/full/10.1073/pnas.1603693113 (visited on 07/07/2022).

Coulson, Tim et al. (July 1, 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. URL: https://www.sciencedirect.com/science/article/pii/S0169534704001466 (visited on 01/06/2023).

Coulson, Tim et al. (2006). Putting evolutionary biology back in the ecological theatre: a demographic framework mapping genes to communities. Evolutionary Ecology Research *8*.7. Publisher: Evolutionary Ecology, Ltd., 1155–1171. ISSN: 1522-0613. URL: https://www.evolutionary-ecology.com/abstracts/v08/2021.html (visited on 02/28/2023).

Crow, James F. and Kimura, Motoo (1970). An introduction to population genetics theory Harper international editions. OCLC: 91086. (New York: Harper & Row). 591 pp.

Débarre, Florence and Otto, Sarah P. (Apr. 1, 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.

URL: https://www.sciencedirect.com/science/article/pii/S0040580915001240 (visited on 07/06/2022).

DeLong, John P. and Cressler, Clayton E. (2023).
Stochasticity directs adaptive evolution toward nonequilibrium evolutionary attractors. Ecology *104*.1. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.3873, e3873. ISSN: 1939-9170. DOI: 10.1002/ecy.3873.
URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.3873 (visited on 01/06/2023).

Dieckmann, Ulf and Law, Richard (May 1, 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.
URL: https://doi.org/10.1007/BF02409751 (visited on 07/13/2022).

Dodd, Peter J. and Ferguson, Neil M. (Sept. 1, 2009).
A Many-Body Field Theory Approach to Stochastic Models in Population Biology. PLOS ONE *4*.9. Publisher: Public Library of Science, e6855. ISSN: 1932-6203.
DOI: 10.1371/journal.pone.0006855. URL:
https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006855 (visited on 03/01/2023).

Doebeli, Michael (2011). Adaptive diversification OCLC: 687685579.
(Princeton, N.J.: Princeton University Press). ISBN: 978-0-691-12894-8.

Doebeli, Michael et al. (Feb. 15, 2017).
Towards a mechanistic foundation of evolutionary theory.
eLife *6*. Wenying Shou, ed. Publisher: eLife Sciences Publications, Ltd, e23804.
ISSN: 2050-084X. DOI: 10.7554/eLife.23804.
URL: https://doi.org/10.7554/eLife.23804 (visited on 02/28/2023).

Doi, M. (Sept. 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.
URL: https://dx.doi.org/10.1088/0305-4470/9/9/008 (visited on 01/21/2023).

Durrett, R. and Levin, S. (Dec. 1, 1994). The Importance of Being Discrete (and Spatial). Theoretical Population Biology *46*.3, 363–394. ISSN: 0040-5809.
DOI: 10.1006/tpbi.1994.1032.

URL: https://www.sciencedirect.com/science/article/pii/S004058098471032X (visited on 03/01/2023).

Etheridge, Alison (2011). Some mathematical models from population genetics: École d'été de Probabilités de Saint-Flour XXXIX-2009 Lecture Notes in Mathematics. OCLC: 920323591. (Berlin: Springer-Verlag). 119 pp. ISBN: 978-3-642-16631-0.

Ewens, Warren J. (Jan. 9, 2004). Mathematical Population Genetics 1: Theoretical Introduction (Springer Science & Business Media). 448 pp. ISBN: 978-0-387-20191-7.

Fisher, Ronald Aylmer (1930). The genetical theory of natural selection OCLC: 18500548. (Oxford: The Clarendon Press). 272 pp. URL: http://link.library.utoronto.ca/eir/EIRdetail.cfm?Resources__ID=538875&T=F (visited on 02/27/2023).

Frank, Steven A. (2012). Natural selection. IV. The Price equation*. Journal of Evolutionary Biology *25*.6. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1420-9101.2012.02498.x, 1002–1019. ISSN: 1420-9101. DOI: 10.1111/j.1420-9101.2012.02498.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1420-9101.2012.02498.x (visited on 02/26/2023).

Frank, Steven A. and Slatkin, Montgomery (Mar. 1, 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. URL: https://www.sciencedirect.com/science/article/pii/016953479290248A (visited on 02/10/2023).

