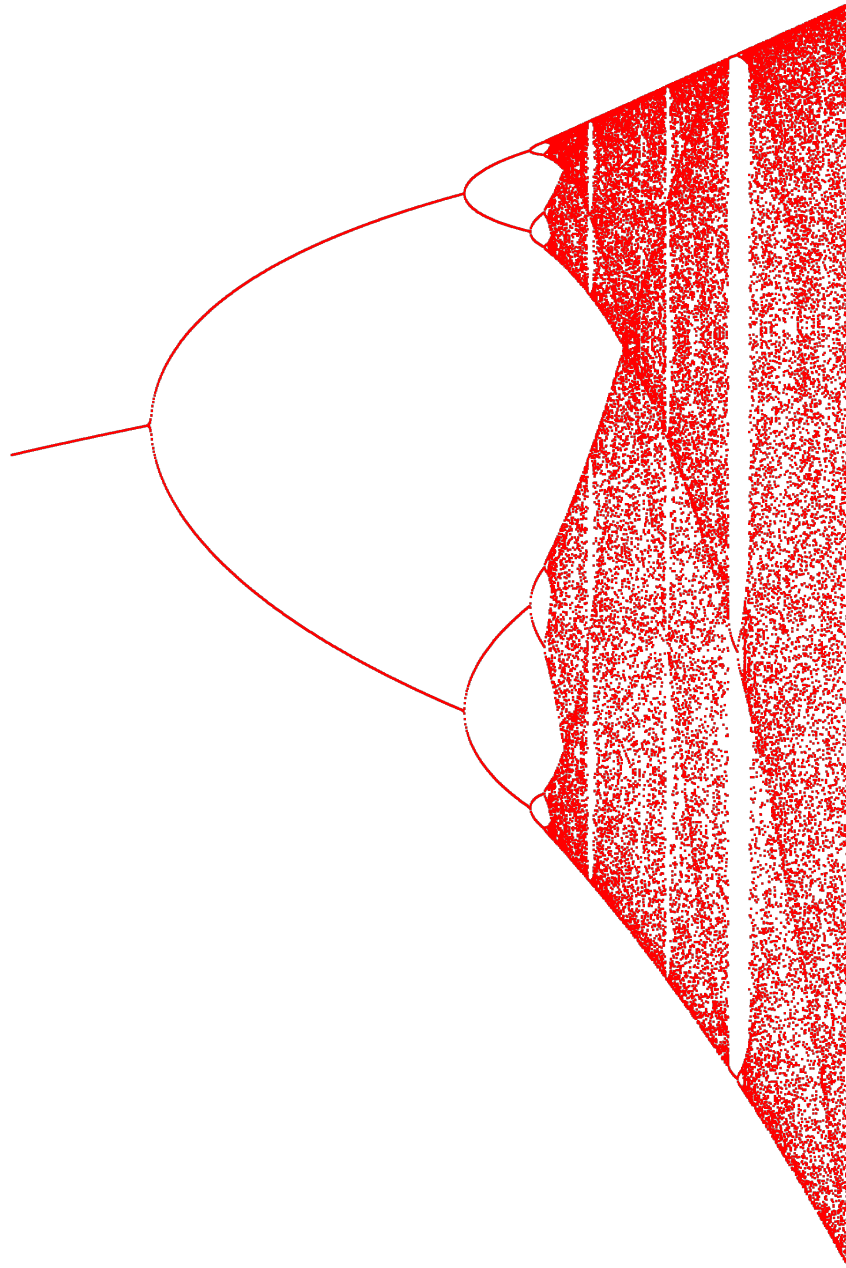


# An EEB theory Napkin

An attempt to construct Evan Chen's **Infinitely Large Napkin** for  
theoretical population biology



Ananda Shikhara Bhat

# Contents

<b>I</b>	<b>Evolution in finite populations</b>	<b>4</b>
<b>1</b>	<b>Demographic roots: birth-death processes for finite populations</b>	<b>5</b>
1.1	Birth-death processes . . . . .	5
1.2	Stochastic Differential Equations . . . . .	6
1.3	Fokker-Planck Equations . . . . .	9
1.3.1	Density-dependence and the intuition for system-size expansions via ecology . . . . .	11
1.3.2	The intuition for the weak noise approximation in ecology . . . . .	12
1.4	Stochastic equations for population size dynamics . . . . .	12
1.4.1	Description of the process and the Master Equation . . . . .	13
1.4.2	The system-size expansion . . . . .	15
1.4.3	Stochastic fluctuations and the weak noise approximation . . . . .	16
1.5	An example: The stochastic logistic equation . . . . .	19
<b>II</b>	<b>Spatial ecology and random walks</b>	<b>23</b>
<b>2</b>	<b>Random walks</b>	<b>24</b>
2.1	A simple random walk . . . . .	24
2.1.1	The mean square displacement . . . . .	26
2.1.2	A general expression for the MSD . . . . .	26
2.1.3	Dimensionality considerations . . . . .	27
2.2	Introducing bias . . . . .	28
2.3	Rate of dispersion . . . . .	29
<b>3</b>	<b>The Diffusion Equation</b>	<b>30</b>
3.1	Typical Behavior of solutions to the diffusion equation . . . . .	31
3.2	Introducing bias: The drift-diffusion equation . . . . .	32
3.3	Typical behavior of solutions to the drift-diffusion equation . . . . .	32
<b>4</b>	<b>Turing Patterns</b>	<b>33</b>
4.1	Linear stability analysis in the mean-field case . . . . .	33
4.2	Linear stability analysis when diffusion is present . . . . .	34

4.3	Stability criteria for a homogeneous perturbation . . . . .	35
4.4	Stability criteria for a periodic perturbation . . . . .	35
4.4.1	The Trace . . . . .	36
4.4.2	The Determinant . . . . .	36
4.4.3	Identifying sufficient criteria for stability . . . . .	37
4.4.4	General criteria for Turing instability (I think?) . . . . .	37
<b>III</b>	<b>Evolutionary Game theory and adaptive dynamics</b>	<b>39</b>
<b>5</b>	<b>Evolutionary game theory for mathematicians</b>	<b>40</b>
5.1	Games, strategies, and populations . . . . .	40
5.2	Nash Equilibria, ESSs, and associated results . . . . .	42
5.3	Replicator dynamics and the folk theorem . . . . .	44
5.4	An example: The ideal free distribution as an evolutionary game . . . . .	45
5.4.1	The game-theoretic perspective: A habitat selection game . . . . .	45
<b>6</b>	<b>Adaptive Diversification</b>	<b>50</b>
6.1	The ‘classic’ adaptive dynamics approach . . . . .	50
6.2	The PDE approach . . . . .	53
6.3	Stochastic individual-based models . . . . .	54
6.4	An example of a prediction of adaptive diversification . . . . .	54
6.4.1	Habitat Filtering . . . . .	55
6.4.2	Competition and character displacement . . . . .	55
6.5	Diversity in multi-dimensional trait space . . . . .	55
	<b>References</b>	<b>57</b>

# Part I

## Evolution in finite populations

# Chapter 1

## Demographic roots: birth-death processes for finite populations

Raymond Pearl, one of the pioneers of mathematical ecology as a discipline, mentioned that “What we want to know is how the biological forces of natality and mortality are so integrated and correlated in their action as to lead to a final result in size of population which follows this particular curve rather than some other one” (Pearl and Slobodkin, 1976). Indeed, as a more recent paper put it, “all paths to fitness lead through demography” (Metcalf and Pavard, 2007). Can we model finite populations using these observations? One potential way to incorporate birth and death analytically is by using a certain class of stochastic processes called ‘birth-death processes’.

### 1.1 Birth-death processes

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

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

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

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

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

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

## 1.2 Stochastic Differential Equations

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

$$X_t = \int_0^t F(s, X_s) ds + \int_0^t G(s, X_s) dW_s \quad (1.1)$$

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

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

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

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

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

The second integral in equation (1.1) is Itô's 'stochastic integral', and is to be interpreted in the following sense: Fix a time  $T > 0$ . For any  $n \geq 2$ , let  $\Pi_n = \{t_1, t_2, \dots, t_n\}$  be a partition of the interval  $[0, T]$ . In other words, the points contained in  $\Pi_n$  divide  $[0, T]$  into  $n$  slices of the form  $[t_i, t_{i+1}]$  such that  $0 = t_0 < t_1 < t_2 < \dots < t_n = T$ . Then, the (Itô) stochastic integral of the function  $G(t, x)$  over the time interval  $[0, T]$  is given by

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

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

Equation (1.1) is often represented in the 'differential' form:

$$dX_t = F(t, X_t)dt + G(t, X_t)dW_t \tag{1.2}$$

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

$$\frac{dx}{dt} = F(t, x) + G(t, x)\eta(t) \tag{1.3}$$

---

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

where  $\eta(t)$  is supposed to be ‘Gaussian white noise’, a ‘function’ that is defined indirectly such that the integral  $\int_0^t G(s, x)\eta(s)ds$  behaves identically to  $\int_0^t G(s, X_s)dW_s$ . However, it is important to remember that these are both purely formal<sup>2</sup> expressions - equation (1.2) is meaningless on its own and is really just shorthand for equation (1.1), which is well-defined as explained above; Equation (1.3) 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 (1.1). SDEs are convenient because they satisfy several ‘nice’ analytical properties. For example, using the fact that the Brownian motion has no expected change in value (*i.e.*  $\mathbb{E}[W_t|W_0] = W_0 = 0$ ), it can be shown that the stochastic integral also has an expectation value of 0 for all  $t$ , *i.e.*:

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

One important but not immediately obvious consequence of the definition of stochastic integrals and the Brownian motion is that we can no longer rely on the normal rules of calculus when dealing with stochastic integrals. In regular calculus, if we had a continuous quantity  $x(t)$  satisfying

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

for two ‘nice’ real functions  $f$  and  $g$ , then, given any real function  $h$ , we can calculate how the quantity  $h(x(t))$  changes over time using the chain rule of differentiation, which says that

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

*i.e.*

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

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

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

with  $gdt$  simply being replaced by  $GdW_t$  on the RHS. However, this does not work. The correct relation is instead given by Itô’s formula<sup>3</sup>:

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

There is now an extra  $h''(X_t)G^2(X_t)/2$  term that does not exist in the deterministic setting(!).

---

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

<sup>3</sup>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



## 1.3 Fokker-Planck Equations

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

Consider a one-dimensional real Itô process given by  $dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t$  that takes values in an open set  $\Omega \subseteq \mathbb{R}$  and admits a probability density function  $P(x, t)$  with compact support<sup>4</sup> in  $\Omega$  at all times  $t \in [0, \infty)$ . Let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be an arbitrary  $C^2(\mathbb{R})$  function. By Itô's formula, we have:

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

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

$$\mathbb{E}[f(X_t)] = \mathbb{E} \left[ \int_0^t \left( \mu f' + \frac{\sigma^2}{2} f'' \right) ds \right] + \mathbb{E} \left[ \int_0^t \sigma f' dW_s \right] \quad (1.4)$$

As long as  $X_t$  and  $\sigma(X_t, t)$  are reasonably 'nice', the stochastic integral in the second term of the RHS of (1.4) will vanish upon taking an expectation. 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<sup>5</sup>, we can now differentiate with respect to time on both sides and use the fundamental theorem of calculus to write

$$\int_{\Omega} f(X_t) \frac{\partial P}{\partial t}(x, t) dx = \underbrace{\int_{\Omega} \mu f' P(x, t) dx}_{M(x, t)} + \underbrace{\int_{\Omega} \frac{\sigma^2}{2} f'' P(x, t) dx}_{N(x, t)} \quad (1.5)$$

We will now use integration by parts to further evaluate the two parts  $M(x, t)$  and  $N(x, t)$ . Recall that for any two functions  $u$  and  $v$  defined on a domain  $\Omega$ , the general formula for integration by parts is given by:

$$\int_{\Omega} \frac{\partial u}{\partial x_i} v d\mathbf{x} = - \int_{\Omega} u \frac{\partial v}{\partial x_i} d\mathbf{x} + \int_{\partial\Omega} uv \gamma_i dS(\mathbf{x}) \quad (1.6)$$

---

<sup>4</sup>For our biological context you can think of this as just meaning that very extreme values in  $\Omega$  are so unlikely that we disallow them entirely

<sup>5</sup>By the Leibniz integral rule, this only requires the functions of  $(x, t)$  inside the integrals to be bounded and  $C^1$  in an open subset of  $\Omega \times [0, \infty)$

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

$$M(x, t) = - \int_{\Omega} f(X_t) \left( \frac{\partial}{\partial x} \mu P(x, t) \right) dx \quad (1.7)$$

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

$$\begin{aligned} N(x, t) &= -\frac{1}{2} \int_{\Omega} f'(X_t) \left( \frac{\partial}{\partial x} \sigma^2 P(x, t) \right) dx \\ &= \frac{1}{2} \int_{\Omega} f(X_t) \left( \frac{\partial^2}{\partial x^2} \sigma^2 P(x, t) \right) dx \end{aligned} \quad (1.8)$$

Substituting (1.7) and (1.8) into (1.5) and collecting terms yields

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

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

$$\boxed{\frac{\partial P}{\partial t}(x, t) = -\frac{\partial}{\partial x} (\mu(x, t) P(x, t)) + \frac{1}{2} \frac{\partial^2}{\partial x^2} ((\sigma(x, t))^2 P(x, t))} \quad (1.9)$$

Equation (1.9) is the Fokker-Planck equation in one dimension. Using the exact same strategy, the multidimensional Fokker-Planck equation for the  $m$ -dimensional Itô Process  $d\mathbf{X}_t = \boldsymbol{\mu}(\mathbf{X}_t, t)dt + \boldsymbol{\sigma}(\mathbf{X}_t, t)d\mathbf{W}_t$  can be easily found to be:

$$\frac{\partial P}{\partial t}(\mathbf{x}, t) = - \sum_{i=1}^m \frac{\partial}{\partial x_i} (\mu_i(\mathbf{x}, t) P(\mathbf{x}, t)) + \frac{1}{2} \sum_{i=1}^m \sum_{j=1}^m \frac{\partial^2}{\partial x_i \partial x_j} (D_{ij} P(\mathbf{x}, t)) \quad (1.10)$$

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

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

In this notation, equation (1.10) can be written in the more compact form:

$$\frac{\partial P}{\partial t} + \nabla \cdot \mathbf{J} = 0 \quad (1.11)$$

where  $\nabla \cdot$  is the divergence operator (also denoted  $\text{div}$  in some texts). Those familiar with physics should immediately recognize that equation (1.11) is in the form of a so-called ‘continuity equation’ for a conserved quantity. Continuity equations turn up in various areas of applied mathematics, most famously in electromagnetism (conservation of charge from Maxwell’s equations), fluid dynamics (continuity equations for mass of a flowing fluid from the Euler equations), and molecular diffusion (Fick’s law). This explains the name ‘probability current’ as an analogy to currents in physics such as electrical current or fluid current. The continuity equation representation also makes it clear that the Fokker-Planck equation describes the ‘flow of probability’ in the system. In particular, equation (1.11) says that the total probability in the system is ‘conserved’, and is thus simply a mathematical formalization of the common-sense idea that whenever the probability of the system of being in a given state decreases, the probability of it being in some other state must increase (and vice versa).

