A Diffusion Approximation of a Three Species Fitness-Dependent Population Model

Liam Peuckert

A Thesis
in
The Department
of
Mathematics and Statistics

Presented in Partial Fulfillment of the Requirements
for the Degree of Master of Science (Mathematics) at
Concordia University
Montreal, Quebec, Canada

November 22, 2015

CONCORDIA UNIVERSITY
SCHOOL OF GRADUATE STUDIES

This is to certify that the thesis prepared
By: **Liam Peuckert**

Entitled: **A Diffusion Approximation of a Three Species Fitness-Dependent Population Model**

and submitted in partial fulfillment of the requirements for the degree of

Master of Science (Mathematics)

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

Approved by _____ Chair
Approved by Dr. X. Zhou

Approved by _____ Examiner
Approved by Dr. W. Sun

Approved by _____ Examiner
Approved by Dr. X. Zhou

Approved by_____ Thesis Supervisor
Approved by Dr. L. Popovic

Approved by _____
Approved by Dr. J. Garrido, Graduate Program Director

October 19th, 2015 _____

Dr. A. Roy, Dean
Faculty of Arts and Science

ABSTRACT

A Diffusion Approximation of a Three Species Fitness-Dependent Population Model

Liam Peuckert

We introduce a continuous time Markov chain to model the ecological competition in the population of 3 species with fitness governed by a modified Moran process based on environmental resources in a limited niche of constant total population N. We run simulations to observe population behavior under different N and initial conditions. We then propose a model approximation which for large N converges to an ODE over most of the population space, with the population following a deterministic trajectory until it reaches an asymptotically stable line. We then prove that the approximation converges to a one dimensional diffusion forced onto the stable line until the first extinction occurs. We use the drift and diffusion coefficients of the diffusion to calculate the expected probability of first extinction for specific species, as well as the expected time until first extinction. Finally, we compare these with data obtained via simulations to show that the approximation is a good fit.

Acknowledgements

I would like to first express my gratitude to my supervisor, Dr Lea Popovic, for her boundless patience, guidance and all the help she has provided during my masters studies, from proposing the model for this paper to suggestions about direction and solutions to overcome roadblocks.

I would also like to thank the staff of the Department of Mathematics and Statistics for all their help.

To my mother, father and sister, I would like to thank them for all the support I have received.

To my friends, thanks for your help and company.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

Deterministic Approach

Many different approaches to modeling populations of interacting species or groups within a species have been developped over time. These models borrow from different branches of mathematics, although many of these approaches end up overlapping in some way. One of the first was the Lotka-Volterra predator-prey model. It was first proposed by Lotka in 1925 by extending one of his previous models applied to chemical reactions [12]. Volterra, who was analysing fluctuations in the fish populations in the Adriatic sea, independently constructed the model in 1926, working on the assumption that fish and sharks were in a predator-prey relationship [19].

Let $x(t)$ be the number of individuals of the prey species, and $y(t)$ the number of predator. Then according to the Lotka-Volterra model, the change in populations over time follows these two equations:

$$\begin{aligned} \frac{dx}{dt} &= x(\lambda - by) \\ \frac{dy}{dt} &= y(-\mu + cx) \end{aligned} \tag{1.1}$$

In the absence of predators, the prey reproduce at a constant per-capita rate λ. Therefore the prey never suffer from shortages of any resources, such as food. Predators, on the other

hand, die at a constant per-capita rate μ in the absence of the prey species, meaning that they are completely dependent on the prey species as a food source. When both species are present, the predators eat the prey, which adds to the growth rate of the predator and reduces that of the prey [2].

The Lotka-Volterra model has two stable points (points where the population does not change over time). One is when $x = 0$ and $y = 0$; in other words when both species are extinct, their populations remain at zero since there is no way for them to recover. The other exists at $x = \mu/c$ and $y = \lambda/b$, so both populations are at perfect equilibrium at this point. At any other population state, the population will exhibit periodic behavior around the non-zero stable point, meaning the populations will over time follow the same trajectory endlessly. Plotting the populations on the x-y plane, the trajectory will form a closed loop around the non-zero stable point.

The Lotka-Volterra predator-prey model makes many assumptions in addition to ones already mentioned, such as having both species' growth rates proportional to their populations and ignoring any environmental or genetic changes. Numerous other models of differential equations have been proposed to better fit different assumptions, to accommodate different numbers and types of species interactions, or to better mimic the population dynamics observed in certain species. For example, if a pair of predator-prey species do not display periodic behavior, such as population trajectories heading towards certain stable points or towards extinction, then the Lotka-Volterra model would be a bad fit.

However, deterministic differential equations by their nature cannot take into account noise or randomness at all, and running a model with a given set of parameters and initial values will always produce the same results. So for some purposes, such as examining probabilities of extinction, stochastic models are more appropriate.

Stochastic Approach

For stochastic population models, both discrete and continuous time Markov chains can be used. One commonly used model of Markov chains are birth-death processes, where

the only transitions are single births (population increases by 1), single deaths (population decreases by 1) or the population remains the same [1]. So in the continuous case of a general, single-type birth and death process, the infinitesimal transition probabilities satisfy

$$\begin{aligned} p_{i+j,i}(\Delta t) &= P(\Delta X(t) = j \mid X(t) = i) \\ &= \begin{cases} \lambda_i \Delta t + o(t), & j = 1 \\ \mu_i \Delta t + o(t), & j = -1 \\ 1 - (\lambda_i + \mu_i)\Delta t + o(t), & j = 0 \\ o(t), & j \neq -1, 0, 1 \end{cases} \end{aligned} \tag{1.2}$$

for Δt sufficiently small. Note that λ_i and μ_i are usually described as functions of the population size $X(t)$.

We can also have multi-type birth and death process, where $X(t)$ is now a vector of population sizes. Then the number of different jumps possible from any point in the population can be much greater, since there are many possible combinations of births and deaths.

Let us consider a two species continuous-time Moran process with neutral drift as an example. We will elaborate further on in the introduction about Moran processes; for now we will think of this as a birth-death process where the sum of the population of both species stays constant at N, and with rate 1 a simultaneous birth-death event occurs in which one member of either species is born, and one member of either species dies. Both the newborn and dying individuals are chosen in proportion to their respective populations. Label the two species as $(X(t), Y(t))$. Since $Y(t) = N - X(t)$, the behavior of the two species are completely dependent on each other, and we can track the model by tracking only species

$X(t)$. In this example, the infinitesimal transition probabilities for species $X(t)$ are

$$
\begin{aligned}
p_{0,0}(\Delta t) &= 1 \\
p_{N,N}(\Delta t) &= 1 \\
p_{i+j,i}(\Delta t) &= \begin{cases} \frac{i}{N}\frac{N-i}{N}\Delta t + o(t), & j = 1 \\ \frac{N-i}{N}\frac{i}{N}\Delta t + o(t), & j = -1 \\ 1 - \frac{2i(N-i)}{N^2}\Delta t + o(t), & j = 0 \\ o(t), & j \neq -1, 0, 1 \end{cases}
\end{aligned} \tag{1.3}
$$

for $0 < i < N$ and Δt sufficiently small. The expected change in population caused by the next simultaneous birth-death event given a certain state, $E(\Delta X(t) \mid X(t) = i)$, is equal to zero, as implied by the name 'neutral drift'. The second moment is $E(\Delta X^2(t) \mid X(t) = i) = \frac{2i(N-1)}{N^2}$.