Gardiner, C. W (2009). Stochastic methods: a handbook for the natural and social sciences OCLC: 277069162. (Berlin: Springer). ISBN: 978-3-540-70712-7.

Geritz, Stefan A. H. and Kisdi, Éva (Dec. 1, 2012). Mathematical ecology: why mechanistic models? Journal of Mathematical Biology *65*.6, 1411–1415. ISSN: 1432-1416. DOI: 10.1007/s00285-011-0496-3. URL: https://doi.org/10.1007/s00285-011-0496-3 (visited on 11/02/2022).

Hastings, Alan (Jan. 1, 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.

URL: https://www.sciencedirect.com/science/article/pii/S0169534703002854 (visited on 01/04/2023).

Hofbauer, Josef and Sigmund, Karl (1998). Evolutionary games and population dynamics OCLC: 800133814. (Cambridge: Cambridge University Press). 323 pp. ISBN: 978-0-521-62570-8.

Houchmandzadeh, Bahram (Jan. 1, 2015). Fluctuation driven fixation of cooperative behavior. Biosystems *127*, 60–66. ISSN: 0303-2647. DOI: 10.1016/j.biosystems.2014.11.006. URL: https://www.sciencedirect.com/science/article/pii/S0303264714001956 (visited on 01/26/2023).

Houchmandzadeh, Bahram and Vallade, Marcel (May 10, 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. URL: https://doi.org/10.1186/1471-2148-12-61 (visited on 01/26/2023).

Hutchinson, G. Evelyn (1965). The ecological theater and the evolutionary play OCLC: 250039. (New Haven: Yale University Press). 139 pp.

Johansson, Jacob and Ripa, Jörgen (Oct. 2006). Will Sympatric Speciation Fail due to Stochastic Competitive Exclusion? The American Naturalist *168*.4. Publisher: The University of Chicago Press, 572–578. ISSN: 0003-0147. DOI: 10.1086/507996. URL: https://www.journals.uchicago.edu/doi/10.1086/507996 (visited on 07/29/2022).

Johnson, Brian et al. (Apr. 21, 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. Publisher: Royal Society, 210182. DOI: 10.1098/rsos.210182. URL: https://royalsocietypublishing.org/doi/full/10.1098/rsos.210182 (visited on 10/20/2022).

Karatzas, Ioannis and Shreve, Steven E. (1998). Brownian motion and stochastic calculus 2nd ed. Graduate texts in mathematics. OCLC: 70184596. (New York: Springer-Verlag). 470 pp. ISBN: 978-0-387-97655-6. URL: http://www.gbv.de/dms/bowker/toc/9780387976556.pdf (visited on 02/25/2023).

Kingsland, Sharon E. (1985). Modeling nature: episodes in the history of population ecology Science and its conceptual foundations. OCLC: 11755083.

(Chicago: University of Chicago Press). 267 pp. ISBN: 978-0-226-43726-2. URL: http://www.gbv.de/dms/bs/toc/016547187.pdf (visited on 01/10/2023).

Kokko, Hanna (Nov. 17, 2021). The stagnation paradox: the ever-improving but (more or less) stationary population fitness. Proceedings of the Royal Society B: Biological Sciences *288*.1963. Publisher: Royal Society, 20212145. DOI: 10.1098/rspb.2021.2145. URL: https://royalsocietypublishing.org/doi/full/10.1098/rspb.2021.2145 (visited on 01/27/2023).

Kokko, Hanna et al. (Mar. 1, 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. URL: https://www.sciencedirect.com/science/article/pii/S0169534716302385 (visited on 02/28/2023).

Kuosmanen, Teemu et al. (July 12, 2022). *Turnover shapes evolution of birth and death rates*. Pages: 2022.07.11.499527 Section: New Results. DOI: 10.1101/2022.07.11.499527. URL: https://www.biorxiv.org/content/10.1101/2022.07.11.499527v1 (visited on 02/10/2023).

Lafuerza, Luis F. and McKane, Alan J. (Mar. 11, 2016). Role of demographic stochasticity in a speciation model with sexual reproduction. Physical Review E *93*.3. Publisher: American Physical Society, 032121. DOI: 10.1103/PhysRevE.93.032121. URL: https://link.aps.org/doi/10.1103/PhysRevE.93.032121 (visited on 07/22/2022).