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 (1.9) corresponds to the solution of an SDE of the form (1.2), though the proof is much more technical and will not be discussed here. This two-way correspondence proves to be extremely useful, since the two formulations are complementary and one may be much easier than the other for solving a particular problem. This correspondence makes it greatly desirable to be able to describe our stochastic process of interest as either the solution to an Itô SDE of the form (1.2) or as the solution to a Fokker-Planck equation of the form (1.9). System-size expansions facilitate such a description for birth-death processes.

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

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

indefinitely if  $K$  is finite, reflecting the empirical fact that the total amount of resources in the world is limited (of course, this should also be reflected in how birth and death rates vary as functions of population density). Thus, in our systems,  $K$  will manifest as some fundamental limit on resources, such as habitat size or carrying capacity. When  $K$  is large, the fact that transitions occur in units of a small value  $1/K$  can be exploited via a Taylor expansion of the transition rates in the Master equation, which then yields a Fokker-Planck equation upon neglecting higher order terms<sup>6</sup>. A similar approximation is well-known (ever since Fisher) in theoretical population genetics, where it goes by the name of the ‘diffusion approximation’ (Ewens, 2004; Barton and Etheridge, 2019) or ‘continuum limit’ (Czuppon and Traulsen, 2021), and has been heavily used by Kimura (Crow and Kimura, 1970) in his stochastic models.

### 1.3.2 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<sup>7</sup> (obtained by letting  $K \rightarrow \infty$ ) and only works if  $K$  is large enough to be able to neglect all but the highest-order terms. This is usually an excellent approximation for populations in which the deterministic trajectory has already reached an attractor (stable fixed point, stable limit cycle, etc.). Since many deterministic eco-evolutionary models are expected to relax to such attractors, such an approximation is a useful first step in increasing the generality of existing models (which are usually studied only in the equilibrium regime) to incorporate the dynamics of finite populations. 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.

## 1.4 Stochastic equations for population size dynamics

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

---

<sup>6</sup>If this sounds handwavy to you, see chapter 11, section 3 in Ethier and Kurtz (1986) for a more rigorous treatment.

<sup>7</sup>This idea can be made much more rigorous via an analog of the central limit theorem for density-dependent Markov chains. See chapter 11, section 2 in Ethier and Kurtz (1986)

### 1.4.1 Description of the process and the Master Equation

Consider a population of identical individuals subject to some ecological rules that affect individuals' birth and death rates. I will neglect any potential factor that could lead to two individuals becoming non-identical in their birth and death rates, like mutations, etc. Since all individuals are identical in their birth and death rates, we only really need to track the total population size through time to know everything there is to know about the population. The population as a whole at any time  $t$  can thus be characterized by a single number - its population size (Figure 1.1). Imagine further that if a population has  $n$  identical individuals, then, from the ecological rules, we can determine a *birth rate*  $b(n)$ , which gives us a measure of the probability that a new individual will be born and the population size becomes  $n + 1$  'per unit time'. One must be slightly precise about what exactly they mean when they say 'per unit time' since there are no discrete 'time steps' for individuals to be born. Here, by 'birth rate', I mean the probability that there will be a birth (and no death) per an *infinitesimal* amount of time. More formally, letting  $N_t$  denote the random variable representing the population size at time  $t$  and letting  $\mathbb{P}(E)$  denote the probability (in the common-sense usage) of an event  $E$ , the birth rate<sup>8</sup>  $b(n)$  of individuals a population with population size  $n$  is the quantity

$$b(n) := \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P}(N_{t+\epsilon} = n + 1 | N_t = n) \quad (1.12)$$

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

$$d(n) := \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \mathbb{P}(N_{t+\epsilon} = n - 1 | N_t = n) \quad (1.13)$$

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

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

and the probability that the event is instead a death is

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

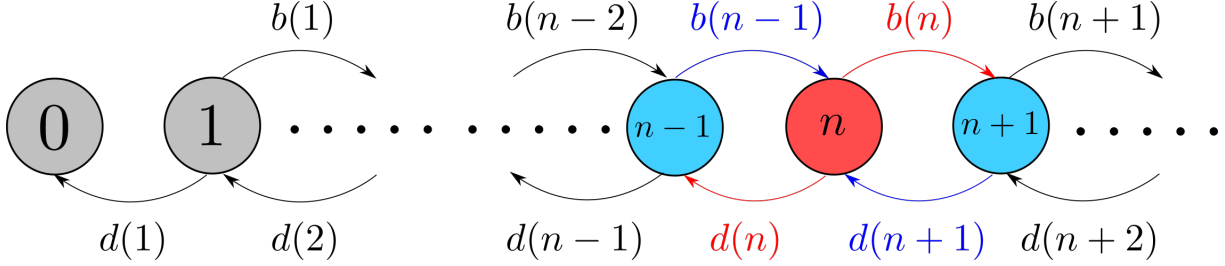
**Example 1.** Consider the case where the per-capita birth rate is a constant  $\lambda > 0$ , *i.e.*,  $b(n) = \lambda n$ , and the per-capita death rate has the linear density-dependence  $(\mu + (\lambda - \mu)\frac{n}{K})$ , *i.e.* the total death rate is  $d(n) = (\mu + (\lambda - \mu)\frac{n}{K})n$ , where  $\mu$  and  $K$  are positive constants. Taking the difference between the birth and death rates, we obtain  $b(n) - d(n) = (\lambda - \mu)n(1 - \frac{n}{K})$ ,

---

<sup>8</sup>Note that unlike usual ecology convention, this is *not* a per-capita birth rate. The ecology per capita birth and death rates can be found by dividing my birth/death rates by the current population size  $n(t)$

where, identifying  $r = \lambda - \mu$ , we obtain the familiar functional form of the density dependence of the logistic equation. 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.



**Figure 1.1: Schematic description of a one-dimensional birth-death process.** Consider a population of identical individuals. The state of the system can be described by a single number, the population size (numbers within the circles). Births and deaths result in changes in the total population size, and the birth and death rates (arrows) are dependent on the current population size. For a given state  $n$ , the blue arrows depict the rate of ‘inflow’ to the state (from the blue states), whereas the red arrows depict the rate of ‘outflow’.

To do this, we imagine a large ensemble of populations. In a large ensemble of copies evolving independently, a fraction  $P(n, t)$  will have population size  $n$  at time  $t$  by definition of probability. We can now simply measure the ‘inflow’ and ‘outflow’ of copies of the population from each state (Figure 1.1). If a population has  $n$  individuals, it could either have gotten there from a population of  $n + 1$  individuals, with a death rate of  $d(n + 1)$ , or from a population of  $n - 1$  individuals, with a birth rate of  $b(n - 1)$ . Thus, the rate of ‘inflow’ to the state  $n$  is given by

$$R_{\text{in}}(n, t) = b(n - 1)P(n - 1, t) + d(n + 1)P(n + 1, t) \quad (1.14)$$

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

$$R_{\text{out}}(n, t) = b(n)P(n, t) + d(n)P(n, t) \quad (1.15)$$

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

$$\begin{aligned} \frac{\partial P}{\partial t}(n, t) &= R_{\text{in}}(n, t) - R_{\text{out}}(n, t) \\ &= b(n - 1)P(n - 1, t) + d(n + 1)P(n + 1, t) - b(n)P(n, t) - d(n)P(n, t) \end{aligned} \quad (1.16)$$

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

$$\frac{\partial P}{\partial t}(n, t) = (\mathcal{E}^- - 1)b(n)P(n, t) + (\mathcal{E}^+ - 1)d(n)P(n, t) \quad (1.17)$$

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 (1.17) directly.

### 1.4.2 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 (1.17) becomes

$$\frac{\partial P}{\partial t}(x, t) = (\Delta^- - 1)Kb_K(x)P(x, t) + (\Delta^+ - 1)Kd_K(x)P(x, t) \quad (1.18)$$

where we now have the new step operators

$$\Delta^\pm f(x, t) = f\left(x \pm \frac{1}{K}, t\right) \quad (1.19)$$

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 (1.18) and neglecting terms of  $\mathcal{O}(K^{-3})$  and higher, we obtain

$$\boxed{\frac{\partial P}{\partial t}(x, t) = -\frac{\partial}{\partial x}\{A^-(x)P(x, t)\} + \frac{1}{2K}\frac{\partial^2}{\partial x^2}\{A^+(x)P(x, t)\}} \quad (1.20)$$

where

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

Equation (1.20) is a Fokker-Planck equation and corresponds to the SDE:

$$dX_t = A^-(X_t)dt + \sqrt{\frac{A^+(X_t)}{K}}dW_t \quad (1.21)$$

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

### 1.4.3 Stochastic fluctuations and the weak noise approximation

If we assume the noise is weak (this will be made precise shortly), then we can go still further with analytic techniques by measuring fluctuations from the deterministic expectations, albeit with some slightly cumbersome calculations to arrive at the final expressions. We will grit our teeth and get through the algebra below, with my promise that the final answer is neat and easy to handle. It is clear that as  $K \rightarrow \infty$ , equation (1.21) describes a deterministic process, obtained as the solution to

$$\frac{dx}{dt} = A^-(x) = b_K(x) - d_K(x) \quad (1.22)$$

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

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



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

$$\begin{aligned}
\frac{\partial \tilde{P}}{\partial t} &= \frac{\partial \tilde{P}}{\partial y} \frac{\partial y}{\partial t} + \frac{\partial \tilde{P}}{\partial s} \frac{\partial s}{\partial t} \\
&= \frac{\partial \tilde{P}}{\partial y} \left( -\sqrt{K} \frac{d\alpha}{dt} \right) + \frac{\partial \tilde{P}}{\partial s} \\
&= -\sqrt{K} A^-(\alpha(s)) \frac{\partial \tilde{P}}{\partial y} + \frac{\partial \tilde{P}}{\partial s}
\end{aligned} \tag{1.23}$$

and

$$\frac{\partial}{\partial y} = \frac{1}{\sqrt{K}} \frac{\partial}{\partial x} \tag{1.24}$$

Reformulating (1.20) in terms of  $y, s$  and  $\tilde{P}$  and substituting (1.23) and (1.24) yields:

$$\begin{aligned}
-A^-(\alpha) \frac{\partial \tilde{P}}{\partial x} + \frac{\partial \tilde{P}}{\partial s} &= -\sqrt{K} \frac{\partial}{\partial y} \left( A^-(\alpha + \frac{y}{\sqrt{K}}) \tilde{P} \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left( A^+(\alpha + \frac{y}{\sqrt{K}}) \tilde{P} \right) \\
\Rightarrow \frac{\partial \tilde{P}}{\partial s} &= -\frac{\partial}{\partial y} \left[ \sqrt{K} \left( A^-(\alpha + \frac{y}{\sqrt{K}}) - A^-(\alpha) \right) \tilde{P} \right] + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left( A^+(\alpha + \frac{y}{\sqrt{K}}) \tilde{P} \right)
\end{aligned} \tag{1.25}$$

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

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

with  $A_0^-(s) = A^-(\alpha(s))$ ,  $A_0^+(s) = A^+(\alpha(s))$ . These could be Taylor expansions, for example, but the exact form of the coefficients is irrelevant as long as it is known to us, so any expansion will work. We can now substitute these series expansions into (1.25) to obtain:

$$\begin{aligned}
\sum_{n=0}^{\infty} \left( \frac{1}{\sqrt{K}} \right)^n \frac{\partial \tilde{P}_n}{\partial s} &= -\frac{\partial}{\partial y} \left[ \sqrt{K} \left( \sum_{n=1}^{\infty} A_n^-(s) \left( \frac{y}{\sqrt{K}} \right)^n \right) \left( \sum_{m=0}^{\infty} \tilde{P}_m \left( \frac{1}{\sqrt{K}} \right)^m \right) \right] \\
&\quad + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left[ \left( \sum_{n=0}^{\infty} A_n^+(s) \left( \frac{y}{\sqrt{K}} \right)^n \right) \left( \sum_{m=0}^{\infty} \tilde{P}_m \left( \frac{1}{\sqrt{K}} \right)^m \right) \right]
\end{aligned} \tag{1.26}$$

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 (1.26) is precisely

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

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

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

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

$$\frac{\partial \tilde{P}_r}{\partial s} = -\frac{\partial}{\partial y} \left( \sum_{m=0}^r \tilde{P}_m A_{r-m+1}^- y^{r-m+1} \right) + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left( \sum_{m=0}^r \tilde{P}_m A_{r-m}^+ y^{r-m} \right) \quad (1.27)$$

We will now make the meaning of ‘weak’ precise. Let us assume that the fluctuations are weak enough that we can obtain a reasonable approximation of the dynamics by retaining only the highest order (in  $1/K$ ) term in equation (1.27) and neglecting all higher-order terms<sup>9</sup>. We are then left with the expression:

$$\frac{\partial \tilde{P}_0}{\partial s} = -A_1^-(s) \frac{\partial}{\partial y} (y \tilde{P}_0) + \frac{A_0^+(s)}{2} \frac{\partial^2 \tilde{P}_0}{\partial y^2} \quad (1.28)$$

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

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

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

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

---

<sup>9</sup>For example, we may imagine this is reasonable if the deterministic trajectory is at a stable fixed point and subject to weak fluctuations

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

$$Y_t = Y_0 \exp \left( \int_0^t A_1^-(s) ds \right) + \int_0^t \exp \left( - \int_s^t A_1^-(v) dv \right) \sqrt{A^+(\alpha(s))} dW_s \quad (1.29)$$

as the zeroth-order weak noise approximation for stochastic fluctuations from the deterministic trajectory due to demographic noise. This equation can be exactly solved via analytical techniques and one can get many insights from the solution. For example, if  $Y_0 = 0$  (*i.e.* we start at the deterministic steady state, a natural assumption for measuring fluctuations from it), then we can show by taking expectations in (1.29) and using results presented in 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 (1.29) using some tools from stochastic calculus, but I will not demonstrate this here.

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

## 1.5 An example: The stochastic logistic equation

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

$$\begin{aligned} b(n) &= \lambda n \\ d(n) &= \left( \mu + (\lambda - \mu)\frac{n}{K} \right) n \end{aligned} \quad (1.30)$$

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

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

Thus, we have

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

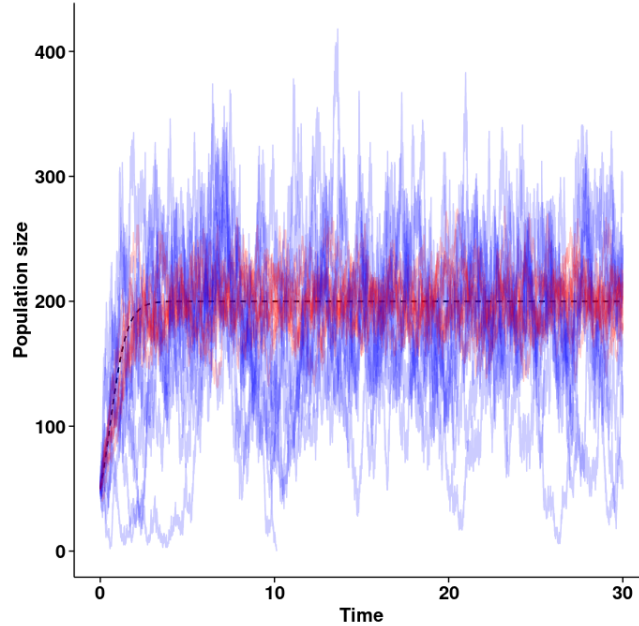
Defining  $r = \lambda - \mu$  and  $v = \lambda + \mu$  and using equation (1.21), we see that the ‘mesoscopic view’ of the system is given by the solution of the SDE

$$dX_t = rX_t(1 - X_t)dt + \sqrt{\frac{X_t(v + rX_t)}{K}}dW_t \quad (1.31)$$

From equation (1.22), we see that the deterministic dynamics are

$$\frac{dx}{dt} = A^-(x) = rx(1-x) \quad (1.32)$$

showing that in the infinite population limit, we obtain the logistic equation. This derivation also makes it clear that two systems with very different stochastic dynamics can nevertheless converge to the same infinite population limit (1.32), since equation (1.32) only depends on the difference  $\lambda - \mu$ .