For continuous time Markov chains, differential equations make a comeback as we can express the rate of change of these transition probabilities with the forward and backward Kolmogorov differential equations. Probabilistic models can also make use of stochastic differential equations and diffusion processes, which are Markov processes with continuous sample paths and additional properties regarding their infinitesimal means and variances which we will elaborate upon later. Ito SDEs are often chosen for interacting populations, and can be derived from discrete Markov chains [1].

For large populations, certain birth-death processes can be approximated by diffusion processes. One way to do this is by manipulating the generator of the Markov process; since generators of Markov processes uniquely define them, showing that one generator approximates another is equivalent to showing that the two processes approximate each other in distribution. If L is an operator on real-valued functions f of the state space and $\beta(x, y)$ denotes the jump rate of a process from state x to state y, then the generator of the Markov

process is [16]

$$(Lf)(x) = \sum_y \beta(x,y)(f(y) - f(x)) \tag{1.4}$$

By taking limits of time or population size of these generators, in some cases generators of a diffusion process can be obtained. One-dimensional diffusions have the following infinitesimal generator:

$$(Lf)(x) = \frac{d}{dt} E_x f(X_t) \mid_{t=0} = \frac{1}{2} a(x) \frac{d^2}{dx^2} f(x) + b(x) \frac{d}{dx} f(x) \tag{1.5}$$

where $b(x)$ and $a(x)$ are known as the infinitesimal mean and infinitesimal variance respectively [5]. For one-dimensional diffusions, formulas exist for finding such values as hitting times and probabilities which might be difficult or impossible to otherwise calculate in the original birth-death process. Some diffusions are degenerate, in which case their infinitesimal variance goes to zero and it behaves like an ordinary differential equation.

Returning to our two species continuous-time Moran process with neutral drift example, we can approximate it with a diffusion after it has been rescaled. The generator the rescaled process $N^{-1}X(Nt)$ is

$$(Lf)(x) = N\left(\frac{i(N-i)}{N^2}(f(i+\frac{1}{N}) - f(i)) + \frac{i(N-i)}{N^2}(f(i-\frac{1}{N}) - f(i))\right) \tag{1.6}$$

By using a Taylor expansion and letting the population go to infinity, we get the following approximation of the generator

$$(\tilde{L}f)(x) = \frac{1}{2N} \frac{2i(N-1)}{N^2} \frac{d^2}{dx^2} f(x) + 0 \frac{d}{dx} f(x) \tag{1.7}$$

in which case we see that the infinitesimal mean and variance are equal to $E(\Delta X(t) \mid X(t) = i)$ and $\frac{1}{N} E(\Delta X^2(t) \mid X(t) = i)$ respectively.

Game Theory Approach

Another approach to interacting populations is to apply game theory to evolution. In

game theory we have players playing a "game" in which they each choose a strategy, and then receive a payoff based off of the interactions of the various strategies chosen. When applied to biology and evolution, each individual plant, animal or bacteria is a player, and the "strategy" they choose is based off of their species or phenotype. The "game" can be a competition for resources, or a mating strategy, or a reaction to a change in the environment, for example. Finally the payoff is a surrogate for fitness [3]. The fit reproduce, and so phenotypes spread or wither within a population. In the game theory approach the payoffs, as well as the evolutionary game mechanisms for selecting births, deaths, invasions and mutation can be either stochastic or deterministic. For example, a species' mating strategy could involve a large game of rock-paper-scissors, with those choosing 'rock' having the best chance of reproducing if most rivals play 'scissors', and so on.

	rock	paper	scissors
rock	0.5	0	1
paper	1	0.5	0
scissors	0	1	0.5

Table 1.1: Example of a fitness payoff matrix for a pairwise comparison rock-paper-scissors model. Two individuals are chosen at random from the population, their fitnesses are compared and one will be replaced by the child of the other. The numbers represent the fitness of the row strategy when paired against the column strategy.

In such a model, depending on the payoff matrix and the evolutionary game dynamics, it is possible for all three strategies to coexist in equilibrium, in which case the disappearance of a phenotype would be highly unlikely. Alternately, a different set of parameters could lead to highly unstable dynamics featuring regular cycles of invasion and extinction. This might seem like a purely theoretical model, yet such a non-transitive dynamic has been found in the mating strategies of the side-blotched lizard [15] [4].

One example of how selection dynamics would work in an evolutionary game with finite population N is the Moran process, such as the Moran process with neutral drift example introduced previously. It stems from a classical model of population genetics and has been

recently applied to game theory. Individuals are replaced at a rate $\frac{1}{N}$. That is, an individual x lives for an exponentially distributed amount of time with mean N, and the population as a whole experiences one "replacement" after another at a rate of 1. So lifespans and deaths are random and independent of fitness.

To replace individual x, we choose an individual proportional to fitness from the population (including x itself) to be the parent of the new individual [5]. Since this is an evolutionary game, fitness in this case is based on the payoff of a "game" involving the different populations. If fitness is completely random, birth chances are equal for all individuals and proportional to population for all types and we obtain neutral drift, as in our example.

Species	Individual fitness	Birth probability
X	1	$X(t)/N$
Y	1	$Y(t)/N$

Table 1.2: Fitness payoff and birth probability in a two species $(X(t), Y(t))$ Moran model with neutral drift and total population size N. We see that the fitness of each individual is constant in all cases, and so the probability of a species reproducing during the next birth-death event is proportional to its population.

Let us apply the rock-paper-scissors dynamics to a Moran model as another brief example. In a population dominated by paper individuals, each individal has an equal chance of being replaced regardless of type, but the new individual will probably be rock. Note that the total population remains constant at N at all times. The Moran process is a birth-death process [18].

Model Dimensions

Both the Lotka-Volterra predator-prey model and the rock-paper-scissors Moran process are 2-dimensional (in the Moran process case, since the sum of all populations remains fixed, the population of the third species is completely dependent on the other two). The Lotka-Volterra model features periodic orbits around a fixed point, a behavior which is not possible in a 1-dimensional ordinary differential equations system. Systems with more dimensions allow for more complex behavior when looking at population drift. Just as rock-paper-

scissors offers the possibility of a attracting or repulsing fixed point, higher dimensions offer the possiblitiy of multi-dimensional manifolds of fixed points exerting attraction or repulsion in the space around them, or of being the centers of convoluted periodic orbits.