Lande, Russell (1976). Natural Selection and Random Genetic Drift in Phenotypic Evolution. Evolution *30*.2. Publisher: [Society for the Study of Evolution, Wiley], 314–334. ISSN: 0014-3820. DOI: 10.2307/2407703. URL: https://www.jstor.org/stable/2407703 (visited on 02/18/2023).

Lehtonen, Jussi (Jan. 2018). The Price Equation, Gradient Dynamics, and Continuous Trait Game Theory. The American Naturalist *191*.1. Publisher: The University of Chicago Press, 146–153. ISSN: 0003-0147. DOI: 10.1086/694891. URL: https://www.journals.uchicago.edu/doi/full/10.1086/694891 (visited on 02/03/2023).

Lehtonen, Jussi (Mar. 9, 2020).
The Price equation and the unity of social evolution theory.
Philosophical Transactions of the Royal Society B: Biological Sciences *375*.1797.
Publisher: Royal Society, 20190362. DOI: 10.1098/rstb.2019.0362.
URL: https://royalsocietypublishing.org/doi/full/10.1098/rstb.2019.0362
(visited on 02/03/2023).

Lion, Sébastien (Jan. 2018). Theoretical Approaches in Evolutionary Ecology:
Environmental Feedback as a Unifying Perspective.
The American Naturalist *191*.1. Publisher: The University of Chicago Press, 21–44.
ISSN: 0003-0147. DOI: 10.1086/694865.
URL: https://www.journals.uchicago.edu/doi/full/10.1086/694865 (visited on
01/27/2023).

Lion, Sébastien et al. (Sept. 2022).
Multimorph Eco-Evolutionary Dynamics in Structured Populations.
The American Naturalist *200*.3. Publisher: The University of Chicago Press, 345–372.
ISSN: 0003-0147. DOI: 10.1086/720439. URL:
https://www.journals.uchicago.edu/doi/10.1086/720439 (visited on 02/14/2023).

Luque, Victor J. and Baravalle, Lorenzo (Dec. 1, 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.
URL: https://doi.org/10.1007/s11229-021-03339-6 (visited on 01/27/2023).

Mazzolini, Andrea and Grilli, Jacopo (May 6, 2022).
Universality of evolutionary trajectories under arbitrary competition dynamics.
Pages: 2021.06.17.448795 Section: New Results. DOI: 10.1101/2021.06.17.448795.
URL: https://www.biorxiv.org/content/10.1101/2021.06.17.448795v3 (visited on
02/14/2023).

McLeod, David V. and Day, Troy (Feb. 4, 2019).
Social evolution under demographic stochasticity.
PLOS Computational Biology *15*.2. Publisher: Public Library of Science, e1006739.
ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1006739. URL: https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1006739
(visited on 07/16/2022).

Melbinger, Anna et al. (Oct. 18, 2010).
Evolutionary Game Theory in Growing Populations.

Physical Review Letters *105*.17. Publisher: American Physical Society, 178101.
DOI: 10.1103/PhysRevLett.105.178101.
URL: https://link.aps.org/doi/10.1103/PhysRevLett.105.178101 (visited on 02/05/2023).

Metcalf, C. Jessica E. and Pavard, Samuel (Apr. 1, 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.
URL: https://www.sciencedirect.com/science/article/pii/S0169534706003806 (visited on 01/09/2023).

— (Nov. 1, 2007b). All paths to fitness lead through demography.
Trends in Ecology & Evolution *22*.11. Publisher: Elsevier, 563–564. ISSN: 0169-5347.
DOI: 10.1016/j.tree.2007.07.003. URL: https://www.cell.com/trends/ecology-evolution/abstract/S0169-5347(07)00255-8 (visited on 02/28/2023).

Nåsell, INGEMAR (July 7, 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.
URL: https://www.sciencedirect.com/science/article/pii/S0022519301923288 (visited on 01/04/2023).

Nordbotten, Jan Martin et al. (Aug. 19, 2020).
The dynamics of trait variance in multi-species communities.
Royal Society Open Science *7*.8. Publisher: Royal Society, 200321.
DOI: 10.1098/rsos.200321.
URL: https://royalsocietypublishing.org/doi/10.1098/rsos.200321 (visited on 02/14/2023).