**Figure 1.2: Comparison of time series produced by Gillespie simulations of the stochastic logistic equation for two different parameter values.** Simulations with the parameter values  $\lambda = 4, \mu = 2$  are plotted in red, and simulations with the parameter values  $\lambda = 31, \mu = 29$  are plotted in blue. Both simulations have  $K = 200$ . The infinite population limit of both simulations is plotted in black dotted lines. The graph shows 10 realizations each for the two sets of parameter values.

To illustrate the effects of this seemingly innocent fact, figure 1.2 compares two simulations which have the same value of  $\lambda - \mu$  but a ten-fold difference in  $\lambda + \mu$ . As is clear from the figure, though both populations have the same behavior at the infinite population limit, populations with a higher value of  $\lambda + \mu$  exhibit much wilder fluctuations and are therefore more prone to stochastic extinction.

Letting  $\alpha(t)$  be the solution of the logistic equation (1.32), We can Taylor expand  $A^\pm(x)$

for the weak noise approximation, and we find:

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

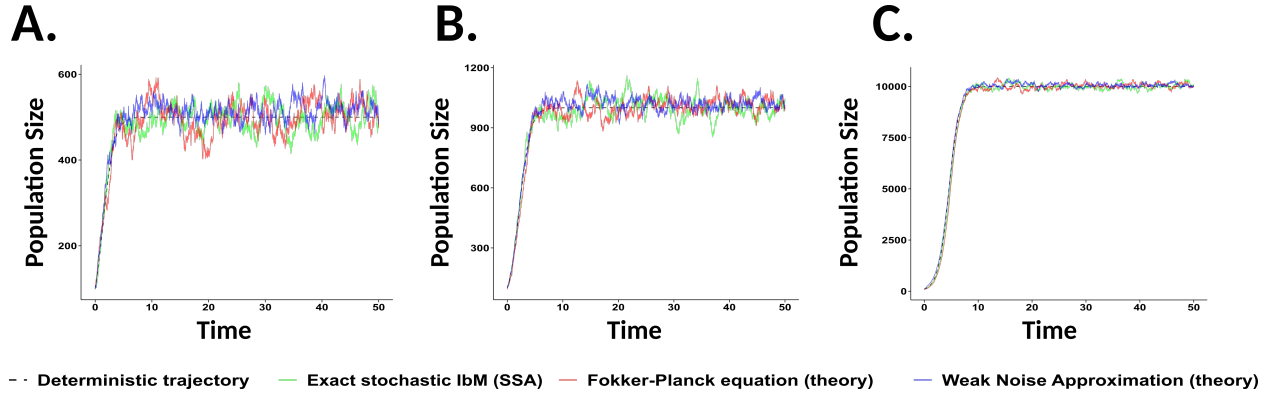
Thus, the weak noise approximation of 1.30 is given by

$$X_t = \alpha(t) + \frac{1}{\sqrt{K}} Y_t \quad (1.33)$$

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

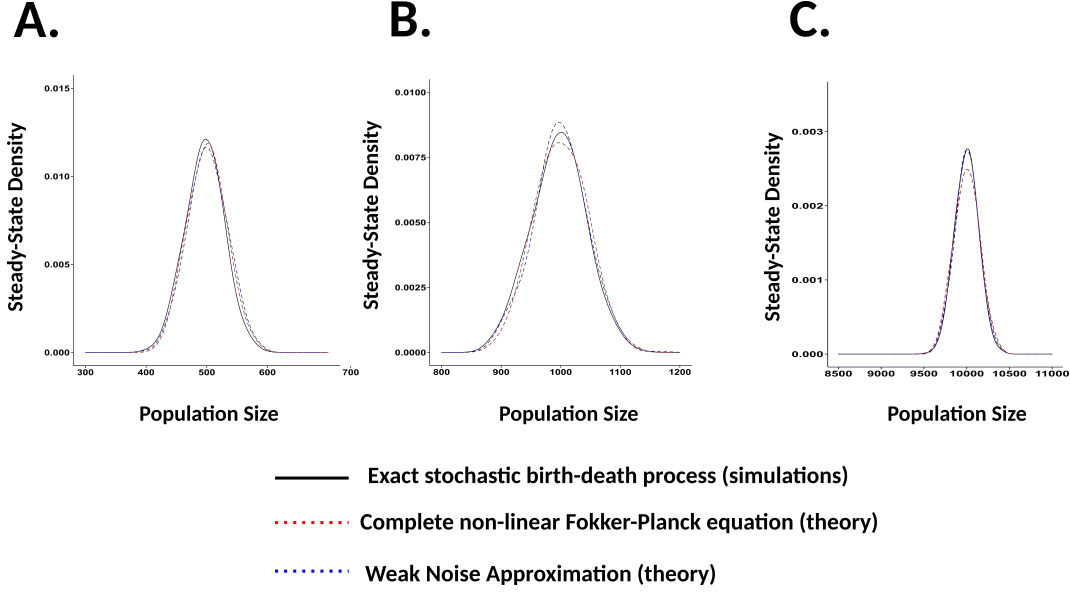
$$\begin{aligned} dY_t &= A_1^-(t)Y_t dt + \sqrt{A_0^+(t)} dW_t \\ \Rightarrow dY_t &= r(1-2\alpha(t))Y_t dt + \sqrt{\alpha(t)(v+r\alpha(t))} dW_t \end{aligned} \quad (1.34)$$

The time series predicted by these three processes look qualitatively similar and all seem to fluctuate about the deterministic steady state (Figure 1.3).



**Figure 1.3: Comparison of a single realization** of the exact birth-death process (1.30), the deterministic trajectory (1.32), the non-linear Fokker-Planck equation (1.31), and the weak noise approximation (1.34) for **(A)**  $K = 500$ , **(B)**  $K = 1000$ , and **(C)**  $K = 10000$ .  $\lambda = 2, \mu = 1$  for all three cases.

The deterministic trajectory (1.32) 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.



**Figure 1.4: Comparison of the steady-state densities** given by (1.30), (1.31), and (1.34) for **(A)**  $K = 500$ , **(B)**  $K = 1000$ , and **(C)**  $K = 10000$ .  $\lambda = 2, \mu = 1$  for all three cases. Each curve was obtained using 1000 independent realizations.

The stochastic dynamics (1.31) and (1.34), however, depend not only on  $r$ , but also on  $v$ , the sum of the birth and death rates. It has been proven that  $X_t = 0$  is the only recurrent state for the full stochastic dynamics (1.31), meaning that every population is guaranteed to go extinct<sup>10</sup> 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 (1.31) 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 (1.34) implicitly assumes non-extinction by only measuring small fluctuations from the deterministic solution to (1.32) and thus, at steady state, naturally describes a quasistationary distribution centered about  $X_t = 1$ . The steady-state density (probability density function as  $t \rightarrow \infty$ ) of the exact birth-death process (1.30) is compared with that predicted by (1.31) and (1.34) for various values of  $K$  in figure 1.4.

<sup>10</sup>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.

## Part II

### Spatial ecology and random walks

# Chapter 2

## Random walks

Individuals often move around for various important reasons such as foraging or finding a mate. A tool that is often used to model such movement is the theory of *random walks*. In this lecture, we look at a relatively non-technical introduction to unbiased random walks.

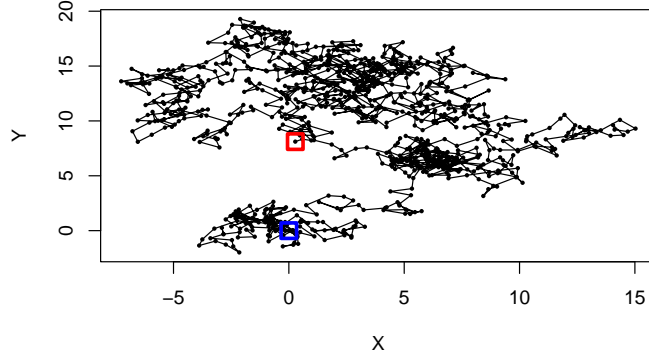
### 2.1 A simple random walk

Consider an organism on a discrete one-dimensional space. Let  $X_t$  be the random variable representing the position of the organism at time  $t$ . Throughout this discussion, we assume that time is discrete. Consider a stochastic process that assigns the following movement rule to the animal:

$$X_{t+1} = \begin{cases} X_t + a & \text{with probability } \frac{1}{2} \\ X_t - a & \text{with probability } \frac{1}{2} \end{cases} \quad (2.1)$$

The organism can be thought of as executing as a 'random walk' over the one-dimensional space. Figure 2.1 shows what a trajectory produced by one such process may look like in two dimensions. In this figure, the blue square indicates the initial position of the organism, at the origin, and the red square indicates the position after 10,000 time steps.





**Figure 2.1:** A sample trajectory of an organism moving according to Equation (2.1).

More generally, a *one-dimensional random walk* is a stochastic process of the form:

$$X_{t+1} = X_t + a \quad (2.2)$$

where  $a$  is a random variable with some probability distribution. In Equation (2.1),  $a$  is a bernoulli(1/2) random variable.

If we were to observe an organism starting at  $X_0$  and moving along one dimension according to (2.1) for  $t$  timesteps, where may we expect it to end up? In other words, what is the expectation value of the *displacement*  $D_t := X_t - X_0$ ? The calculation for the displacement is straightforward if one notices that  $X_t$  can be written in terms of  $X_{t-1}$  as follows:

$$\begin{aligned} \langle X_t \rangle &= \frac{1}{2}(\langle X_{t-1} \rangle + a) + \frac{1}{2}(\langle X_{t-1} \rangle - a) \\ &= \langle X_{t-1} \rangle \end{aligned}$$

Thus, applying this recursively  $t$  times, we obtain:

$$\begin{aligned} \langle X_t \rangle &= X_0 \\ \implies \langle D_t \rangle &= 0 \end{aligned}$$

In other words, the expectation value is 0, which sounds like the organism hasn't moved at all! It is easy to see that this argument is independent of the dimensionality of the walk. However, a cursory glance at 2.1 tells us that the organism has moved away from the origin after all. Where, then, have we gone wrong?

The issue is that since the displacement can be either positive or negative, all our calculation tells us is that if we had many such systems, the displacement would be zero *on average* (Note: such a walk is often called an 'unbiased' random walk). However, one could get  $\langle D_t \rangle = 0$  even if the organism went very far off in the positive  $x$  direction in half of the times we tried the process, and very far off in the negative  $x$  direction in the other half of the times. The expectation value thus fails to capture the notion of 'spread away from the starting point'. For this, we require a slightly different quantity.

### 2.1.1 The mean square displacement

Intuitively, the expectation value can be zero because realizations with a high positive distance covered 'cancel out' with those which had a high negative distance covered, since both are equally likely to occur. To account for this, we will instead define the 'mean square displacement' (MSD) of a random walk as:

$$MSD(t) := \langle (X_t - X_0)^2 \rangle$$

We can slightly simplify the term on the right using the linearity of the expectation value as follows:

$$MSD(t) = \langle X_t^2 \rangle - 2X_0\langle X_t \rangle + X_0^2 \quad (2.3)$$

Using the fact that  $\langle X_t \rangle = X_0$  for an unbiased random walk, we then see that:

$$MSD(t) = \langle X_t^2 \rangle - X_0^2 \quad (2.4)$$

Equation (2.4) is very general, and applies to any unbiased one-dimensional random walk whatsoever. For the random walk given in Equation (2.1), we have:

$$X_t^2 = \begin{cases} X_{t-1}^2 + a^2 + 2aX_{t-1} & \text{with probability } \frac{1}{2} \\ X_{t-1}^2 + a^2 - 2aX_{t-1} & \text{with probability } \frac{1}{2} \end{cases}$$

and thus

$$\langle X_t^2 \rangle = \langle X_{t-1}^2 + a^2 \rangle$$

This is a recursive equation. One can write  $X_{t-1}^2$  in terms of  $X_{t-2}^2$ ,  $X_{t-2}^2$  in terms of  $X_{t-3}^2$ , and so on. After doing this  $t$  times, we arrive at:

$$\langle X_t^2 \rangle = X_0^2 + ta^2$$

Substituting this into Equation (2.4), we arrive at:

$$MSD(t) = ta^2 \quad (2.5)$$

This is telling us something profound. It tells us that as the time increases, we expect the MSD to linearly increase, meaning that in any given realization, an organism executing such a walk is likely to be found further away from the origin as time increases. Note that this is despite the fact that the organism is *equally likely* to go in either the positive or the negative direction at any given instance(!).

### 2.1.2 A general expression for the MSD

Consider a random walk of the form (2.2). Let the random variable  $a$  have mean  $\mu$  and variance  $\sigma^2$ . We will now obtain an expression for the MSD of such a process. Since we are

not assuming that the walk is unbiased, we must use equation (2.3) for the definition of the MSD.

Since  $\langle X_t \rangle = \langle X_{t-1} \rangle + \langle a \rangle$ , we can apply this relation recursively to obtain:

$$\langle X_t \rangle = X_0 + t\mu \quad (2.6)$$

We can also calculate  $\langle X_t^2 \rangle$  in a similar manner:

$$\begin{aligned} \langle X_t^2 \rangle &= \langle X_{t-1}^2 \rangle + \langle a^2 \rangle + 2\langle aX_{t-1} \rangle \\ &= \langle X_{t-1}^2 \rangle + (\sigma^2 + \mu^2) + 2[\text{cov}(a, X_{t-1}) + \mu\langle X_{t-1} \rangle] \end{aligned}$$

Substituting equation (2.6), we obtain:

$$\langle X_t^2 \rangle = \langle X_{t-1}^2 \rangle + (\sigma^2 + \mu^2) + 2\text{cov}(a, X_{t-1}) + \mu(X_0 + t\mu)$$

This, too, can be applied recursively to obtain:

$$\langle X_t^2 \rangle = X_0^2 + t(\sigma^2 + \mu^2) + 2t\mu(X_0 + (t-1)\mu) + 2\sum_{k=1}^{t-1} \text{cov}(a, X_k) \quad (2.7)$$

Finally, we can plug (2.6) and (2.7) into (2.3) to obtain our final expression:

$$MSD(t) = t[\sigma^2 + \mu^2(2t-1)] + 2\sum_{k=1}^{t-1} \text{cov}(a, X_k) \quad (2.8)$$

### 2.1.3 Dimensionality considerations

So far, we have only been considering a one-dimensional walk. What happens when an organism moves in multiple dimensions? Assuming that the choice of direction made in each dimension is independent, the MSD simply adds linearly. To see this, note that for a  $D$  dimensional random walk, the MSD is given by:

$$MSD_D(t) = \sum_{d=1}^D \langle (X_d(t) - X_d(0))^2 \rangle$$

where  $X_d(t)$  denotes the random variable for the value of the  $d^{\text{th}}$  dimension of the position at time  $t$ . However, each term in the summation is simply the MSD of a one-dimensional walk along that dimension. Thus, we see that:

$$MSD_D(t) = \sum_{d=1}^D MSD_d(t)$$

where  $MSD_d(t)$  is the MSD for a one-dimensional random walk along the  $d^{\text{th}}$  dimension at time  $t$ .