From a real-life perspective, additional dimensions permits the inclusions of more species or additional groups with different characteristics, phenotypes or strategies. Complex ecosystems can feature multiple species and long food chain. Starting off with a higher-dimensional model allows us to capture this complexity. However, it is also useful to approximate a higher-dimension model by a lower-dimension one. For example, if a 2-dimensional model forces the population state onto a curve, it can be possible to approximate it by a one-dimensional model. This can simplify calculations greatly while providing insight into the model behavior, both of which occur when we approximated this paper's model with a diffusion of a lower dimension.

This Paper's Model

In an attempt to find a good approximation to a model of three interacting groups, this thesis touches upon most of the processes and approaches mentioned so far. The intial model introduced in Chapter 3 is a continuous-time Moran process. While the focus of the fitness of most Moran processes in population biology is on the interaction of different groups, such as in the rock-paper-scissors example, in this case the fitness is based on an environment which fluctuates between two distinct states. Two of the groups are specialists which have high fitness in one environment and zero fitness in the other, while a third group is a generalist that maintains a constant, lower fitness regardless of the environmental conditions. By adding a simple environmental component to a Moran process, this model has the potential to be a stepping stone towards models seeking to combine the interactive fitness of the Moran process with outside factors that influence survival.

To make the model feel more concrete, the environmental states were termed plant season and meat season, and the three groups labeled as a carnivore, herbivore and omnivore species. However, these terms were arbitrarily chosen to make the model intuitive, and can

be substituted with whichever environments and groups the reader prefers.

Species	Individual fitness		Birth Probability	
	Meat season	Plant season	Meat season	Plant season
Carnivore $C(t)$	1	0	$2C(t)/(2C(t)+M(t))$	0
Herbivore $H(t)$	0	1	0	$2H(t)/(2H(t)+M(t))$
Omnivore $M(t)$	0.5	0.5	$M(t)/(2C(t)+M(t))$	$M(t)/(2H(t)+M(t))$

Table 1.3: Fitness payoff and birth probability for this paper's Moran model. Fitness of each species (Carnivore, Herbivore and Omnivore, represented by $(C(t), H(t), M(t))$) is season-dependent (seasons being meat season and plant season). Total population remains constant at $C(t) + H(t) + M(t) = N$.

A discrete-time birth death process is used as an analogue of the model for computer simulations in Chapter 4 to gain a better understanding of the population dynamics. In Chapter 5 we use an approximation on the generator to find that for large populations the model can be approximated by a two-dimensional diffusion which is degenerate over most of the population space. However, it is not degenerate in the case when both species of specialists have equal populations, and the simulations from the previous Chapter spent most of their time in or close to this state. The degenerate diffusion forces the process along its ODE curves and onto this line in the population space. In Chapter 6, we use Ito's formula to show that the two-dimensional mostly degenerate diffusion converges to a one-dimensional diffusion process along this line. This diffusion approximation is shown in Chapter 7 to be a good fit for the simulation results.

This paper also serves as a a useful demonstration for other higher-dimensional diffusion models which are degenerate, forced onto lines and could also be shown to converge to one-dimensional diffusion processes. Such degenerate diffusions might arise from model which, like ours, feature symmetrical equilibriums among some groups (in our case, the two specialists).

Chapter 2

Background

Here are some definitions that will be used throughout the thesis.

2.1 Markov Chains

Most of the definitions in this chapter are based off of ones given in the book *Probability and Random Processes* by Geoffrey Grimmett and David Stirzaker [8].

A Markov process is a special type of stochastic process which is memoryless: only the current state of the process influences what the next state will be.

Let us first examine discrete-time processes. Let $\{X_0, X_1, ...\}$ be a sequence of random variables which take values in *State Space* S, which is a countable set. Each X_n is a discrete random variable that takes one of N possible values, where $N = |S|$.

Definition 2.1.1. The process X is a **Markov chain** if it satisfies the **Markov condition**:

$$P(X_n = s \mid X_0 = x_0, X_1 = x_1, ..., X_{n-1} = x_{n-1}) = P(X_n = s \mid X_{n-1} = x_{n-1}) \tag{2.1}$$

for all $n \geq 1$ and all $s, x_1, ..., x_{n-1} \in S$.

Definition 2.1.2. The Markov chain X is **Homogeneous** if

$$P(X_{n+1} = j \mid X_n = i) = P(X_1 = j \mid X_0 = i) \tag{2.2}$$

for all n, i, j.

In this thesis all Markov processes, both discrete and continuous, are homogeneous.

In the case of homogeneous Markov chains, we can denote the **transition probabilities** as

$$p_{ij} = P(X_{n+1} = j \mid X_n = i) \tag{2.3}$$

For a Markov process, an absorbing state is a state that, once entered, can never be left.

Definition 2.1.3. State i is an **absorbing state** if and only if

$$p_{ii} = 1, p_{ij} = 0 \ \ \forall i \neq j \tag{2.4}$$

Markov chains can also be continuous-time. Let $X = \{X(t) : t \geq 0\}$ be a family of random variables taking values in a countable state space S and indexed by the half-line $[0, \infty)$.

Definition 2.1.4. The process X is a **continuous-time Markov chain** if it satisfies the **Markov property**:

$$P(X(t_n) = j \mid X(t_1) = i_1, ..., X(t_{n-1}) = i_{n-1}) = P(X(t_n) = j \mid X(t_{n-1}) = i_{n-1}) \tag{2.5}$$

for all $j, i_1, ..., i_{n-1} \in S$ and any sequence $t_1 < t_2 < ... < t_n$ of times.

Theorem 2.1.1. *Let T_n be the time of the nth change in value of the continuous-time Markov chain X. Then the holding time $T_{n+1} - T_n$ has the exponential distribution.*

If we concentrate on the changes of state of a continuous-time Markov chain $X(t_n)$ at the times of jumps, we can approach it as a discrete-time Markov chain.

Theorem 2.1.2. *Let T_n be the time of the nth change in value of the continuous-time Markov chain X. The values $Z_n = X(T_n+)$ of X immediately after its jumps constitute a discrete-time Markov chain Z with transition probabilities proportional to the various jump rates.*

Definition 2.1.5. If L is an operator on real-valued functions f of the state space and $\beta(x, y)$ denotes the jump rate of a Markov chain from state x to state y, then the generator of the Markov chain is [16]

$$(Lf)(x) = \sum_y \beta(x, y)(f(y) - f(x)) \tag{2.6}$$

2.2 Diffusions

Most of this section is based off of books by Øksendal [13] as well as Revuz and Yor [14].