Ó Laighléis, E. et al. (Dec. 18, 2018).
Minimal stochastic field equations for one-dimensional flocking.
Physical Review E *98*.6. Publisher: American Physical Society, 062127.
DOI: 10.1103/PhysRevE.98.062127. URL:
https://link.aps.org/doi/10.1103/PhysRevE.98.062127 (visited on 01/15/2023).

Øksendal, B. K (1998). Stochastic differential equations: an introduction with applications OCLC: 680820282. (Berlin; New York: Springer). ISBN: 978-3-662-03620-4.
URL: http://books.google.com/books?id=XGIPAQAAMAAJ (visited on 08/04/2022).

Page, Karen M. and Nowak, Martin A. (Nov. 7, 2002). Unifying Evolutionary Dynamics. Journal of Theoretical Biology *219*.1, 93–98. ISSN: 0022-5193. DOI: 10.1006/jtbi.2002.3112. URL: https://www.sciencedirect.com/science/article/pii/S0022519302931127 (visited on 12/07/2022).

Parsons, Todd L. and Rogers, Tim (Sept. 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. Publisher: IOP Publishing, 415601. ISSN: 1751-8121. DOI: 10.1088/1751-8121/aa86c7. URL: https://dx.doi.org/10.1088/1751-8121/aa86c7 (visited on 01/25/2023).

Peliti, L. (Sept. 1, 1985). Path integral approach to birth-death processes on a lattice. Journal de Physique *46*.9. Publisher: Société Française de Physique, 1469–1483. ISSN: 0302-0738, 2777-3396. DOI: 10.1051/jphys:019850046090146900. URL: http://dx.doi.org/10.1051/jphys:019850046090146900 (visited on 01/21/2023).

Potochnik, Angela (2018). Idealization and the aims of science OCLC: 1101830222. (). ISBN: 978-0-226-50705-7.

Price, George R. (1972). Fisher's 'fundamental theorem' made clear. Annals of Human Genetics *36*.2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1469-1809.1972.tb00764.x, 129–140. ISSN: 1469-1809. DOI: 10.1111/j.1469-1809.1972.tb00764.x. URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1469-1809.1972.tb00764.x (visited on 02/09/2023).

Proulx, S. R. and Day, Troy (July 22, 2005). What can Invasion Analyses Tell us about Evolution under Stochasticity in Finite Populations? Selection *2*.1. Publisher: Akadémiai Kiadó Section: Selection, 2–15. ISSN: 1585-1931, 1588-287X. DOI: 10.1556/select.2.2001.1-2.2. URL: https://akjournals.com/view/journals/076/2/1-2/article-p2.xml (visited on 01/25/2023).

Queller, David C. (Apr. 2017). Fundamental Theorems of Evolution. The American Naturalist *189*.4. Publisher: The University of Chicago Press, 345–353. ISSN: 0003-0147. DOI: 10.1086/690937. URL: https://www.journals.uchicago.edu/doi/full/10.1086/690937 (visited on 01/27/2023).

Rice, Sean H. (Sept. 25, 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. URL: https://doi.org/10.1186/1471-2148-8-262 (visited on 02/14/2023).

— (Mar. 9, 2020). Universal rules for the interaction of selection and transmission in evolution. Philosophical Transactions of the Royal Society B: Biological Sciences *375*.1797. Publisher: Royal Society, 20190353. DOI: 10.1098/rstb.2019.0353. URL: https://royalsocietypublishing.org/doi/10.1098/rstb.2019.0353 (visited on 01/29/2023).

Rogers, Tim and McKane, Alan J. (Sept. 14, 2015). Modes of competition and the fitness of evolved populations. Physical Review E *92*.3. Publisher: American Physical Society, 032708. DOI: 10.1103/PhysRevE.92.032708. URL: https://link.aps.org/doi/10.1103/PhysRevE.92.032708 (visited on 08/31/2022).

Rogers, Tim et al. (Feb. 2012a). Demographic noise can lead to the spontaneous formation of species. Europhysics Letters *97*.4. Publisher: IOP Publishing, 40008. ISSN: 0295-5075. DOI: 10.1209/0295-5075/97/40008. URL: https://doi.org/10.1209/0295-5075/97/40008 (visited on 07/22/2022).