## 2.2 Introducing bias

In the random walk we have seen thus far, the organism is *equally likely* to move either left or right at each time step. This process can also be written as:

$$\begin{aligned} X_{t+1} &= X_t + aY_t - a(1 - Y_t) \\ &= X_t + 2a(Y_t - \frac{1}{2}) \end{aligned} \tag{2.9}$$

where  $Y_t$  is a Bernoulli( $\frac{1}{2}$ ) random variable, i.e

$$Y_t = \begin{cases} 1 & \text{with probability } \frac{1}{2} \\ 0 & \text{with probability } \frac{1}{2} \end{cases}$$

The *final* position after  $t$  time-steps will then be captured by a Binomial( $t, \frac{1}{2}$ ) distribution, i.e. given  $X_0 = 0$ , we have:

$$\begin{aligned} X_t &= X_0 + aZ_t - ta + aZ_t \\ &= 2a(Z_t - \frac{t}{2}) \end{aligned} \tag{2.10}$$

where  $Z_t := \sum_{i=1}^t Y_i \sim \text{Bin}(t, \frac{1}{2})$ . What if the probabilities of moving left and right were *not* equal? What if, instead, the organism moved left with probability  $p$  and right with probability  $q$  at each time step?

We will reframe this process in terms of Equations (2.9) and (2.10) for analytical ease. Thus, the position at time  $t + 1$  given that the position at time  $t$  is  $X_t$  is given by:

$$X_{t+1} = X_t + 2a(Y_t - \frac{1}{2})$$

where, now,  $Y_t \sim \text{Bernoulli}(p)$ . Thus, using equation (10), the displacement after  $t$  timesteps is given by:

$$D_t = 2a(Z - \frac{t}{2})$$

where  $Z \sim \text{Bin}(t, p)$ . The expected displacement is then given by:

$$\begin{aligned} \langle D_t \rangle &= 2a(\langle Z_t \rangle - \frac{t}{2}) \\ &= 2at(p - \frac{1}{2}) \\ &= at(2p - 1) \\ &= at(p + (1 - q) - 1) \\ &= at(p - q) \end{aligned} \tag{2.11}$$

And the mean squared displacement is given by:

$$\begin{aligned}
MSD(t) &= 4a^2(\langle Z_t^2 \rangle - t\langle Z_t \rangle + \frac{t^2}{4}) \\
&= 4a^2 \left( \text{Var}(Z_t) + \langle Z_t \rangle^2 - t\langle Z_t \rangle + \frac{t^2}{4} \right) \\
&= 4a^2 \left( tp(1-p) + t^2p^2 - t^2p + \frac{t^2}{4} \right) \\
&= 4a^2t \left( p^2(t-1) - p(t-1) + \frac{t}{4} \right) \\
&= 4a^2t \left( p(p-1)(t-1) + \frac{t}{4} \right) \\
&= 4a^2t \left( \frac{t}{4} - (t-1)pq \right)
\end{aligned} \tag{2.12}$$

Equation (2.11) tells us that when the walk is biased, the mean displacement is non-zero, and is proportional to  $t$  and to the ‘bias’  $p - q$  of the process. The displacement is thus in the direction for which probability of movement in each time step is higher, as one would expect.

## 2.3 Rate of dispersion

How quickly can we expect an organism to ‘disperse’ from its starting position? We say that a random walk has a dispersion rate of  $\alpha$  if  $MSD \propto t^\alpha$ . Given actual data,  $\alpha$  can be obtained by plotting  $\log(MSD)$  vs  $\log(t)$  and finding the slope of the line.

From equations (2.5) and (2.12), we see that for an unbiased random walk,  $\alpha = 1$ , and for a biased random walk,  $\alpha = 2$ . Thus, an organism executing a biased random walk moves away from its initial location ‘faster’ than one executing an unbiased random walk. In general, processes with  $\alpha \neq 1$  are said to exhibit ‘anomalous’ diffusion. Processes with  $\alpha < 1$  are said to be ‘sub-diffusive’ and processes with  $\alpha > 1$  are said to be ‘super-diffusive’. In the next class, we will see why the word ‘diffusion’ is used here and how the discrete processes we have been studying here relate to the continuous process of diffusion.

# Chapter 3

## The Diffusion Equation

Consider a very large number of organisms starting from the same point and executing an unbiased random walk. We would like to know what we would expect to see if we were to observe this system for a long time. Let  $\phi(x, t)$  denote the number density of particles at position  $x$  at time  $t$ . Consider a small time interval  $\tau$ , comparable to the length of a single time step of the random walk. We note that any particle that is at position  $x$  at time  $t$  must have either been there to begin with, or must have been somewhere else at time  $t - \tau$  and moved a small distance  $\delta x$  (small enough to be comparable to the step size of the random walk) to arrive at its current position. Thus, we can write:

$$\phi(x, t+\tau) = \underbrace{\phi(x, t)}_{\substack{\text{Number of} \\ \text{particles} \\ \text{present at} \\ x}} + \underbrace{\frac{1}{2}}_{\substack{\text{Probability of} \\ \text{entering } x \\ \text{from } x - \delta x}} \underbrace{(\phi(x - \delta x, t))}_{\substack{\text{Number of} \\ \text{particles} \\ \text{present at} \\ x - \delta x}} + \underbrace{\frac{1}{2}}_{\substack{\text{Probability of} \\ \text{entering } x \\ \text{from } x + \delta x}} \underbrace{(\phi(x + \delta x, t))}_{\substack{\text{Number of} \\ \text{particles} \\ \text{present at} \\ x + \delta x}} - \underbrace{\phi(x, t)}_{\substack{\text{Number of} \\ \text{particles} \\ \text{present at} \\ x}} \underbrace{\left(\frac{1}{2} + \frac{1}{2}\right)}_{\substack{\text{Probability of} \\ \text{leaving } x}} \quad (3.1)$$

Thus, we have:

$$\phi(x, t + \tau) - \phi(x, t) = \frac{1}{2} [\phi(x - \delta x, t) + \phi(x + \delta x, t) - 2\phi(x, t)] \quad (3.2)$$

We can now Taylor expand the first two terms of equation (3.2) as:

$$\begin{aligned} \phi(x - \delta x, t) &= \phi(x, t) - \delta x \frac{\partial \phi}{\partial x} + \frac{(\delta x)^2}{2} \frac{\partial^2 \phi}{\partial x^2} + \dots \\ \phi(x + \delta x, t) &= \phi(x, t) + \delta x \frac{\partial \phi}{\partial x} + \frac{(\delta x)^2}{2} \frac{\partial^2 \phi}{\partial x^2} + \dots \end{aligned}$$

Thus, adding them together, we have:

$$\phi(x - \delta x, t) + \phi(x + \delta x, t) = 2\phi(x, t) + (\delta x)^2 \frac{\partial^2 \phi}{\partial x^2} + \mathcal{O}(4)$$

Neglecting 4th order terms and higher, we can now substitute this back into equation (3.2) to obtain:

$$\phi(x, t + \tau) - \phi(x, t) = (\delta x)^2 \frac{\partial^2 \phi}{\partial x^2}$$

Dividing throughout by  $\tau$ , we arrive at:

$$\frac{\phi(x, t + \tau) - \phi(x, t)}{\tau} = \frac{(\delta x)^2}{\tau} \frac{\partial^2 \phi}{\partial x^2}$$

We now recall that both  $\delta x$  and  $\tau$  are very small. Thus, taking limits, we have

$$\lim_{\tau \rightarrow 0} \frac{\phi(x, t + \tau) - \phi(x, t)}{\tau} = \lim_{\tau \rightarrow 0, \delta x \rightarrow 0} \frac{(\delta x)^2}{\tau} \frac{\partial^2 \phi}{\partial x^2}$$

Now, the left hand side of this equation is the definition of a derivative. For the right hand side, we recall that for a one-dimensional unbiased random walk,  $MSD \propto t$ . Thus, we would expect  $\frac{(\delta x)^2}{\tau}$  to be a constant, since  $\tau$  is of the order of a single time-step of the random walk.

Defining the diffusion constant  $D := \frac{(\delta x)^2}{\tau}$ , we arrive at:

$$\frac{\partial \phi(x, t)}{\partial t} = D \frac{\partial^2 \phi(x, t)}{\partial x^2} \quad (3.3)$$

or, in multiple dimensions,

$$\boxed{\frac{\partial \phi(\mathbf{r}, t)}{\partial t} = D \nabla^2 \phi(\mathbf{r}, t)} \quad (3.4)$$

Equation (3.4) is the famous ‘diffusion equation’ (sometimes also referred to as Fick’s Law), and was historically used to model the diffusion of particles in a fluid. Thus, our derivation tells us that particles (or animals) that execute a random walk at a microscopic scale appear to behave like a diffusing fluid on macroscopic scales.

### 3.1 Typical Behavior of solutions to the diffusion equation

Suppose we have  $N$  one-dimensional animals all starting out at the origin. Then, the initial density is given by  $\phi(x, 0) = \delta(x)$ , where  $\delta(x)$  is the dirac delta function given by:

$$\delta(x) = \begin{cases} 1 & x = 0 \\ 0 & \text{otherwise} \end{cases}$$

The solution to equation (3.3) given this initial condition is:

$$\phi(x, t) = N \sqrt{\frac{1}{4\pi Dt}} e^{-\frac{x^2}{4Dt}}$$

This is the form of a Normal distribution. Specifically, the equation for  $\phi(x, t)$  is in form of a  $N(0, 2Dt)$  distribution. Thus, we see that the number density of animals is given by a Gaussian centered at the origin, with the spread of the Gaussian increasing linearly with time. This confirms our intuition of what we would expect when we say particles ‘diffuse’ from the focal point.

## 3.2 Introducing bias: The drift-diffusion equation

Instead of the unbiased population we have been considering so far, consider a population of one dimensional animals each of whom are executing a move right with probability  $p$  and move left with probability  $q = 1 - p$ . How would the diffusion equation change in this case? Well, equation (3.1) still applies, but the probabilities of moving in from  $x - \delta x$  and  $x + \delta x$  are no longer half, but are instead given by  $p$  and  $q$  respectively. Thus, Taylor expanding as earlier, we arrive at:

$$\lim_{\tau \rightarrow 0} \frac{\phi(x, t + \tau) - \phi(x, t)}{\tau} = \lim_{\tau \rightarrow 0, \delta x \rightarrow 0} \frac{(\delta x)^2}{\tau} \frac{\partial^2 \phi}{\partial x^2} + (q - p) \frac{\delta x}{\tau} \frac{\partial \phi}{\partial x}$$

Defining  $D = \lim_{\tau \rightarrow 0} \frac{(\delta x)^2}{\tau}$  and  $u = \lim_{\tau \rightarrow 0} (q - p) \frac{\delta x}{\tau}$ , we arrive<sup>1</sup> at the ‘drift-diffusion equation’:

$$\frac{\partial \phi(x, t)}{\partial t} = \underbrace{D_{p,q} \frac{\partial^2 \phi(x, t)}{\partial x^2}}_{\text{“Diffusion”}} + \underbrace{u \frac{\partial \phi(x, t)}{\partial x}}_{\text{“Drift”}} \quad (3.5)$$

or, in multiple spatial dimensions:

$$\boxed{\frac{\partial \phi(\mathbf{r}, t)}{\partial t} = D_{p,q} \nabla^2 \phi(\mathbf{r}, t) + u \nabla \cdot \phi(\mathbf{r}, t)} \quad (3.6)$$

The ‘drift’ term in equation (3.5) adds a bias to the solutions, and this bias is proportional to  $(q - p)$ .

## 3.3 Typical behavior of solutions to the drift-diffusion equation

As before, consider an initial population of  $N$  animals given by  $\phi(x, 0) = \delta(x)$ . The solution to the drift-diffusion equation given this initial condition is:

$$\phi(x, t) = N \sqrt{\frac{1}{4\pi Dt}} e^{\frac{-(x-ut)^2}{4Dt}}$$

As before, we have an ever-spreading Gaussian. However, this time, the mean of the Gaussian changes linearly with time. The solution is thus a ‘travelling wave’ of ever-diminishing magnitude. The direction and speed of travel depend on the value of  $q - p$ , by the definition of  $u$ .

---

<sup>1</sup>The astute reader will notice that  $\frac{\delta x}{\tau}$  and  $\frac{(\delta x)^2}{\tau}$  cannot really both simultaneously tend to non-zero constants. To make it work, we say that we only look at weak drift ( $q \rightarrow p$ ) so that  $(q - p)\delta x/\tau$  and  $(\delta x)^2/\tau$  both go to non-zero constants as  $q \rightarrow p, \delta x \rightarrow 0, \tau \rightarrow 0$



# Chapter 4

## Turing Patterns

Consider a system consisting of two species of one-dimensional animals. Let the number of individuals of species 1 in the total population be given by  $N_1$  and let the number of individuals of species 2 in the population be given by  $N_2$ . Let the dynamics of this system be given by the coupled reaction-diffusion equations:

$$\begin{aligned}\frac{\partial N_1}{\partial t} &= f_1(N_1, N_2) + D_1 \frac{\partial^2 N_1}{\partial x^2} \\ \frac{\partial N_2}{\partial t} &= f_2(N_1, N_2) + D_2 \frac{\partial^2 N_2}{\partial x^2}\end{aligned}\tag{4.1}$$

where  $D_i$  represents the diffusion constant for the  $i$ th species. We are interested in knowing whether a fixed point  $(N_1^*, N_2^*)$  which is stable in the mean-field case (when  $D_1 = D_2 = 0$ ) remains stable when diffusion takes place.

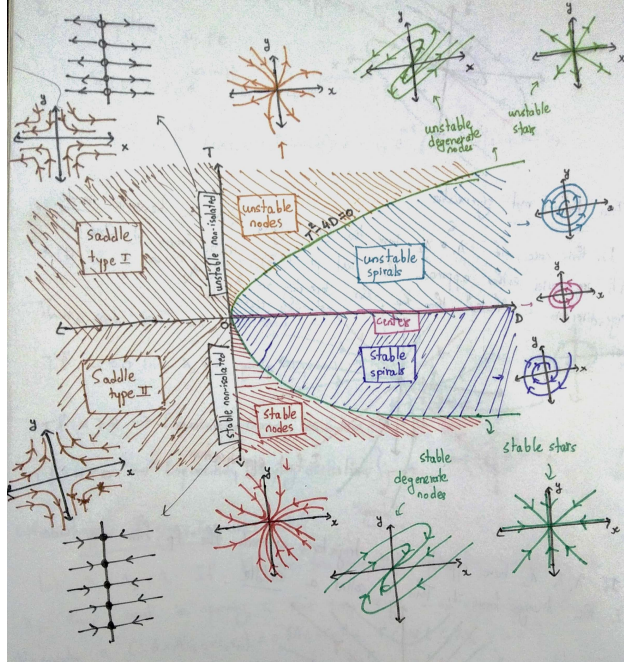
### 4.1 Linear stability analysis in the mean-field case

Recall that for any two-dimensional dynamical system of the form:

$$\begin{aligned}\dot{x} &= g(x, y) \\ \dot{y} &= h(x, y)\end{aligned}$$

with Jacobian  $\mathbf{J}$ , a fixed point  $(x^*, y^*)$  is stable iff:

$$\begin{aligned}\text{Tr}(\mathbf{J}|_{(x^*, y^*)}) &< 0 \\ \text{Det}(\mathbf{J}|_{(x^*, y^*)}) &> 0\end{aligned}$$



**Figure 4.1:** Classification of the fixed points of a linearized 2D system.  $T$  is the trace of the Jacobian and  $D$  is the determinant of the Jacobian of the system.