Definition 2.2.1. A diffusion is a stochastic process $X_t(\omega) = X(t, \omega) : [0, \infty) \times \Omega \to \mathbb{R}^n$ satisfying a stochastic differential equation of the form

$$dX_t = b(X_t)dt + \sigma(X_t)dB_t, \;\; t \geq s; \; X_s = x \tag{2.7}$$

where B_t is a m-dimensional Brownian motion and $b : \mathbb{R}^n \to \mathbb{R}^n, \sigma : \mathbb{R}^n \to \mathbb{R}^{n \times m}$ are Lipschitz continuous.

Definition 2.2.2. Let $\{X_t\}$ be a diffusion in $\mathbb{R}^n$. The (infinitesimal) generator L of X_t is defined by

$$Lf(x) = \lim_{t \to 0} \frac{E_x[f(X_t)] - f(x)}{t}; \;\; x \in \mathbb{R}^n \tag{2.8}$$

Theorem 2.2.1. *Let X_t be the diffusion* $\quad dX_t = b(X_t)dt + \sigma(X_t)dB_t$.

If $f \in C_0^2(\mathbb{R}^n)$ then the diffusion X_t has generator

$$Lf(x) = \frac{1}{2} \sum_{i,j}^{n} a_{i,j}(x) \frac{\partial^2}{\partial x_i \partial x_j} f + \sum_{i}^{n} b_i(x) \frac{\partial}{\partial x_i} f \tag{2.9}$$

where $a = \sigma \times \sigma^T$.

Theorem 2.2.2. *Assume* b *and* σ *are bounded and Lipschitz continuous, and that* a *is uniformly positive definite. Let* $y(t)$ *be a Markov process with values in* $\mathbb{R}^n$ *and with generator* L *defined previously. Then first-order generators define deterministic processes. [16]*

Theorem 2.2.3. *If* X *is a diffusion with covariance* a *and drift* b*, then there exists a predictable process* σ *such that* $a = \sigma \times \sigma^T$ *and a Brownian motion* B *on an enlargement of the probability space such that [14]*

$$X_t = X_0 + \int_0^t \sigma(s) dB_s + \int_0^t b(s) ds \tag{2.10}$$

Theorem 2.2.4. ***Ito's Formula****: Let* $F \in C^2(\mathbb{R}^d, \mathbb{R})$ *and* $X = (X^1, ..., X^d)$ *be a continuous vector semimartingale; then* $F(X)$ *is a continous semimartingale and*

$$F(X_t) = F(X_0) + \sum_{i=1}^{d} \int_0^t \partial_i F(X_s) dX_s^i + \frac{1}{2} \sum_{1 \le i,j \le d} \int_0^t \partial_i \partial_j F(X_s) d\langle X^i, X^j \rangle_s \tag{2.11}$$

Theorem 2.2.5. *Theorem: Maximal inequality (Brownian Motion). If* $\int_0^t \mathbb{E}[Y_s^2] \mathrm{d}s < \infty$*, then*

$$P\{\max_{0 \le s \le t} | \int_0^t Y_s \mathrm{d}B_s | \ge a\} \le \frac{\mathbb{E}[(\int_0^t Y_s \mathrm{d}B_s)^2]}{a^2} = \frac{\int_0^t \mathbb{E}[Y_s^2] \mathrm{d}s}{a^2} \tag{2.12}$$

2.3 ODEs and Dynamical systems

The following definitions are based on those from the book *Ordinary Differential Equations and Dynamical Systems* by Gerard Teschl [17]. Definitions were modified to accomodate manifolds.

Let $y = (d(t), m(t))$ represent values at time t of the ODE system. Denote the derivative

of this function with respect to time by

$$g(y) = \dot{y}(t) \tag{2.13}$$

Definition 2.3.1. A point y^* with $g(y^*) = 0$ is then a **fixed point**.

Definition 2.3.2. A **Lyapunov function** V of a manifold W of fixed points is a continuous function over an open neighborhood U of W which is zero for all fixed points $y^* \in W$, positive for $y \neq y^*$ and satisfies

$$V(y(t_0)) \geq V(y(t_1)),\ t_0 < t_1,\ \text{for } y(t_j) \in U \backslash W \tag{2.14}$$

Definition 2.3.3. A manifold W of fixed points is called **stable** if for any given neighborhood $U \supseteq W$ there exists another neighborhood S, $W \subseteq U \subseteq S$ such that any solution starting in S remains in U for all $t \geq 0$.

Definition 2.3.4. Similarly, a manifold W of fixed points of $y(t)$ is called **asymptotically stable** if it is stable and if there is a neighborhood $U \supseteq W$ such that

$$\lim_{t \to \infty} |y(t) - y^*| = 0 \ \forall\ y(t) \in U,\ y^* \in W \tag{2.15}$$

Theorem 2.3.1. *Suppose W is a manifold of fixed points. If there is a Lyapunov function V of manifold W, then W is an asymptotically stable manifold.*

2.4 Matrices

Theorem 2.4.1. *Let M be a 2 by 2 matrix with values*

$$M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

Define δ as the determinant $\delta = ad - bc$, $q = \pm\sqrt{\delta}$, $r = \pm\sqrt{a + d + 2q}$. Then if $r \neq 0$, a square root of matrix M is

$$R = \frac{1}{r}\begin{pmatrix} a+q & b \\ c & d+q \end{pmatrix} \tag{2.16}$$

2.5 Convergence of a Stochastic Differential Equation Forced onto a Manifold

The following is a slightly modified version of theorem 6.3 from Katzenberger's paper detailing the behavior of a stochastic differential equation forced onto a manifold by large drift. [9] It provides a more rigorous approach to the convergence of our 2-dimensional diffusion to a 1-dimensional diffusion. However, since the following theorem is not straightforward, a more accessible proof of convergence is used in this paper.

Let $U \subset \mathbb{R}^d$ be open and $F : U \to \mathbb{R}^d$ be a C^1 vector field. Assume $\Gamma = \{x|F(x) = 0\}$ is a C^0 submanifold of U of dimension m.