Rogers, Tim et al. (Oct. 2012b). Spontaneous genetic clustering in populations of competing organisms. Physical Biology *9*.6. Publisher: IOP Publishing, 066002. ISSN: 1478-3975. DOI: 10.1088/1478-3975/9/6/066002. URL: https://doi.org/10.1088/1478-3975/9/6/066002 (visited on 07/25/2022).

Sasaki, Akira and Dieckmann, Ulf (Oct. 1, 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. URL: https://doi.org/10.1007/s00285-010-0380-6 (visited on 01/01/2023).

Schreiber, Sebastian J. et al. (2022). Does deterministic coexistence theory matter in a finite world? Ecology. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.3838, e3838. ISSN: 1939-9170. DOI: 10.1002/ecy.3838.

URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.3838 (visited on 11/02/2022).

Shoemaker, Lauren G. et al. (2020).
Integrating the underlying structure of stochasticity into community ecology. Ecology *101*.2. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ecy.2922, e02922.
ISSN: 1939-9170. DOI: 10.1002/ecy.2922.
URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/ecy.2922 (visited on 09/07/2022).

Strang, Alexander G. et al. (Dec. 1, 2019). How to avoid an extinction time paradox. Theoretical Ecology *12*.4, 467–487. ISSN: 1874-1746.
DOI: 10.1007/s12080-019-0416-5.
URL: https://doi.org/10.1007/s12080-019-0416-5 (visited on 02/09/2023).

Tao, Yi and Cressman, Ross (May 1, 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.
URL: https://doi.org/10.1007/s11538-006-9170-0 (visited on 01/04/2023).

Van Kampen, N. G. (1981). Stochastic processes in physics and chemistry
Open Library ID: OL4265375M. (Amsterdam, New York, New York: North-Holland). xiv, 419. ISBN: 978-0-444-86200-6.

Vellend, Mark (Aug. 23, 2016). The Theory of Ecological Communities
Publication Title: The Theory of Ecological Communities (MPB-57).
(Princeton University Press). ISBN: 978-1-4008-8379-0.
URL: https://www.degruyter.com/document/doi/10.1515/9781400883790/html (visited on 08/30/2021).

Veller, Carl et al. (Oct. 1, 2017).
Drift-Induced Selection Between Male and Female Heterogamety.
Genetics *207*.2, 711–727. ISSN: 1943-2631. DOI: 10.1534/genetics.117.300151.
URL: https://academic.oup.com/genetics/article/207/2/711/5930680 (visited on 03/01/2023).

Wakano, Joe Yuichiro and Iwasa, Yoh (Jan. 1, 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.
URL: https://doi.org/10.1534/genetics.112.144980 (visited on 07/08/2022).

Weber, Markus F. and Frey, Erwin (Mar. 2017).
Master equations and the theory of stochastic path integrals.
Reports on Progress in Physics *80*.4. Publisher: IOP Publishing, 046601.
ISSN: 0034-4885. DOI: 10.1088/1361-6633/aa5ae2.
URL: https://dx.doi.org/10.1088/1361-6633/aa5ae2 (visited on 03/01/2023).

Week, Bob et al. (July 21, 2021). A white noise approach to evolutionary ecology.
Journal of Theoretical Biology *521*, 110660. ISSN: 0022-5193.
DOI: 10.1016/j.jtbi.2021.110660.
URL: https://www.sciencedirect.com/science/article/pii/S0022519321000825
(visited on 07/16/2022).

Wickman, Jonas et al. (Nov. 9, 2022).
A theoretical framework for trait-based eco-evolutionary dynamics: population structure, intraspecific variation, and community assembly.
The American Naturalist. Publisher: The University of Chicago Press. ISSN: 0003-0147.
DOI: 10.1086/723406.
URL: https://www.journals.uchicago.edu/doi/abs/10.1086/723406 (visited on 01/01/2023).

Zuk, Marlene and Travisano, Mike (July 1, 2018).
Models on the Runway: How Do We Make Replicas of the World?
The American Naturalist *192*.1. Publisher: The University of Chicago Press, 1–9.
ISSN: 0003-0147. DOI: 10.1086/697508.
URL: https://www.journals.uchicago.edu/doi/full/10.1086/697508 (visited on 01/19/2022).