Substituting  $D_1 = D_2 = 0$  into equation (4.1), we see that in the mean-field case, the Jacobian is given by:

$$J = \begin{bmatrix} \frac{\partial f_1}{\partial N_1} & \frac{\partial f_1}{\partial N_2} \\ \frac{\partial f_2}{\partial N_1} & \frac{\partial f_2}{\partial N_2} \end{bmatrix}$$

Letting  $\alpha_{ij} = \left. \frac{\partial f_i}{\partial N_j} \right|_{(N_1^*, N_2^*)}$  for notational ease, we see that a fixed point  $(N_1^*, N_2^*)$  is stable iff:

$$\begin{aligned} \alpha_{11} + \alpha_{22} &< 0 \\ \alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21} &> 0 \end{aligned} \tag{4.2}$$

## 4.2 Linear stability analysis when diffusion is present

We will now consider a situation where  $D_1$  and  $D_2$  are not zero. Let  $(N_1^*, N_2^*)$  be a fixed point for equation (4.1). We consider a small perturbation  $(N_1, N_2) \rightarrow (N_1^* + \varepsilon_1, N_2^* + \varepsilon_2)$ . Plugging this into the expression for  $\frac{\partial N_1}{\partial t}$  in equation (4.1) yields:

$$\frac{\partial}{\partial t}(N_1^* + \varepsilon_1) = f_1(N_1^* + \varepsilon_1, N_2^* + \varepsilon_2) + D_1 \frac{\partial^2}{\partial x^2}(N_1^* + \varepsilon_1) \tag{4.3}$$

We can now Taylor expand the first term on the RHS as:

$$f_1(N_1^* + \varepsilon_1, N_2^* + \varepsilon_2) = \underbrace{f_1(N_1^*, N_2^*)}_{=0} + \varepsilon_1 \left. \frac{\partial f_1}{\partial N_1} \right|_{(N_1^*, N_2^*)} + \varepsilon_2 \left. \frac{\partial f_1}{\partial N_2} \right|_{(N_1^*, N_2^*)} + \dots$$

Substituting this back into equation (4.3) yields:

$$\frac{\partial \varepsilon_1}{\partial t} = D_1 \frac{\partial^2 \varepsilon_1}{\partial x^2} + \varepsilon_1 \left. \frac{\partial f_1}{\partial N_1} \right|_{(N_1^*, N_2^*)} + \varepsilon_2 \left. \frac{\partial f_1}{\partial N_2} \right|_{(N_1^*, N_2^*)} + \dots$$

Following the same procedure for  $\frac{\partial N_2}{\partial t}$  and neglecting 2nd order and higher terms gives us:

$$\begin{aligned} \frac{\partial \varepsilon_1}{\partial t} &= \alpha_{11}\varepsilon_1 + \alpha_{12}\varepsilon_2 + D_1 \frac{\partial^2 \varepsilon_1}{\partial x^2} \\ \frac{\partial \varepsilon_2}{\partial t} &= \alpha_{21}\varepsilon_1 + \alpha_{22}\varepsilon_2 + D_2 \frac{\partial^2 \varepsilon_2}{\partial x^2} \end{aligned} \tag{4.4}$$

### 4.3 Stability criteria for a homogeneous perturbation

If our perturbation is spatially homogeneous, then  $\frac{\partial^2 \varepsilon_1}{\partial x^2} = \frac{\partial^2 \varepsilon_2}{\partial x^2} = 0$ . Substituting these conditions into equation (4.4) yields a system of PDEs that is linear in  $\varepsilon_1$  and  $\varepsilon_2$ . Thus, the stability criterion for such a system is given by equation (4.2). We are therefore led to conclude that any fixed point that is stable in the mean-field case remains stable to spatially homogeneous perturbations.

### 4.4 Stability criteria for a periodic perturbation

Consider instead a perturbation that is periodic in space. Specifically, consider:

$$\begin{aligned} \varepsilon_1 &= a_{1k} \sin\left(\frac{k\pi x}{L}\right) \\ \varepsilon_2 &= a_{2k} \sin\left(\frac{k\pi x}{L}\right) \end{aligned} \tag{4.5}$$

For notational ease, let  $\beta = \frac{k\pi}{L}$ . We would like to see how the amplitudes  $a_{i_k}$  of the perturbations change with time. We would conclude that a fixed point is stable if perturbations die out with time. We will find out how perturbations change with time by substituting (4.5) back into (4.4). For  $i = 1, 2$ , we have:

$$\begin{aligned} \frac{\partial \varepsilon_i}{\partial t} &= \sin(\beta x) \frac{da_{i_k}}{dt} \\ D_i \frac{\partial^2 \varepsilon_i}{\partial x^2} &= -D_i a_{i_k} \beta^2 \sin(\beta x) = -D_i \beta^2 \varepsilon_i \end{aligned} \tag{4.6}$$

Substituting (4.5) and (4.6) into (4.4) gives us:

$$\begin{aligned}\frac{da_{1_k}}{dt} &= (\alpha_{11} - D_1\beta^2)a_{1_k} + \alpha_{12}a_{2_k} \\ \frac{da_{2_k}}{dt} &= \alpha_{21}a_{1_k} + (\alpha_{22} - D_2\beta^2)a_{2_k}\end{aligned}\tag{4.7}$$

or, in more compact notation:

$$\dot{\mathbf{a}} = \mathbf{J}_{\mathbf{p}} \mathbf{a}$$

where  $\mathbf{a} = \begin{bmatrix} a_{1_k} \\ a_{2_k} \end{bmatrix}$  and  $\mathbf{J}_{\mathbf{p}}$  is the Jacobian of the system, given by:

$$\mathbf{J}_{\mathbf{p}} = \begin{bmatrix} \alpha_{11} - D_1\beta^2 & \alpha_{12} \\ \alpha_{21} & \alpha_{22} - D_2\beta^2 \end{bmatrix}\tag{4.8}$$

This is a simple system of ODEs that is linear in  $a_{1_k}$  and  $a_{2_k}$ . Thus, as before, we require  $\text{Tr}(\mathbf{J}_{\mathbf{p}}) < 0$  and  $\text{Det}(\mathbf{J}_{\mathbf{p}}) > 0$  for stability. Thus, we require:

$$\alpha_{11} - D_1\beta^2 + \alpha_{22} - D_2\beta^2 < 0\tag{4.9}$$

$$(\alpha_{11} - D_1\beta^2)(\alpha_{22} - D_2\beta^2) - \alpha_{12}\alpha_{21} > 0\tag{4.10}$$

Let us examine these two criteria in detail.

#### 4.4.1 The Trace

From (4.2), we know that  $\alpha_{11} + \alpha_{22} < 0$ . Further, by definition, we have  $D_1 > 0$  and  $D_2 > 0$ . Thus, rearranging equation (4.9):

$$\alpha_{11} - D_1\beta^2 + \alpha_{22} - D_2\beta^2 = \underbrace{\alpha_{11} + \alpha_{22}}_{< 0} - \underbrace{(D_1 + D_2)\beta^2}_{> 0}$$

we see that (4.9) is always satisfied for a point that was stable in the mean-field equilibrium.

#### 4.4.2 The Determinant

For equation (4.10), we use (4.2) and make the following observation:

$$(\alpha_{11} - D_1\beta^2)(\alpha_{22} - D_2\beta^2) - \alpha_{12}\alpha_{21} = \underbrace{(\alpha_{11}\alpha_{22} - \alpha_{12}\alpha_{21})}_{> 0} + \underbrace{D_1D_2\beta^4}_{> 0} - \beta^2(\alpha_{22}D_1 + \alpha_{11}D_2)$$

Though the first two parts are always positive, the last term need not be positive. Thus, the determinant need not be positive, and stability in the mean field equilibrium here does *not* guarantee stability in the diffusive case (!). Diffusion can thus destabilize points that were previously stable when diffusion was disallowed. This sort of diffusion-induced destabilization is called ‘Turing instability’.

#### 4.4.3 Identifying sufficient criteria for stability

When, then, is stability of a fixed point guaranteed? Note that we only need to examine the effect of the last term of the previous equation for this. Thus, we require:

$$\beta^2(\alpha_{22}D_1 + \alpha_{11}D_2) < 0 \quad (4.11)$$

Now, since  $D_1 > 0$ , we are allowed to take out a factor of  $D_1$  from the addition term. Thus, the condition:

$$\alpha_{22} + \alpha_{11} \frac{D_2}{D_1} < 0 \quad (4.12)$$

is sufficient to ensure stability of the fixed point. What does this mean biologically? Recall that  $\alpha_{11}$  and  $\alpha_{22}$  represent the ‘self-regulation’ (or ‘autocatalytic’) rates of species 1 and 2 respectively. Consider the case in which  $\alpha_{22} < 0$  and  $\alpha_{11} > 0$ . In this case, for (4.12) to be satisfied, we require that  $\frac{D_2}{D_1}$  is small, i.e  $D_2 < D_1$ . On the other hand, if  $\alpha_{22} > 0$  and  $\alpha_{11} < 0$ , we would require that  $\frac{D_2}{D_1}$  is large, i.e  $D_2 > D_1$ .

Thus, equation (4.12) is telling us that for a MFE stable fixed point to remain stable, the species with a positive self-regulation ( $\alpha_{ii} > 0$ ) should have a high diffusion constant, whereas the species with negative self-regulation ( $\alpha_{ii} < 0$ ) should have a low diffusion constant. If the two species diffuse at comparable rates, i.e if  $D_1 \approx D_2$ , then this condition is always satisfied. To see this, we can substitute  $D_1 = D_2$  into equation (4.12) and recall from (4.2) that  $\alpha_{22} + \alpha_{11} < 0$  for any fixed point that is stable in the mean-field equilibrium.

#### 4.4.4 General criteria for Turing instability (I think?)

In general, letting  $\lambda = \beta^2$ , we see that the determinant  $\Delta$  is a quadratic in  $\lambda$  given by:

$$\Delta = D_1 D_2 \lambda^2 - (\alpha_{11} D_2 + \alpha_{22} D_1) \lambda + (\alpha_{11} \alpha_{22} - \alpha_{12} \alpha_{21}) \quad (4.13)$$

Since the trace of the Jacobian is always non-positive, the system can only move from stability to instability if we go from  $\Delta > 0$  to  $\Delta < 0$ . A few insights can be gained from this. For one, for a system to always be stable, it would need to have at most one real root for  $\lambda$  in equation (4.13). However, calculating the discriminant for this equation, we find:

$$\underbrace{(\alpha_{11} D_2 + \alpha_{22} D_1)^2}_{\geq 0} - 4 D_1 D_2 \underbrace{(\alpha_{11} \alpha_{22} - \alpha_{12} \alpha_{21})}_{< 0 \text{ by equation (4.2)}} > 0$$

Thus, if  $D_1 D_2 > 0$  (*i.e* one of the species is diffusing faster than the other), equation (4.13) always has at least one root, meaning that Turing instability can *always* be induced in such reaction diffusion systems by eliciting a suitable perturbation (*i.e* one with a suitable  $\lambda$ ). What kind of perturbation? Well, the switch from stability to instability occurs at  $\Delta = 0$ . In particular, at this point, we must have  $\frac{\partial \Delta}{\partial \lambda} = 0$ , which from equation (4.13) implies

$$\lambda = \frac{D_2 \alpha_{11} + D_1 \alpha_{22}}{2D_1 D_2}$$

$$\Rightarrow \frac{k\pi}{L} = \left( \frac{D_2 \frac{\partial f_1}{\partial N_1} + D_1 \frac{\partial f_2}{\partial N_2}}{2D_1 D_2} \right)^{\frac{1}{2}}$$

## Part III

# Evolutionary Game theory and adaptive dynamics

# Chapter 5

## Evolutionary game theory for mathematicians

Nature is full of situations in which the fitness of organisms is frequency-dependent, in the sense that the ‘fitness payoff’ to an individual adopting a ‘strategy’ depends on the strategies adopted by other individuals in the population. Evolutionary game theory tries to capture the dynamics of such systems using a class of models called ‘evolutionary games’. Below, I will present (without proof) some notions and results from evolutionary game theory. Proofs of the results can be found in Broom and Rychtář, 2022 and sources cited therein.

### 5.1 Games, strategies, and populations

**Definition.** (*Evolutionary Game*) An  $n$ -player evolutionary game is a collection  $\{S_1, \dots, S_n; g_1, \dots, g_n\}$  of sets  $S_i$  and functions  $g_i : (S_1 \times S_2 \times \dots \times S_n) \rightarrow \mathbb{R}$ .  $S_i$  is called the pure strategy set available to player  $i$ , and  $g_i : (x_1, \dots, x_n) \mapsto g_i(x_1, \dots, x_n)$  is called the payoff function for player  $i$  when player  $j$  is playing strategy  $x_j \in S_j$ . Unless specified otherwise, we will assume that every player has access to the same set of pure strategies, in which case the evolutionary game can be characterized by the collection  $(S, G)$ , where  $S$  is the pure strategy set and  $G : (S \times S \times \dots \times S) \rightarrow \mathbb{R}$  is the payoff function. Thus, the game  $(S, G)$  is modelling a situation in which every organism must choose some action from a set  $S$ , and the net benefit, or ‘payoff’, of every action in  $S$  is determined by the function  $G$ . We will in our discussion here restrict ourselves to the situation in which  $S$  is finite<sup>1</sup>. The finiteness of the pure strategy set is meant to model situations in which organisms have to ‘choose’ between  $m$  distinct, discrete alternatives. If  $S$  is a finite set  $S = \{S_1, \dots, S_m\}$ , we

---

<sup>1</sup>In general,  $S$  need not be finite or even countable, it only needs to be measurable. In this setting, we must be careful with assigning payoffs and probabilities, and generally proceed by identifying a suitable  $\sigma$ -algebra  $\mathcal{F}$  of  $S$  over which to analyze the game (for example, if  $S$  is also a topological space, the space of Borel sets of  $S$  is usually a good choice). Then, each mixed strategy is a probability measure  $\mu \in \mathbb{M}$  on  $(S, \mathcal{F})$ , payoffs are  $\mu$ -measurable real functions, and average payoffs are in terms of Lebesgue integrals over  $(S, \mathcal{F}, \mathbb{M})$ . This more general setting is tricky to handle and is likely overkill for our sake, so we stick to cases where  $S$  is a finite set and everything is ‘nice’.



say that the game is played on  $m$  ‘pure’ strategies. But what does this mean? What exactly does it mean to adopt a strategy?

**Definition.** (*Strategy*) Given a game with pure strategy set  $S = \{S_1, \dots, S_m\}$ , A *strategy* is defined as a probability vector  $p = (p_1, \dots, p_m)$  such that a player playing strategy  $p$  plays  $S_i \in S$  with probability  $p_i$ . A player playing strategy  $p$  is called a  $p$ -player. Note that given a set of  $m$  pure strategies, all strategies lie on the  $m$ -simplex  $\Delta^m = \{(x_1, \dots, x_m) : x_i > 0, \sum_{i=1}^m x_i = 1\}$ . This is sometimes called the strategy simplex.