For $n \geq 1$, let $(\Omega^n, \mathcal{F}^n, \{\mathcal{F}_t^n\}_{t\geq 0}, P)$ be a flitered probability space, let Z_n be a an $\mathbb{R}^e$-valued cadlag $\{\mathcal{F}_t^n\}$-semimartingale with $Z_n(0) = 0$ and let A_n be a real-valued cadlag $\{\mathcal{F}_t^n\}$-semimartingale satisfying

$$X_n(t) = X_n(0) + \int_0^t \sigma_n(X_n)dZ_n + \int_0^t F(X_n)dA_n \tag{2.17}$$

for all $t \leq \lambda_n(K)$ and all compact $K \subset U$, where

$$\lambda_n(K) = \inf\{t \geq 0|X_n(t-) \notin K or X_n(t) \notin K\} \tag{2.18}$$

Assume that Γ is C^2 and, for every $y \in \Gamma$, the matrix $\partial F(y)$ has $d - m$ negative eigenvalues. Assume the martingale part of Z_n is uniformly integrable, that the process A_n is

asymptotically continuous, Φ is C^2 (or F is LC^2) and $X_n \Rightarrow X(0) \in U_\Gamma$. Let

$$Y_n(t) = X_n(t) - \psi(X_n(0), A_n(t)) + \Phi(X_n(0)) \tag{2.19}$$

and, for compact $K \subset U_\Gamma$, let

$$\mu_n(K) = \inf\{t \geq 0 | Y_n(t-) \notin K or Y_n(t) \notin K\} \tag{2.20}$$

Then, for every compact $K \subset U_\Gamma$, the sequence $\{Y_n^{\mu_n(K)}, Z_n^{\mu_n(K)}, \mu_n(K)\}$ is relatively compact in $D_{\mathbb{R}^d \times \mathbb{R}^e}[0,\infty) \times [0,\infty]$. If (Y, Z, μ) is a limit point of this sequence, then (Y, Z) is a continuous semimartingale, $Y(t) \in \Gamma$ for every t a.s., $\mu \geq \inf\{t \geq 0 | Y(t) \notin K\}$ a.s. and

$$\begin{aligned} Y(t) =& Y(0) + \int_0^{t\wedge\mu} \partial\Phi(Y)\sigma(Y)dZ \\ &+ \frac{1}{2}\sum_{ijkl} \int_0^{t\wedge\mu} \partial_{ij}\Phi(Y)\sigma^{ij}(Y)d[Z^k, Z^l] \end{aligned} \tag{2.21}$$

The following remark from Katzenberger's paper helps to explain the significance of the previous theorem:

"Basically this theorem says that X_n follows the flow of F according to the clock $A_n(t)$ until X_n is close to Γ, then it stays close to Γ and moves according to the SDE given in (previous eqn). Notice that $\psi(X_n(0), A_n(t)) - \Phi(X_n(0))$ is small for t bounded away from 0. The theorem implies that $X_n(t) - \psi(X_n(0), A_n(t))$ is small for t close to 0." [9]

Chapter 3

Model

3.1 Introducing the Model

We define this model as Moran model represented by a continuous time, vector valued Markov chain $X(t) = (C(t), H(t), M(t))$, $t \geq 0$, on the state space $\{0, 1, ..., N\}^3$. Each coordinate of the vector represents the population of a species (carnivores, herbivores and omnvores, respectively), and $C(t) + H(t) + M(t) = N$. In other words, the total population size of all three species is constant at N.

The population undergoes a simultaneous birth and death at a constant rate β. Since this rate is independent of time t, our Markov chain is time homogeneous, otherwise known as stationary. Thus if we let T_i denote the amount of time between the ith birth-death and the next one, T_i's are exponentially distributed with mean $\frac{1}{\beta}$ and are mutually independent. For simplicity, we will set $\beta = 1$.

We can break down this jump rate based on which species give birth or die. Let $\mu_K(X)$ be the death rate of species K at population state $X(t)$. Similarly, let $\lambda_K(X)$ be the birth rate of species K at population state $X(t)$. Note that

$$\begin{aligned} \mu_C(X) + \mu_H(X) + \mu_M(X) &= \beta = 1 \\ \lambda_C(X) + \lambda_H(X) + \lambda_M(X) &= \beta = 1 \end{aligned} \tag{3.1}$$

since the sum of the different death rates at a given population state $X(t)$ must equal the total jump rate, and the same applies to the sum of the birth rates.

Since this is a Moran model, deaths are completely random. During a death event, each individual is equally likely to die regardless of species, thus the death rate of the whole species is proportionate to the number of individuals alive. To simplify notation, we will denote $C(t)$,$H(t)$ and $M(t)$ by C,H and M respectively.

For $C \neq N, H \neq N$, $M \neq N$,

$$\begin{aligned} \mu_C(X) &= \frac{C}{N} \\ \mu_H(X) &= \frac{H}{N} \\ \mu_M(X) &= \frac{M}{N} \end{aligned} \tag{3.2}$$

Birth rates in a Moran model, however, are fitness dependent: random, but proportional to fitness.

For $C \neq N, H \neq N$, $M \neq N$,

$$\begin{aligned} \lambda_C(X) &= q * \frac{2C}{2C+M} \\ \lambda_H(X) &= (1-q) * \frac{2H}{2H+M} \\ \lambda_M(X) &= q * \frac{M}{2C+M} + (1-q) * \frac{M}{2H+M} \end{aligned} \tag{3.3}$$

The reasoning behind these rates is that fitness is determined by the "food season" (meat season and plant season) and the diet of the species. The type of season is randomly determined at the time of the birth-death; it is a meat season with probability q and a plant season with probability $1-q$. For simplicity, we have set $q = \frac{1}{2}$ for the rest of this paper. In a meat season, herbivores are totally outcompeted and have no chance of birthing, whereas carnivores are twice as likely as omnivores to birth (proportional to population). In a plant season, carnivores can't reproduce and herbivores are twice as likely as omnivores to birth. An exception is made when only one species remains: if herbivores are the sole surviving

species, they cannot be outcompeted and thus birth even in the unfavorable meat season. Thus the total population remains stable no matter how many species coexist. There is no mutation, so extinction is permanent.

Note that although the model has been termed with labels pertaining to food and diet, it could conceivably work with other criteria such as weather conditions, and the 3 "species" could be simple phenotypes within a species, thus mutations between phenotypes could be introduced.

3.2 Additional Notation: 2 Variable Representation and Proportional Population

While a 3 variable notation has been used to define the mode and program the simulation, it is possible to use only two variables to define the model because of the constraint $C(t) + H(t) + M(t) = N$. For mathematical analysis it is preferable to use a 2 variable notation, not least because this permits the representation the population space in 2 dimensions. Therefore most of the rest of this article will use the 2-variable notation.

Denote the difference in population between carnivores and herbivores at time t by $D(t) = C(t) - H(t)$. Instead of $X(t)$, the model can be represented by the 2-dimensional vector

$$Y(t) = (D(t), M(t)) \tag{3.4}$$

It can be useful to use proportions of populations rather than absolute populations. To

this end we will use lower-case letters to distinguish the two.

$$\begin{aligned} x(t) &= X(t)/N = (c(t), h(t), m(t)) \\ c(t) &= C(t)/N \\ h(t) &= H(t)/N \\ m(t) &= M(t)/N \\ d(t) &= c(t) - h(t) = D(t)/N \end{aligned} \tag{3.5}$$

3.3 Infinitesimal Transition Probabilities

This section deals with non-absorbing states. For absorbing states, see the next section.