**Definition.** (*Support of a strategy*) The *support* of a strategy  $p$  is the set of indices of non-zero elements of  $p$ , i.e  $\text{supp}(p) = \{i : p_i > 0\}$ .

**Definition.** (*Pure and mixed strategies*) Let  $p \in \Delta^m$  be a strategy. If  $|\text{supp}(p)| = 1$ , we say that  $p$  is a pure strategy, and otherwise, we say that  $p$  is a mixed strategy. If  $p$  uses every pure strategy (i.e  $p_i > 0 \forall i$ , or  $|\text{supp}(p)| = |S| = m$ ), we say that  $p$  is an internal strategy, since it is in the interior of  $\Delta^m$ .

In biology, we are generally interested in a population of individuals, all of which are playing some game. Thus, we are interested in the payoff for an individual in a particular population. This idea is made precise below.

**Definition.** (*Population*) We will use  $\delta_p$  to represent a population in which a randomly chosen player is almost surely a  $p$ -player. Note that we can write any population  $\Pi$  as  $\Pi = \sum \delta_{S_i} p_i$ , where each  $S_i$  is a pure strategy, and the probability of encountering an  $S_i$ -player is  $p_i$ . Unless otherwise specified, we will assume that populations are well-mixed and infinitely large.

We will use  $\mathbb{E}[p; \Pi]$  to denote the expected payoff of a  $p$ -player in a population  $\Pi$ . When modeling dynamics, this is usually assumed to be equivalent (up to a constant) to the fitness of  $p$ -players in a population  $\Pi$ .

The most well-known evolutionary games are matrix games, i.e games for which one can define a (fixed) payoff matrix  $A$  such that for any strategy  $p$  in any population  $\delta_q$ , the expected payoff can be written as  $\mathbb{E}[p; \delta_q] = qAp^T$ . Note that this is equivalent to assuming that individuals are engaged in pairwise contests. Furthermore, note that  $\mathbb{E}$  is linear in both  $p$  and  $q$ . However, in many biologically relevant games, one cannot make this assumption. A more general class of evolutionary games are the so-called ‘playing-the-field’ games, in which an organism is envisioned as playing against the population ‘as a whole’. This notion is made precise below.

**Definition.** (*Playing the field*) An evolutionary game is said to be a playing-the-field game if for any strategy  $p = (p_1, \dots, p_n)$  in any population  $\Pi$ , we can write

$$\mathbb{E}[p; \Pi] = \sum_{i=1}^n p_i f_i(\Pi)$$

where each  $f_i : \Delta^n \rightarrow \mathbb{R}$  can be any (fixed) nonlinear function.

Note that in playing-the-field games,  $\mathbb{E}$  is linear in the focal player's strategy  $p$ , but need not be linear in the population strategy  $q$ . It is also clear that every matrix game with payoff matrix  $A = (a_{ij})_{n \times n}$  is a playing-the-field game, with  $f_i(q) = \sum_{j=1}^n a_{ji}q_j = (A^T q^T)_i$ . Playing-the-field games are thus one of the two natural ways to try and generalize matrix games to model more varied biological phenomena<sup>2</sup>

## 5.2 Nash Equilibria, ESSs, and associated results

I will now define two important notions which will characterize the types of strategies we are interested in.

**Definition.** (*Nash Equilibrium*) Consider a game on  $n$  strategies. A strategy  $p^*$  is said to be a *Nash equilibrium* for the game if:

$$\mathbb{E}[p^*; \delta_{p^*}] \geq \mathbb{E}[q; \delta_{p^*}] \quad \forall q \neq p^* \in \Delta^n$$

This definition is saying that if  $p^*$  is a Nash equilibrium, then no  $q$ -player can do better than a  $p^*$ -player in a population of (almost surely)  $p^*$ -players, unless  $q = p^*$ . However, in evolutionary game theory, this is not a strong enough condition. We are generally interested in a stronger notion that allows for the possibility of small ‘invasions’ of other ‘mutant’ strategies that occur with *non-zero* probability.

**Definition.** (*ESS*) Consider a game on  $n$  strategies. A strategy  $p^*$  is said to be an *evolutionarily stable strategy (ESS)* for the game if  $\forall q \neq p^* \in \Delta^n \exists \epsilon_q > 0$  such that:

$$\mathbb{E}[p^*; (1 - \epsilon)\delta_{p^*} + \epsilon\delta_q] > \mathbb{E}[q; (1 - \epsilon)\delta_{p^*} + \epsilon\delta_q] \quad \forall \epsilon \in (0, \epsilon_q)$$

Thus, if  $p^*$  is an ESS, it is *strictly* better than any other strategy  $q$  when the population is comprised mostly of  $p^*$ -players, with some sufficiently small fraction  $\epsilon$  of ‘mutant’  $q$  players. Note that given any ESS  $p^*$  and any strategy  $q \neq p^*$ , by taking the limit  $\epsilon \rightarrow 0$  in the definition of an ESS, we see that  $\mathbb{E}[p^*; \delta_{p^*}] \geq \mathbb{E}[q; \delta_{p^*}]$ . Thus, if a strategy is an ESS, it must also be a Nash equilibrium.

Below, I will list some results pertaining to ESSs.

**Theorem 5.2.1.** (Characterization) *Let  $h_{p,q,u}$  be the function<sup>3</sup>:*

$$h_{p,q,u} = \mathbb{E}[p; (1 - u)\delta_p + u\delta_q] - \mathbb{E}[q; (1 - u)\delta_p + u\delta_q]$$

---

<sup>2</sup>the other being to make payoffs linear in the population strategy but not the player strategy. My admittedly limited understanding is that this latter approach has found lesser applicability

<sup>3</sup> $h_{p,q,u}$  is often called the ‘incentive function’, since it can be interpreted as the ‘incentive’ of switching from strategy  $q$  to strategy  $p$  in a population where a fraction  $u$  of the individuals are  $q$ -players and the rest are  $p$ -players.

If  $h_{p,q,u}$  is differentiable with respect to  $u$  at  $u = 0$  for every  $p, q \in \Delta^n$ , then  $p$  is an ESS iff for every  $q \neq p$ , the following are true<sup>4</sup>:

1.  $\mathbb{E}[p; \delta_p] \geq \mathbb{E}[q; \delta_p]$

2. If  $\mathbb{E}[p; \delta_p] = \mathbb{E}[q; \delta_p]$ , then

$$\left. \frac{\partial h_{p,q,u}}{\partial u} \right|_{u=0} > 0$$

How else can we identify ESSs? The following theorem examines how an ESS behaves when playing against strategies that are ‘very close’ to the ESS.

**Theorem 5.2.2.** (Local superiority) *Consider a playing-the-field game on  $n$  strategies.  $p$  is an ESS for this game iff  $\exists \epsilon > 0$  s.t.*

$$\mathbb{E}[p; \delta_q] > \mathbb{E}[q; \delta_q] \quad \forall q \neq p \in B_n(p, \epsilon) \subseteq \Delta^n$$

where  $B_n(p, \epsilon)$  is the  $n$ -dimensional ball with radius  $\epsilon$  centered at  $p$ .

This theorem says that in the case of playing-the-field games,  $p$  is an ESS if and only if it is ‘locally superior’, in the sense that if one considers any strategy  $q$  that is sufficiently similar to  $p$ , then  $p$  performs (strictly) better than  $q$  in a population of  $q$ -players. Sometimes, finding several ESSs is much easier if you already know that a certain strategy must be an ESS. In the case of matrix games (*i.e* games for which you can write down a constant payoff matrix), the following theorem is extremely useful:

**Theorem 5.2.3.** (Bishop-Cannings theorem) *Consider a matrix game with pure strategies  $S_1, \dots, S_n$ . Let  $p$  be an ESS for this game. Let  $T(p)$  be the set of indices of all pure strategies which have the same payoff against  $p$  as  $p$  has against itself, *i.e*  $T = \{i : \mathbb{E}[S_i, p] = \mathbb{E}[p, p]\}$ . Then, if  $q$  is any strategy such that  $\text{supp}(q) \subseteq T(p)$ ,  $q$  is not an ESS for the matrix game.*

Thus, given a single ESS, the Bishop-Cannings theorem dictates that certain other strategies cannot also be ESSs. This greatly reduces the total number of possible ESSs in matrix games, and thus helps analyze such models. For example, an immediate and powerful corollary of the Bishop-Cannings theorem is the following:

**Corollary 5.2.1.** (Uniqueness of the internal ESS) *If an internal strategy  $p$  is an ESS for a matrix game, then it is the only ESS for that game.*

Thus, in the case of matrix games, it is sufficient to find a *single* internal ESS, since the Bishop-Cannings theorem will then ensure that this ESS is unique.

---

<sup>4</sup>condition 1 is often called the ‘equilibrium condition’, and condition 2 is often called the ‘stability condition’. Maynard Smith, working only with matrix games, originally defined the notion of an ESS using these two criteria, with the stability condition as  $\mathbb{E}[p; \delta_p] = \mathbb{E}[q; \delta_p] \Rightarrow \mathbb{E}[p; \delta_q] > \mathbb{E}[q; \delta_q]$ . This characterization in terms of the incentive function shows that our definition in terms of  $\epsilon$ s generalizes his original definition.

## 5.3 Replicator dynamics and the folk theorem

Often, we are not interested in particular strategies, but are instead interested in predicting the fate of a given strategy in a given population of individuals. For any game, if one assumes that the fitness of a strategy in a population is given by its expected payoff, then the population dynamics are given by the associated ‘replicator equation’ on  $\Delta^n$ :

$$\dot{p}_i = p_i(\pi_i(\mathbf{p}) - \bar{\pi}(\mathbf{p})) \quad (5.1)$$

where  $\mathbf{p}(t) = (p_1(t), \dots, p_n(t)) \in \Delta^n$  is the strategy vector that we are tracking,  $\pi_i(\mathbf{p}) = \mathbb{E}[S_i; \sum_{i=1}^n p_i \delta_{S_i}]$  is the expected payoff of an  $S_i$ -player in a population where the probability of encountering an  $S_i$ -player is  $p_i$ , and  $\bar{\pi}(\mathbf{p}) = \sum_{i=1}^n p_i \pi_i(\mathbf{p})$  is the average payoff/fitness of the population.

Equation (5.1) defines a system of  $n$  coupled ODEs, which has always has a unique solution<sup>5</sup> given an initial condition  $\mathbf{p}(0) = \mathbf{p}_0$ . Biologically, we are interested in stable fixed points, *i.e.* points  $\mathbf{p}^*$  at which  $\dot{p}_i^* = 0 \forall i$  and trajectories which are ‘close enough’ to  $\mathbf{p}^*$  converge to  $\mathbf{p}^*$ . If we were to find such points, then we could reason that populations which contain such a mix of individuals will maintain them forever, even in the face of small perturbations from the fixed point (due to, say, mutations or invasions from other populations). In the case of matrix games, the following theorems prove very useful and provide a connection between replicator dynamics and the notions introduced in section 5.2:

**Theorem 5.3.1.** (Folk theorem of evolutionary game theory) *If  $p$  is an ESS for a matrix game, then  $p$  is also a locally asymptotically stable fixed point for the replicator dynamics associated with that game. Further, if  $p^*$  is any fixed point for the replicator equations such that it is also the limit of at least one interior trajectory<sup>6</sup>, then  $p^*$  must be a Nash equilibrium for that game. In particular, every stable fixed point for the replicator dynamics is a Nash equilibrium for the game.*

**Theorem 5.3.2.** (Zeeman, 1980) *If  $p$  is an internal ESS for a matrix game, then it is a global attractor for the associated replicator equation in  $\text{int}(\Delta^n)$ , *i.e.* every trajectory that starts within the interior of  $\Delta^n$  converges to  $p$ .*

Thus, in the case of matrix games, these theorems along with the Bishop-Cannings theorem tell us that identifying ESSs and Nash equilibria lets us characterize almost all the interesting properties of the game that we are interested in. This is why ESSs have played such a central role in evolutionary game theory. The first part of the Folk theorem (ESS  $\Rightarrow$  stable fixed point) also holds for playing-the-field games (though the other results presented here do not). It is thus desirable to seek out ESSs even when trying to analyze a playing-the-field game.

<sup>5</sup>Assuming all functions are ‘nice’: Specifically, we require that each  $\pi_i$  is Lipschitz continuous in  $\Delta^n$ . In almost all practical cases, this will be satisfied.

<sup>6</sup>*i.e.*  $\exists p_0 \in \text{int}(\Delta^n)$  s.t.  $p(0) = p_0 \Rightarrow \lim_{t \rightarrow \infty} p(t) = p^*$ . Alternatively, it is also sufficient if the fixed point  $p^*$  is instead Lyapunov stable

## 5.4 An example: The ideal free distribution as an evolutionary game

The ideal free distribution was first formulated by Fretwell and Lucas in 1970 to study the distribution of birds in patchy environments. Fretwell and Lucas considered an environment with  $n$  patches of unequal resources, where the  $i^{\text{th}}$  patch had resources of value  $B_i$ . Without loss of generality, we can renumber the patches such that  $B_1 > B_2 > \dots > B_n$ . They further assumed that the value of the resource to each individual in the  $i^{\text{th}}$  patch was dependent on the interspecific competition on the  $i^{\text{th}}$  patch, which they modeled using a continuous increasing function  $f_i(x_i)$ , where  $x_i$  is the density of individuals on the  $i^{\text{th}}$  patch, and  $f_i(0) = 0$ . Then, the net of the  $i^{\text{th}}$  patch when the  $i^{\text{th}}$  patch has a population density of  $x_i$  is given by:

$$R_i(x_i) = B_i - f_i(x_i) \quad (5.2)$$

They further assumed that individuals had complete knowledge of the value of every patch, that individuals were free to move from patch to patch with no cost, and that individuals moved between patches in order to maximize the net benefit  $R_i$ . Under these conditions, Fretwell and Lucas reasoned as follows: For any given organism in a given patch, if moving to any other patch would have been better for this individual, it would immediately do so. Thus, the only way to attain an ‘equilibrium’ is if the organism *cannot do any better* in any other patch than it is currently doing in the patch that it is on. Further, this should be true for every organism, and thus in every patch. This leads us to the conclusion that we require  $R_1 = R_2 = \dots = R_n$  if every patch is occupied. If only some number  $l < n$  of patches are occupied at equilibrium, then these patches must be the ones with the  $l$  greatest values of  $B_i$ , and we instead require  $R_1 = R_2 = \dots = R_l \geq B_{l+1} \geq B_{l+2} \geq \dots \geq B_n$ , since no patches past the  $l^{\text{th}}$  patch are occupied and patches are arranged in descending order of  $B_i$ . They dubbed such a distribution the ‘ideal free distribution’ (IFD), and were able to study some properties of such a distribution.

### 5.4.1 The game-theoretic perspective: A habitat selection game

This idea of a scenario in which no organism can do better by moving to a different patch should immediately have you thinking of definition 5.2. In fact, even though Fretwell and Lucas were unaware of this, it was quickly established that the IFD is likely a Nash Equilibrium for a cleverly defined game. In the early 2000s, Cressman, Krivan, and other game theorists defined the following ‘habitat selection game’:

**Definition.** (*Habitat selection game*) Define a game on  $n$  pure strategies where the  $i^{\text{th}}$  pure strategy  $S_i$  is to stay on the  $i^{\text{th}}$  patch, and where the payoff of the  $i^{\text{th}}$  pure strategy in a population  $(x_1, \dots, x_n)$  given by  $R_i = B_i - f_i(x_i)$ . This game is called the ‘habitat selection game’.

In the habitat selection game, the expected payoff of a strategy  $p$  in a population  $\Pi = \sum q_i \delta_{S_i}$

is given by:

$$\mathbb{E}[p; \Pi] = \sum_{i=1}^n p_i (B_i - f_i(q_i)) \quad (5.3)$$

Note that this is linear in  $p$  (but not necessarily linear in  $\Pi$ ), and the habitat selection game is thus a playing-the-field game. In this new game-theoretic framework, we have the following definition:

**Definition.** (*IFD strategy*) A strategy  $p$  for the habitat selection game on  $n$  pure strategies is said to be an IFD strategy if there exists  $l \leq n$  such that:

1.  $p_i > 0 \iff i \leq l$
2.  $R_1 = R_2 = \dots = R_l \geq B_{l+1} \geq \dots \geq B_n$

This puts the IFD that Fretwell and Lucas recognized in game-theoretic terms. Though the game-theoretic aspect of the original IFD as proposed by Fretwell and Lucas was implicitly recognized for a long time, it was only in 2006 that the habitat selection game was defined, and it was proved both that the IFD is in fact an ESS for the habitat selection game, and that *any* ESS for the habitat selection game must be an IFD strategy. We state and prove this as a theorem below.

**Theorem 5.4.1.** (Cressman *et al.*, 2006) *A strategy  $p$  is an ESS for the habitat selection game iff it is an IFD strategy.*

*Proof.* Assume that  $p = (p_i)_i$  is an ESS for the habitat selection game. We will prove that it is an IFD, *i.e.* that it satisfies both criteria listed in definition 5.4.1.

To see that it satisfies the first criterion, assume the contrary, *i.e.* that there exist  $j < k \leq n$  such that  $p_j = 0$  and  $p_k > 0$ . Then,

$$R_j = B_j - 0 = B_j \geq B_k > B_k - f_k(p_k) = R_k \quad (5.4)$$

Define  $\tilde{p}$  as:

$$\tilde{p}_i = \begin{cases} p_i & i \neq j, k \\ p_k & i = j \\ p_j & i = k \end{cases}$$

Thus,  $p_i = \tilde{p}_i$  everywhere except at  $j$  and  $k$ , where the values are switched. The incentive function  $h_{p, \tilde{p}, 0}$  is given by:

$$h_{p, \tilde{p}, 0} = \mathbb{E}[p; \delta_p] - \mathbb{E}[\tilde{p}; \delta_p]$$

Using equation (5.3), we get:

$$\begin{aligned} h_{p, \tilde{p}, 0} &= \sum_{i=1}^n (p_i - \tilde{p}_i)(R_i) \\ &= -\tilde{p}_j R_j + p_k R_k \end{aligned}$$

But since  $\tilde{p}_j = p_k$ , using equation (5.4), we get:

$$h_{p, \tilde{p}, 0} = p_k(R_k - R_j) < 0$$

which contradicts the necessary part of theorem 5.2.1, thus yielding a contradiction. If payoffs of the occupied patches are unequal, *i.e* if there exist  $j, k \leq l$  such that  $R_j > R_k$ , we can similarly define

$$\tilde{p}_i = \begin{cases} p_i & i \neq j, k \\ p_j + p_k & i = j \\ 0 & i = k \end{cases}$$

to yield the same contradiction, by following the same steps. Lastly, if  $l$  patches are occupied in the population and there are  $j \leq l < k$  with  $R_l < B_k$ , we can define

$$\tilde{p}_i = \begin{cases} p_i & i \neq j, k \\ p_k & i = j \\ p_j & i = k \end{cases}$$

to arrive at a contradiction through the same steps, leading us to the conclusion that  $R_1 = \dots = R_l \geq B_{l+1} \dots B_n$ . We have thus proved that every ESS is an IFD strategy. We will now prove the converse.

Assume that  $p = (p_i)_i$  is an IFD strategy with  $l$  patches occupied. We will use the local superiority property to prove that  $p$  is an ESS. Consider any strategy  $q \neq p$ .

**Case 1.** All patches are occupied in the IFD (*i.e*  $l = n$ )

Let  $R_1 = \dots = R_n = R^*$ . We have:

$$\sum_{i=1}^n (p_i - q_i)(B_i - f_i(p_i)) = \sum_{i=1}^n (p_i - q_i)R_i(p) \quad (5.5)$$

$$\begin{aligned} &= R^* \sum_{i=1}^n (p_i - q_i) \\ &= R^* \left( \sum_{i=1}^n p_i - \sum_{i=1}^n q_i \right) \\ &= R^*(1 - 1) = 0 \end{aligned} \quad (5.6)$$

Now, consider the quantity  $\mathbb{E}[p; \delta_q] - \mathbb{E}[q; \delta_q]$ . Subtracting the LHS of (5.6) from

this, we have:

$$\begin{aligned}
\mathbb{E}[p; \delta_q] - \mathbb{E}[q; \delta_q] &= \sum_{i=1}^n (p_i - q_i)(B_i - f_i(q_i)) \\
&= \sum_{i=1}^n (p_i - q_i)(B_i - f_i(q_i)) - \sum_{i=1}^n (p_i - q_i)(B_i - f_i(p_i)) \\
&= \sum_{i=1}^n (p_i - q_i)(f_i(p_i) - f_i(q_i)) \tag{5.7}
\end{aligned}$$

By definition, each  $f_i$  is assumed to be a strictly increasing function. It is easy to see that this implies that  $(p_i - q_i)(f_i(p_i) - f_i(q_i)) > 0 \forall i$ . Using this in (5.7), we see that  $\mathbb{E}[p; \delta_q] - \mathbb{E}[q; \delta_q] > 0 \forall q \neq p$ . Thus,  $p$  is locally superior, and by the necessary part of theorem 5.2.2, it must be an ESS.

**Case 2.** Some patches are unoccupied in the IFD (*i.e*  $l < n$ )

Let  $R_1 = \dots = R_l = R^*$ . As before, we examine the quantity:

$$\begin{aligned}
\sum_{i=1}^n (p_i - q_i)(B_i - f_i(p_i)) &= \sum_{i=1}^l (p_i - q_i)(B_i - f_i(p_i)) + \sum_{i=l+1}^n (p_i - q_i)(B_i - f_i(p_i)) \\
&= \sum_{i=1}^l (p_i - q_i)R^* - \sum_{i=l+1}^n q_i B_i \tag{5.8}
\end{aligned}$$

Where the last equality comes from the definition of an IFD strategy on  $l$  patches as having  $p_i > 0 \iff i \leq l$  and the definition of each  $f_i$  as satisfying  $f_i(0) = 0$ . Using the fact that  $R^* \geq B_k \forall k > l$  in 5.8, we now get:

$$\begin{aligned}
\sum_{i=1}^n (p_i - q_i)(B_i - f_i(p_i)) &\geq \sum_{i=1}^l (p_i - q_i)R^* - \sum_{i=l+1}^n q_i R^* \\
&= R^* \left( \sum_{i=1}^n p_i - \sum_{i=1}^n q_i \right) \\
&= R^*(1 - 1) = 0
\end{aligned}$$



We can now proceed as in case 1 and write

$$\begin{aligned}
\mathbb{E}[p; \delta_q] - \mathbb{E}[q; \delta_q] &= \sum_{i=1}^n (p_i - q_i)(B_i - f_i(q_i)) \\
&> \sum_{i=1}^n (p_i - q_i)(B_i - f_i(q_i)) - \sum_{i=1}^n (p_i - q_i)(B_i - f_i(p_i)) \\
&= \sum_{i=1}^n (p_i - q_i)(f_i(p_i) - f_i(q_i)) \\
&\geq 0 \quad \forall q \neq p \quad (\because \text{every } f_i \text{ is increasing})
\end{aligned}$$

and thus,  $p$  is locally superior and therefore an ESS by theorem 5.2.2. This concludes our proof.

□

Thus, for the habitat selection game as defined above, every ESS is an IFD strategy, and every IFD strategy is an ESS. It is also possible to prove that for the game defined above, there is a single unique ESS (and thus IFD). The habitat selection game and the IFD concept have been extended in various directions, such as to include multiple species (each with a different  $B_i$  and  $f_i$  set), migration dynamics, costs to migration, and incomplete information.

# Chapter 6

## Adaptive Diversification

Adaptive diversification refers to the phenomenon an initially monomorphic population evolving to become polymorphic due to (disruptive) frequency-dependent selection induced by ecological interactions (Doebeli, 2011). The framework has been widely used in evolutionary ecology and evolutionary game theory, especially in the context of sympatric speciation (Dieckmann and Doebeli, 1999). Models of adaptive diversification can be broadly classified into three classes of models which differ starkly in their approach and assumptions (Doebeli, 2011). I introduce all three approaches below.

### 6.1 The ‘classic’ adaptive dynamics approach

The classic approach to modelling adaptive diversification was first articulated in the late 90s by J.J. Metz’s group (Geritz et al., 1998). The crux of the approach relies on the observation that evolutionary stability (in the ESS sense of evolutionary game theory) and asymptotic convergence stability (in the sense of dynamical systems) need not always coincide.

More concretely, consider an infinitely large asexual population that is monomorphic for some quantitative trait such that every individual of the population has the trait value  $x \in \mathcal{T} \subseteq \mathbb{R}^n$ . We are interested in the dynamics of  $x$  over evolutionary time. There is assumed to be a separation of timescales between ecology and evolution such that whenever we observe the population, it is at ecological equilibrium (This is equivalent to a strong-selection + weak-mutation setting). The evolutionary dynamics of the trait value in the population are modelled as following the equation:

$$\frac{dx}{dt} = g(x) = B(x) \left( \nabla_y f(y; x) \Big|_{y=x} \right) \quad (6.1)$$

Here,  $B(x)$  describes the mutational process of the trait and is intended to model mutational biases. For the sake of simplicity, I will assume that  $B(x) = 1$  below, but the essential results are not greatly changed by more complicated forms.  $f(y; x)$  is the *invasion fitness* function, and describes the expected fitness of a mutant type  $y$  appearing in a population that is

monomorphic  $x$ -valued.  $x$  is typically called the ‘resident’ trait value. Thus, evolution is modelled as following the gradient  $\nabla_y f(y; x)$  of the invasion fitness, a quantity sometimes called the selection gradient. (6.1) can be interpreted to mean that at each time step, the population ‘samples’ all local mutations, and if a mutant can invade this resident population, then the trait rapidly spreads in the population, and by the time we next ‘observe’ it, the entire population has adopted this mutant trait. Equation (6.1) is called the *canonical equation of adaptive dynamics*. Fixed points for this equation, *i.e* points  $x^*$  for which  $g(x^*) = 0$ , are called *evolutionary singularities*. These singularities can be characterized by two different stability notions, as we will see below, and the difference between the two notions captures the important phenomenon of *evolutionary branching*, which we will encounter shortly.

**Definition.** (*Convergence stability*) A singularity  $x^*$  is said to be convergent stable (CS) if it is an asymptotically stable fixed point for equation (6.1).

Thus, singularities which are convergence stable are local attractors, in the sense that nearby populations will evolve towards this state.

**Definition.** (*Evolutionary stability*) A singularity  $x^*$  is said to be evolutionarily stable (ES) if it cannot be invaded by any nearby mutants.

In the one dimensional case, evolutionary stability requires that the invasion fitness be such that no mutant can invade the population. Since  $g(x^*) = \frac{\partial}{\partial y} f(y; x^*) = 0$  by definition at a singularity, the condition for evolutionary stability is controlled by the second derivative, and is given by:

$$\left. \frac{\partial^2}{\partial y^2} f(y; x) \right|_{y=x=x^*} < 0 \quad (6.2)$$

On the other hand, from elementary non-linear dynamics, we know that convergence stability requires  $\frac{dg}{dx} < 0$ , *i.e*:

$$\left. \frac{\partial^2}{\partial x \partial y} f(y; x) \right|_{y=x=x^*} + \left. \frac{\partial^2}{\partial y^2} f(y; x) \right|_{y=x=x^*} < 0 \quad (6.3)$$

Note that neither (6.2) nor (6.3) imply the other, and thus, we arrive at the following classification of evolutionary singularities:

	ES	not ES
CS	Ⓐ	Ⓑ
not CS	Ⓒ	Ⓓ

Points which are not convergent stable are not of interest to us because populations can only attain such a state if they begin there. Thus, Ⓒ and Ⓓ can be ignored for our

purposes.<sup>1</sup>

Points of type (A) are both evolutionarily stable and convergent stable. Such points represent ‘endpoints’ for evolution, since populations are attracted to such points and also cannot evolve away from them since they cannot be invaded by any nearby mutants. Points of type (B) are CS but not ES. Populations are attracted to such points, but once they have arrived, they are susceptible to invasion by nearby mutants. Let  $x^*$  be such a point, and let  $x_1 < x^* < x_2$  be two points in the immediate vicinity of  $x^*$ .

**Claim.** (Protected Polymorphism) Each of  $x_1$  and  $x_2$  can invade the other

*Proof.* We can Taylor expand the invasion fitness function as

$$f(x_1; x_2) = \underbrace{f(x_2; x_2)}_{=0} + \underbrace{(x_1 - x_2)}_{<0} \underbrace{\frac{\partial f}{\partial x_1}(x_1; x_2)}_{=g(x_2)} \bigg|_{x_1=x_2} + \frac{(x_1 - x_2)^2}{2} \frac{\partial^2 f}{\partial x_1^2}(x_1; x_2) \bigg|_{x_1=x_2}$$

where we have neglected terms that are higher than second order. Since  $x^*$  is convergent stable,  $\frac{dg}{dx} \bigg|_{x=x^*} < 0$  and we therefore see that  $g$  is a decreasing function of  $x$  in the immediate vicinity of  $x^*$ . Thus, since  $x_2 > x^*$ , we must have  $g(x_2) < g(x^*) = 0$ , and we can conclude that the second term in the RHS must be positive. Lastly, since  $x^*$  is evolutionarily unstable, we must have  $\frac{\partial^2 f}{\partial x_1^2}(x_1; x^*) \bigg|_{x_1=x^*} > 0$  by the stability criterion. Thus, assuming  $f$  is smooth,

for  $x_2$  sufficiently close to  $x^*$ , it must be the case that  $\frac{\partial^2 f}{\partial x_1^2}(x_1; x_2) \bigg|_{x_1=x_2} > 0$ . Thus, the third term of the RHS is also positive, meaning that  $f(x_1; x_2) > 0$ , and that  $x_1$  can thus invade  $x_2$ . An exactly analogous argument shows that  $f(x_2; x_1) > 0$ , thus completing the proof.  $\square$

Mutual invasibility of  $x_1$  and  $x_2$  results in a so-called ‘protected polymorphism’ in which the population harbours some members with trait value  $x_1$  and some members with trait value  $x_2$ . This can be shown in some cases to lead to further divergence where the polymorphisms grow further apart in trait space while both being maintained in the population. Thus, the population appears to have ‘branched’ from a monomorphic state to a dimorphic state in trait space. Due to this reason, points of type (B) are called *branching points*, and the population is said to have undergone *evolutionary branching* once it has gone from a monomorphism to a dimorphism. Note that once a population has branched, equation (6.1) no longer describes the population, since it is no longer monomorphic - We instead need a system of *two* equations, one for each morph.