For the three variable notation, denote the infinitesimal transition probabilities by

$$p_{X+j,X}(\Delta t) = P(\Delta X(t) = j \mid X(t) = X) \tag{3.6}$$

For $C \neq N, H \neq N$, $M \neq N$ (i.e., for non-absorbing states)

$$p_{X+j,X}(\Delta t) = \begin{cases} \lambda_C(X)\mu_M(X)\Delta t + o(t), & j = (1, 0, -1) \\ \lambda_C(X)\mu_H(X)\Delta t + o(t), & j = (1, -1, 0) \\ \lambda_H(X)\mu_M(X)\Delta t + o(t), & j = (0, 1, -1) \\ \lambda_H(X)\mu_C(X)\Delta t + o(t), & j = (-1, 1, 0) \\ \lambda_M(X)\mu_H(X)\Delta t + o(t), & j = (0, -1, 1) \\ \lambda_M(X)\mu_C(X)\Delta t + o(t), & j = (-1, 0, 1) \\ 1 - (1 - \lambda_C(X)\mu_C(X) - \lambda_H(X)\mu_H(X) - \lambda_M(X)\mu_M(X))\Delta t + o(t), & j = (0, 0, 0) \\ o(t) & otherwise \end{cases} \tag{3.7}$$

From these we can calculate the infinitesimal transition probabilities in the 2-variable notation.

For $D \neq \pm N, M \neq N$ (i.e., for non-absorbing states)

$$p_{Y+j,Y}(\Delta t) = \begin{cases} \frac{(N+D-M)(N-D-M)}{4N(N+D)}\Delta t + o(t), & j = (2,0) \\ \frac{(N+D-M)(N-D-M)}{4N(N-D)}\Delta t + o(t), & j = (-2,0) \\ \frac{M(N-D-M)}{2(N+D)(N-D)}\Delta t + o(t), & j = (1,1) \\ \frac{M(N+D-M)}{2(N+D)(N-D)}\Delta t + o(t), & j = (-1,1) \\ \frac{M(N+D-M)}{2N(N+D)}\Delta t + o(t), & j = (1,-1) \\ \frac{M(N-D-M)}{2N(N-D)}\Delta t + o(t), & j = (-1,-1) \\ 1 - (1 - \frac{(N+D-M)^2+2M^2}{4N(N+D)} - \frac{(N-D-M)^2+2M^2}{4N(N-D)})\Delta t + o(t), & j = (0,0) \\ o(t) & otherwise \end{cases} \tag{3.8}$$

The previous transition probabilities show that $Y(t)$, like $X(t)$, is still a Markov chain, since for filtration $\mathcal{F}_t$ and for every $A \in S, \quad P(Y(t) \in A \mid \mathcal{F}_s) = P(Y(t) \in A \mid Y(s))$

3.4 Absorbing States

Recall that for a Markov Process, an absorbing state is a state that, once entered, can never be left. In this model, there are 3 absorbing states: $\{X(t) = (N,0,0), (0,N,0)$ or $(0,0,N)\}$, corresponding to the extinction of 2 species. In 2-variable notation, these states are represented as $\{Y(t) = (-N,0), (N,0)$ or $(0,N)\}$. Not only are these expected due to extictions, they are also designed into the model with their own transition probabilities.

For $C = N, H = N,\ M = N$

$$p_{X+j,X}(\Delta t) = \begin{cases} 1, & j = (0,0,0) \\ 0 & otherwise \end{cases} \tag{3.9}$$

For $D(t) = \pm N$ and $M(t) = N$, and thus for $d(t) = \pm 1$ and $m = 1$, all moments and variances are 0 since there is no longer any change in population.

3.5 Moments of simultaneous birth-death changes in population

Here we examine the moments of the change in absolute population caused by the next simultaneous birth-death event given a certain state. Recalling theorem 2.1.2, we can treat the nth change of value as a discrete-time Markov chain with transition probabilities proportional to the jump rates. Since in our case the sum of our jump rates (if we include as a jump the event that a birth-death event occurs in which the same species type is born and dies, leaving the populations unchanged) sum to $\beta = 1$, the transition probabilities of the discrete Markov chains are equal to the jump rates and so we can use the jump rates as probabilities to calculate the moments of the simultaneous birth-death changes in population.

Note that to represent the moments as functions of proportional populations, we factored out the total population N from every variable. For all of the following moments, the N's in the numerator and denominator cancel out, so the moments are not affected by the size of the total population, merely the proportions of each species.

For $D(t) \neq \pm N, M(t) \neq N$ (i.e., for non-absorbing states)

$$E(\Delta D(t) \mid Y(t)) = \frac{-D(N^2 - NM - D^2)}{N(N+D)(N-D)} = \frac{-d(1-m-d^2)}{(1+d)(1-d)} \tag{3.10}$$

Note that $E(\Delta D(t) \mid Y(t))$=0 if and only if $D(t) = 0$. The other factor, $(N^2 - NM - D^2)$,

equals zero when $M(t) = N$, which implies $D(t) = 0$, or when $D(t) = \pm N$, which are two absorbing states which are defined as having moments equal to zero.

$$E(\Delta M(t) \mid Y(t)) = \frac{D^2 M}{N(N+D)(N-D)} = \frac{d^2 m}{(1+d)(1-d)} \tag{3.11}$$

Note that $E(\Delta M(t) \mid Y(t))$=0 if and only if $D(t) = 0$ or $M(t) = 0$. Also, since N is positive, M is non-negative and $|D(t)| \leq N$, $E(\Delta M(t)) \geq 0$ over the whole population space.

$$E(\Delta D^2(t) \mid Y(t)) = \frac{2N^3 - 2N^2 M - 2ND^2 - MD^2}{N(N+D)(N-D)} = \frac{2 - 2m - 2d^2 - md^2}{(1+d)(1-d)} \tag{3.12}$$

$$E(\Delta M^2(t) \mid Y(t)) = \frac{M(2N^2 - 2NM - D^2)}{N(N+D)(N-D)} = \frac{m(2 - 2m - d^2)}{(1+d)(1-d)} \tag{3.13}$$

$$E[\Delta D(t) * \Delta M(t) \mid Y(t)] = \frac{-DM(N+M)}{N(N+D)(N-D)} = \frac{-dm(1+m)}{(1+d)(1-d)} \tag{3.14}$$

$$Cov(\Delta D(t), \Delta M(t) \mid Y(t)) = \frac{-MD(N^4 + N^3 M - 2N^2 D^2 + D^4)}{N^2(N+D)^2(N-D)^2} = \frac{-dm(d^4 - 2d^2 + m + 1)}{(1+d)^2(1-d)^2} \tag{3.15}$$

Similarly, the variances of $D(t)$ and $M(t)$ are not functions of N. Since they are long and awkward, they have not been included.

3.6 Moments of simultaneous birth-death changes in proportional population

Here we examine the moments of the change in proportional population caused by the next simultaneous birth-death event given a certain state, similar to the previous section.

Moments of proportional variables are equal to the moments of their corresponding population variables divided by the appropriate factor of N. This makes sense intuitively: one step changes in absolute population are not a function of N, since the increment is always of 1 individual. One step changes in proportion, however, depend on N: the larger the total population, the smaller the increment in proportional population when a single individual is born or dies.

For $d(t) \neq \pm 1, m \neq 1$ (i.e., for non-absorbing states)

$$E(\Delta d(t) \mid Y(t)) = \frac{E(\Delta D(t) \mid Y(t))}{N} = \frac{1}{N} * \frac{-d(1-m-d^2)}{(1+d)(1-d)} \tag{3.16}$$

$$E(\Delta d^2(t) \mid Y(t)) = \frac{E(\Delta D^2(t) \mid Y(t))}{N^2} = \frac{1}{N^2} * \frac{2-2m-2d^2-md^2}{(1+d)(1-d)} \tag{3.17}$$

$$E(\Delta m(t) \mid Y(t)) = \frac{E(\Delta M(t) \mid Y(t))}{N} = \frac{1}{N} * \frac{d^2 m}{(1+d)(1-d)} \tag{3.18}$$

$$E(\Delta m^2(t) \mid Y(t)) = \frac{E(\Delta M^2(t) \mid Y(t))}{N^2} = \frac{1}{N^2} * \frac{m(2-2m-d^2)}{(1+d)(1-d)} \tag{3.19}$$

$$E[\Delta d(t) * \Delta m(t) \mid Y(t)] = \frac{E[\Delta D(t) * \Delta M(t) \mid Y(t)]}{N^2} = \frac{1}{N^2} \frac{-dm(1+m)}{(1+d)(1-d)} \tag{3.20}$$

$$Cov(\Delta d(t), \Delta m(t) \mid Y(t)) = \frac{Cov(\Delta D(t), \Delta M(t) \mid Y(t))}{N^2} = \frac{1}{N^2} \frac{-dm(d^4-2d^2+m+1)}{(1+d)^2(1-d)^2} \tag{3.21}$$

3.7 Graphs

Graphs over the population space deal with proportional populations instead of absolute populations, have $d(t)$ on the x axis and $m(t)$ on the y axis, and treat the population as continuous (i.e. $N \to \infty$). The population space is thus displayed as an isoceles triangle, bounded by $m(t) \geqslant 0$, $m(t) + d(t) \leq 1$, $m(t) - d(t) \leq 1$, since total proportions must equal 1.

This first graph shows the expected change from a simultaneous birth-death in vector form over the proportional population space. In other words, it shows the direction of the "drift" in the model. $d(t) = 0$ is clearly an attracting line. $d(t)$ always tends towards 0, and $m(t)$ is always increasing except at $m(t) = 0$.

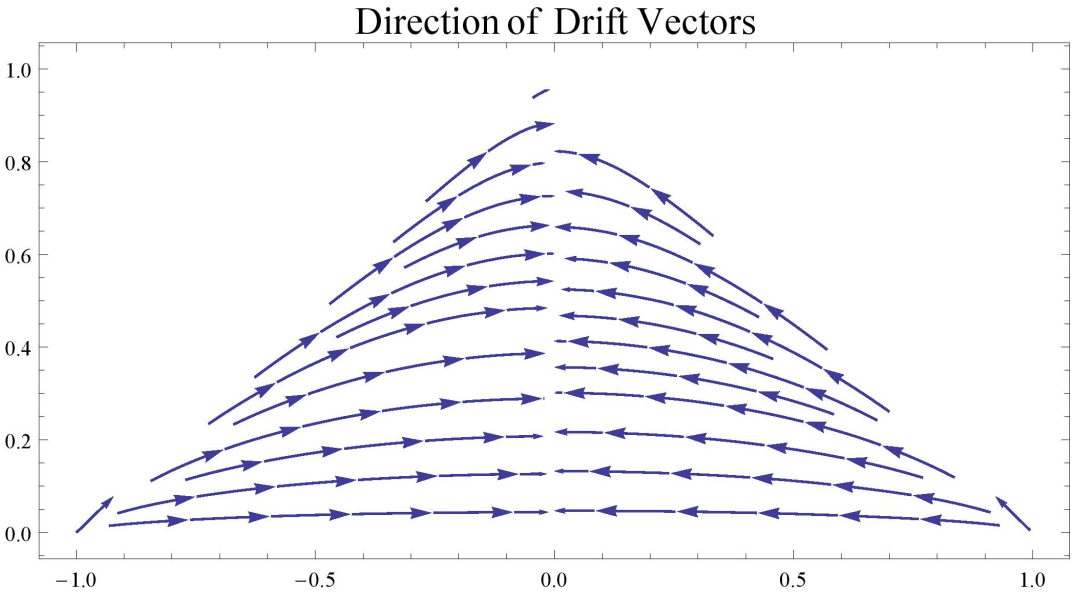

Figure 3.1: Direction of drift vectors. Line $d(t) = 0$ is clearly attracting.

To better understand the second graph, the following density graph shows the magnitude of the vector at each point, that is $\sqrt{E[d(t)]^2 + E[m(t)]^2}$. It shows that the drift decreases as the population approaches the attracting line $d(t) = 0$, and that the drift is strongest close to $d(t) = \pm 1$.