In adaptive dynamics, the name of the game is thus formulating reasonable guesses for  $B(x)$

---

<sup>1</sup>Singularities of type (C) are sometimes called ‘garden of Eden’ points, since a population that begins at such a point will remain there and cannot be invaded by nearby mutants, but if a population does not begin there (or is somehow driven out by external forces), it can never find its way back to the singularity.

and  $f(y; x)$ . For example, one common choice for modelling asexual resource competition is  $B(x) \equiv 1$  and the invasion fitness function:

$$f(y : x) = 1 - \frac{\alpha(x, y)K(x)}{K(y)} \quad (6.4)$$

where  $\alpha(x, y)$  is called the *competition kernel* and  $K(x)$  is called the *carrying capacity function* (formulated in analogy with Lotka-Volterra competition or the logistic equation). Invasion fitnesses are sometimes derived from more mechanistic processes (such as explicitly formulating birth/death functions), but the conceptual idea is the same regardless of how complicated these functions may be, and evolutionary branching easily arises in a very wide range of ecological scenarios as long as the frequency-dependence of the selective force is strong enough (Doebeli and Dieckmann, 2000; Doebeli, 2011).

## 6.2 The PDE approach

A slightly more general approach to adaptive diversification is to relax the assumption of separation of timescales while keeping the assumption of infinite population size. In this case, we instead wish to formulate an equation for the distribution  $\phi(u)$  of trait values in  $\mathcal{T}$ . We thus seek PDEs of the form:

$$\frac{\partial \phi(u)}{\partial t} = F(u, \phi(u), t)$$

In PDE models, diversification shows up as patterns (in the Turing sense, as covered in Chapter 4), and the existence of a polymorphism corresponds to a multimodal solution for  $\phi(x)$ . The functional form of  $F$  is often formulated through biological principles. For example, in analogy to the logistic equation, one could postulate the continuous version:

$$\frac{\partial \phi(u)}{\partial t} = r\phi(u) \left( 1 - \frac{\phi(u)}{K(u)} \int_{\mathcal{T}} \alpha(u, v)\phi(v)dv \right) \quad (6.5)$$

where  $r$  is a growth rate,  $K(u)$  gives the carrying capacity of the environment for individuals with phenotype  $u$ , and  $\alpha(u, v)$  is a competition kernel which determines the effect of an individual with phenotype  $v$  on the growth of an individual with phenotype  $u$ , and the integral thus gives a measure of the effective density experienced by individuals with phenotype  $u$ .  $\alpha(u, v)$  is generally chosen such that strength of competition decreases with phenotypic distance (for example,  $\alpha(u, v) = \exp(|u - v|)$ ), in analogy with niche partitioning. Note that this is an integrodifferential equation, and as such, is usually not easy to solve for most functional forms of  $\alpha(u, v)$ . We thus need to resort to numerical methods to solve such equations. PDE models can also incorporate space more easily than the classic adaptive dynamics approach. For example, Doebeli *et al.* propose a spatial model of resource competition in

which we track the density  $\phi(x, u)$  of  $u$ -phenotype individuals at the location  $x$  given by the PDE:

$$\begin{aligned} \frac{\partial \phi(x, u)}{\partial t} = r \left( \int_{\mathcal{T}} B(v) \phi(x, v) dv - \frac{\phi(x, u)}{K(x, u)} \int_S \int_{\mathcal{T}} \alpha_p(u, v) \alpha_s(x, y) \phi(y, v) dv dy \right) \\ + m \left( \int_S D(x, y) \phi(y, u) dy - \phi(x, u) \right) \end{aligned} \quad (6.6)$$

where  $B(v)$  is a birth kernel that specifies births and mutations,  $K(x, u)$  is a carrying capacity function that varies with both phenotype and geographic location,  $\alpha_p(u, v)$  describes how strength of competition varies with phenotypic distance,  $\alpha_s(x, y)$  describes how strength of competition varies with spatial (geographic) distance,  $m$  is a migration rate, and  $D(x, y)$  is a dispersion kernel that describes the probability that an individual at location  $y$  will migrate into location  $x$ .

PDE models sacrifice interpretability for increased generality. While these models can include space as well as incorporate polymorphic populations more easily, the equations themselves generally have to be solved numerically, and it is more difficult to gain understanding as to why a certain behavior is observed. Nevertheless, mathematically sophisticated techniques have shown that multimodality remains a robust phenomenon for a very wide class of PDE models (Elmhirst et al., 2008; Doebeli, 2011).

### 6.3 Stochastic individual-based models

The most general approach to modelling adaptive diversification is through individual-based models. In such models, birth rates, death rates, and interaction rules are specified for each individual, and system-level properties are observed by letting the system evolve. Tools such as the Gillespie algorithm are useful here in order to effectively implement simulations. Here, one can include all sorts of complications if they want, and view the outcome. This affords flexibility that PDE models and the classic invasion fitness approach do not, but also leads to even less interpretability. In theory, IbMs can be analyzed as stochastic processes using the ideas covered in chapter 1.4, though depending on the model, the mathematics could get very complicated. As such, depending on complexity, in many cases, IbMs may be best viewed (in my opinion) as *in silico* ‘experiments’.

### 6.4 An example of a prediction of adaptive diversification

Consider the continuous version of the logistic equation given by the invasion fitness (6.4) in adaptive dynamics and the PDE (6.5) for PDE models. Assume that we have  $\alpha(u, v) =$

$\exp\{-(u-v)^2/(2\sigma_\alpha^2)\}$  and  $K(u) = K_0 \exp\{-(u)^2/(2\sigma_K^2)\}$ , *i.e.* Gaussian carrying capacity and Gaussian competition kernel. Then, it is known that diversification fails (*i.e.* the population remains monomorphic) if and only if:

$$\sigma_K < \sigma_\alpha \tag{6.7}$$

. This result arises from basic calculus using adaptive dynamics, but what does it mean? I argue here that there are two distinct biological interpretations to be made, which I present below:

### 6.4.1 Habitat Filtering

One way that 6.7 can be satisfied is if  $\sigma_K$  is very small. Biologically, this means that the environment is such that only some very particular morphs are viable, and the limit where  $\sigma_K = 0$  corresponds to a case where only a single morph is environmentally viable. Ecologists are well-acquainted with this notion, and refer to it as ‘habitat filtering’, the phenomenon in which the habitat itself ‘selects’ for certain phenotypes due to particular limiting abiotic factors.

### 6.4.2 Competition and character displacement

A second way to satisfy 6.7 is if  $\sigma_\alpha$  is very large. Biologically, this means that competition occurs across a wide range of phenotypes, and the limit where  $\sigma_\alpha = \infty$  corresponds to frequency-independent competition. In other words, diversification can fail if competition cannot be alleviated through character displacement, *i.e.* selection is not disruptive (or has a very weak disruptive component). This could happen if, for example, the morphs are competing for a resource that has no alternatives and can only be acquired in a single way (Ex: Competition for sunlight in plants).

## 6.5 Diversity in multi-dimensional trait space

Organisms rarely (if ever) vary along a single independent phenotypic axis. However, adaptive dynamics in multi-dimensional trait space is not very easy to deal with due to mathematical technicalities that arise with identifying necessary and sufficient criteria for evolutionary singularities to be branching points (Doebeli, 2011; Leimar, 2009). As such, adaptive diversification in multi-dimensional trait spaces is usually studied through PDE models or individual-based models (Ispolatov et al., 2016). The motivation for these studies arises from the classic notions of competitive exclusion and limiting similarity, with position in trait space being thought of as analogous to the (Hutchinsonian) ecological niche.

Broadly, the results of these studies align with intuitive expectations: As the dimensionality of the trait space increases, expected diversity increases. For one-dimensional adaptive diversification, strong frequency-dependence is needed to ensure coexistence of multiple morphs/branches at equilibrium. Using the continuous logistic equation (6.5) with Gaussian

forms of  $K(x)$  and  $\alpha(x, y)$ , Doebeli and Ispolatov have shown that as the dimensionality of the trait space increases, progressively weaker frequency dependence is sufficient to maintain multiple morphs at equilibrium (Doebeli and Ispolatov, 2010). Furthermore, they have also shown that for a fixed level of frequency-dependence, the probability of having multiple morphs at equilibrium increases as the dimensionality of the trait space increases. Other authors have since arrived at broadly similar conclusions in more general settings (Débarre et al., 2014; Svardal et al., 2014). While these studies say that coexistence is *easier* in higher dimensions, they do not comment on how many morphs are expected to coexist at equilibrium. This has only been addressed by a recent study (Doebeli and Ispolatov, 2017) which used adaptive dynamics for a continuous analogue of the Lotka-Volterra competition equations to attack the question. They find that all else being equal, the number of coexisting morphs should increase exponentially with the dimensionality of the trait space that the organisms compete in. These authors also repeat this analysis using PDE models (Equation (6.5) in particular) and computational individual-based simulations in place of adaptive dynamics and claim to find the same broad results.



# References

- Barton, N. H. and Etheridge, A. (2019). “Mathematical Models in Population Genetics”. en.  
In: *Handbook of Statistical Genomics*.  
Section: 4 \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781119487845.ch4>.  
John Wiley & Sons, Ltd, pp. 115–20. ISBN: 978-1-119-48784-5. DOI: [10.1002/9781119487845.ch4](https://doi.org/10.1002/9781119487845.ch4).
- Broom, M. and Rychtář, J. (2022). *Game-Theoretical Models in Biology*. 2nd ed.  
New York: Chapman and Hall/CRC. ISBN: 978-1-00-302468-2. DOI: [10.1201/9781003024682](https://doi.org/10.1201/9781003024682).
- Crow, J. F. and Kimura, M. (1970). *An introduction to population genetics theory*. eng.  
Harper international editions. OCLC: 91086. New York: Harper & Row.
- Czuppon, P. and Traulsen, A. (2021).  
“Understanding evolutionary and ecological dynamics using a continuum limit”. en. In: *Ecology and Evolution* 11.11. \_eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/ece3.7205>, pp. 5857–5873.  
ISSN: 2045-7758. DOI: [10.1002/ece3.7205](https://doi.org/10.1002/ece3.7205).
- Débarre, F., Nuismer, S. L., and Doebeli, M. (2014). “Multidimensional (Co)Evolutionary Stability.”  
In: *The American Naturalist* 184.2. Publisher: The University of Chicago Press, pp. 158–171.  
ISSN: 0003-0147. DOI: [10.1086/677137](https://doi.org/10.1086/677137).
- Dieckmann, U. and Doebeli, M. (1999). “On the origin of species by sympatric speciation”. en.  
In: *Nature* 400.6742. Number: 6742 Publisher: Nature Publishing Group, pp. 354–357. ISSN: 1476-4687.  
DOI: [10.1038/22521](https://doi.org/10.1038/22521).
- Doebeli, M. (2011). *Adaptive diversification*. English. OCLC: 687685579.  
Princeton, N.J.: Princeton University Press. ISBN: 978-0-691-12894-8.
- Doebeli, M. and Dieckmann, U. (2000). “Evolutionary Branching and Sympatric Speciation Caused by Different Types of Ecological Interactions.”  
In: *The American Naturalist* 156.S4. Publisher: The University of Chicago Press, S77–S101.  
ISSN: 0003-0147. DOI: [10.1086/303417](https://doi.org/10.1086/303417).
- Doebeli, M. and Ispolatov, I. (2010). “Complexity and Diversity”.  
In: *Science* 328.5977. Publisher: American Association for the Advancement of Science, pp. 494–497.  
DOI: [10.1126/science.1187468](https://doi.org/10.1126/science.1187468).
- (2017). “Diversity and Coevolutionary Dynamics in High-Dimensional Phenotype Spaces”.  
In: *The American Naturalist* 189.2. Publisher: The University of Chicago Press, pp. 105–120.  
ISSN: 0003-0147. DOI: [10.1086/689891](https://doi.org/10.1086/689891).
- Elmhirst, T., Stewart, I., and Doebeli, M. (2008). “Pod systems: an equivariant ordinary differential equation approach to dynamical systems on a spatial domain”. en.  
In: *Nonlinearity* 21.7. Publisher: IOP Publishing, pp. 1507–1531. ISSN: 0951-7715.  
DOI: [10.1088/0951-7715/21/7/008](https://doi.org/10.1088/0951-7715/21/7/008).
- Ethier, S. N. and Kurtz, T. G. (1986). *Markov processes: characterization and convergence*. eng.  
Wiley series in probability and mathematical statistics. OCLC: 12134755. New York: Wiley.  
ISBN: 978-0-471-08186-9.
- Ewens, W. J. (2004). *Mathematical Population Genetics 1: Theoretical Introduction*. en.  
Springer Science & Business Media. ISBN: 978-0-387-20191-7.
- Gardiner, C. W. (2009). *Stochastic methods: a handbook for the natural and social sciences*. English.  
OCLC: 277069162. Berlin: Springer. ISBN: 978-3-540-70712-7.

- Geritz, S., Kisdi, E., Meszéna, G., and Metz, J. (1998). “Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree”. en. In: *Evolutionary Ecology* 12.1, pp. 35–57. ISSN: 1573-8477. DOI: [10.1023/A:1006554906681](https://doi.org/10.1023/A:1006554906681).
- Hastings, A. (2004). “Transients: the key to long-term ecological understanding?” en. In: *Trends in Ecology & Evolution* 19.1, pp. 39–45. ISSN: 0169-5347. DOI: [10.1016/j.tree.2003.09.007](https://doi.org/10.1016/j.tree.2003.09.007).
- Ispolatov, I., Madhok, V., and Doebeli, M. (2016). “Individual-based models for adaptive diversification in high-dimensional phenotype spaces”. en. In: *Journal of Theoretical Biology* 390, pp. 97–105. ISSN: 0022-5193. DOI: [10.1016/j.jtbi.2015.10.009](https://doi.org/10.1016/j.jtbi.2015.10.009).
- Leimar, O. (2009). “Multidimensional convergence stability”. English. In: *Evolutionary Ecology Research* 11.2. Publisher: Evolutionary Ecology, Ltd., pp. 191–208. ISSN: 1522-0613.
- Metcalf, C. J. E. and Pavard, S. (2007). “All paths to fitness lead through demography”. English. In: *Trends in Ecology & Evolution* 22.11. Publisher: Elsevier, pp. 563–564. ISSN: 0169-5347. DOI: [10.1016/j.tree.2007.07.003](https://doi.org/10.1016/j.tree.2007.07.003).
- Nåsell, I. (2001). “Extinction and Quasi-stationarity in the Verhulst Logistic Model”. en. In: *Journal of Theoretical Biology* 211.1, pp. 11–27. ISSN: 0022-5193. DOI: [10.1006/jtbi.2001.2328](https://doi.org/10.1006/jtbi.2001.2328).
- Pearl, R. and Slobodkin, L. (1976). “The Growth of Populations”. In: *The Quarterly Review of Biology* 51. Publisher: University of Chicago Press, pp. 6–24. ISSN: 0033-5770.
- Svardal, H., Rueffler, C., and Doebeli, M. (2014). “Organismal Complexity and the Potential for Evolutionary Diversification”. en. In: *Evolution* 68.11. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/evo.12492>, pp. 3248–3259. ISSN: 1558-5646. DOI: [10.1111/evo.12492](https://doi.org/10.1111/evo.12492).