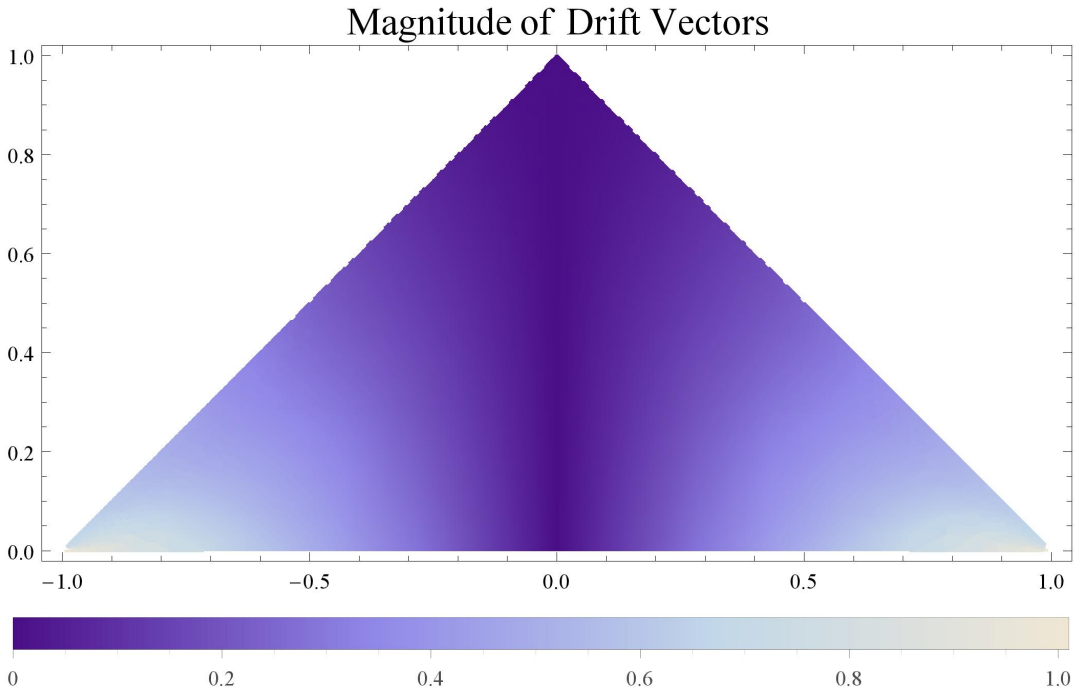

Figure 3.2: $\sqrt{E[d(t)]^2 + E[m(t)]^2}$, Magnitude of drift vectors

Chapter 4

Simulation Results

4.1 Introducing the Discrete-Time Markov Chain Simulations

The purpose of simulations for this model are three-fold: to better understand population behavior of the Markov process over time, to help choose an appropriate approximation, and to judge the accuracy of that approximation.

In order to simplify computations, simulations were run for a discrete-time Markov chain with the exact same birth-death probabilities as the continuous-time model. Recalling theorem 2.1.2, we know that the only difference is in the time of the occurences of the birth-deaths: while in the original model birth-deaths occur in mutually independent exponential time intervals with mean $\frac{1}{\beta}$, in the simulations birth-deaths occur at constant time-intervals of length 1. We shall see that the 2 quantities we will use to compare the simulations and the approximation of the original model are identical for the discrete-time and continous-time models.

4.2 Variables of Interest

For the continuous-time Markov chain, let $T_K = min\{t \geq 0 : K(t) = 0\}$, where $K = C, H$ or M. Define $T_K = \infty$ if K never goes extinct. Thus T_C, T_H and T_M represent the time until extinction of specific species.

Let $\tau = T_C \wedge T_H \wedge T_M$, thus the time to first extinction.

Finally let $h(K) = P(T_K < T_J \; \forall J \neq K) = P(\tau = T_K)$. This is the probability that the first extinction is of species K.

For the discrete-time Markov chain, let T_K^* and τ^* be analogous to their continuous-time counterparts. Therefore

$$T_K = \sum_{i=1}^{T_K^*} Z_i \text{ and } \tau = \sum_{i=1}^{\tau^*} Z_i \tag{4.1}$$

where Z_i 's are i.i.d. exponential random variables with mean $\frac{1}{\beta}$.

Since $h(K)$ relates to the birth-death occurences instead of time, it is identically distributed for both discrete- and continous-time Markov chains and as such will not be subject to a different notation.

The 2 quantities we will be using to evaluate the approximation are $h(M)$ and $E[\tau]$. Note that

$$E[\tau] = E[\sum_{i=1}^{\tau^*} Z_i] = E[E[\sum_{i=1}^{\tau^*} Z_i \mid \tau^*]] = \beta * E[\tau^*] \tag{4.2}$$

so for $\beta = 1$ $E[\tau]$ should be equal to its discrete-time counterpart $E[\tau^*]$, allowing a proper comparison between simulations and the approximation.

4.3 Notation for Observed Data

Since this chapter will be dealing with the data observed from the simulations, we will add new notations to distinguish these from the random variables.

Tilde will be used to denote a single outcome of observed data (ex: $\tilde{\tau}^*$).

Bar will be used to denote the average of the observed data points (ex: $\overline{\tau}^* = \sum_{i=1}^{N} \tilde{\tau}_i^*/n$).

4.4 Effect of the Size of N

Simulations were run until time τ^* with N=100 (1000 runs), N=1000 (1000 runs) and N=10 000 (100 runs), with the inital point $x(t) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. A population of size 10 000 is the upper limit of what a personal computer can handle in a reasonable time frame.

Probability that first extinction is of species K: $\overline{h(K)}$

First extinction probabilities $\overline{h(K)}$ remained similar across all 3 total population sizes N. $\overline{h(M)} > \overline{h(C)}$ and $> \overline{h(H)}$, so M is more likely to go extinct first for all N when starting from $x(t) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. We should note, however, that this does not necessarily mean that M is the least likely sole survivor. Recall that the model stays near $D = 0$, so when the first species to go extinct is C we can expect $M \approx N$ and $H \approx 0$. Similarly when the first species to go extinct is H almost the entire population consists of omnivores. So the probability that M is the sole surviving species is $\approx 1 - h(M) = h(C) + h(H)$, while the probability that C is the sole surviving species is $\approx h(M)/2$ and similarly for H.

$\overline{h(K)}$	N=100	N=1000	N=10000
M	0.419	0.417	0.38
C	0.287	0.295	0.28
H	0.294	0.288	0.34

Table 4.1: Probabilities of first extinction for different N

Times of first extinction: $\tilde{\tau}^*$

By scaling $\tilde{\tau}^*$ by N^{-2}, we arrive at very similar values for different total population sizes N. $\overline{\tau}^* \approx 0.48 * N^2$, and the square factor explains why scaling up N affects the time it takes to

run simulations so drastically. The scaling by a factor of N^2 is the same as that of Brownian motion, suggesting that a diffusion approximation might be appropriate. Note that only 100 simultions were performed for $N = 10000$ instead of the usual 1000 runs because of how slow simulations are at high populations.

For all sizes of N, $\overline{T_M^*} < \overline{\tau}^*$. In other words, if the omnivores go extinct first, they tend to do so in less time than if either the carnivores or herbivores go extinct first. We also see that $\overline{T_C^*}/N^2$ and $\overline{T_H^*}/N^2$ grow as N becomes larger, despite the inverse factor of N^2. Although this might partly be due to chance (since the number of runs for $N = 10000$ was much smaller than for other N because of computational time), it also suggests that scaling by N^{-2} is not exact.

N	$\overline{\tau^*}/N^2$	$\overline{T_M^*}/N^2$	$\overline{T_C^*}/N^2$	$\overline{T_H^*}/N^2$
N=100	0.4574	0.4192	0.4749	0.4950
N=1000	0.4862	0.4208	0.5252	0.5409
N=10000	0.5121	0.3215	0.6400	0.6198

Table 4.2: Average scaled times of first extinction for different N

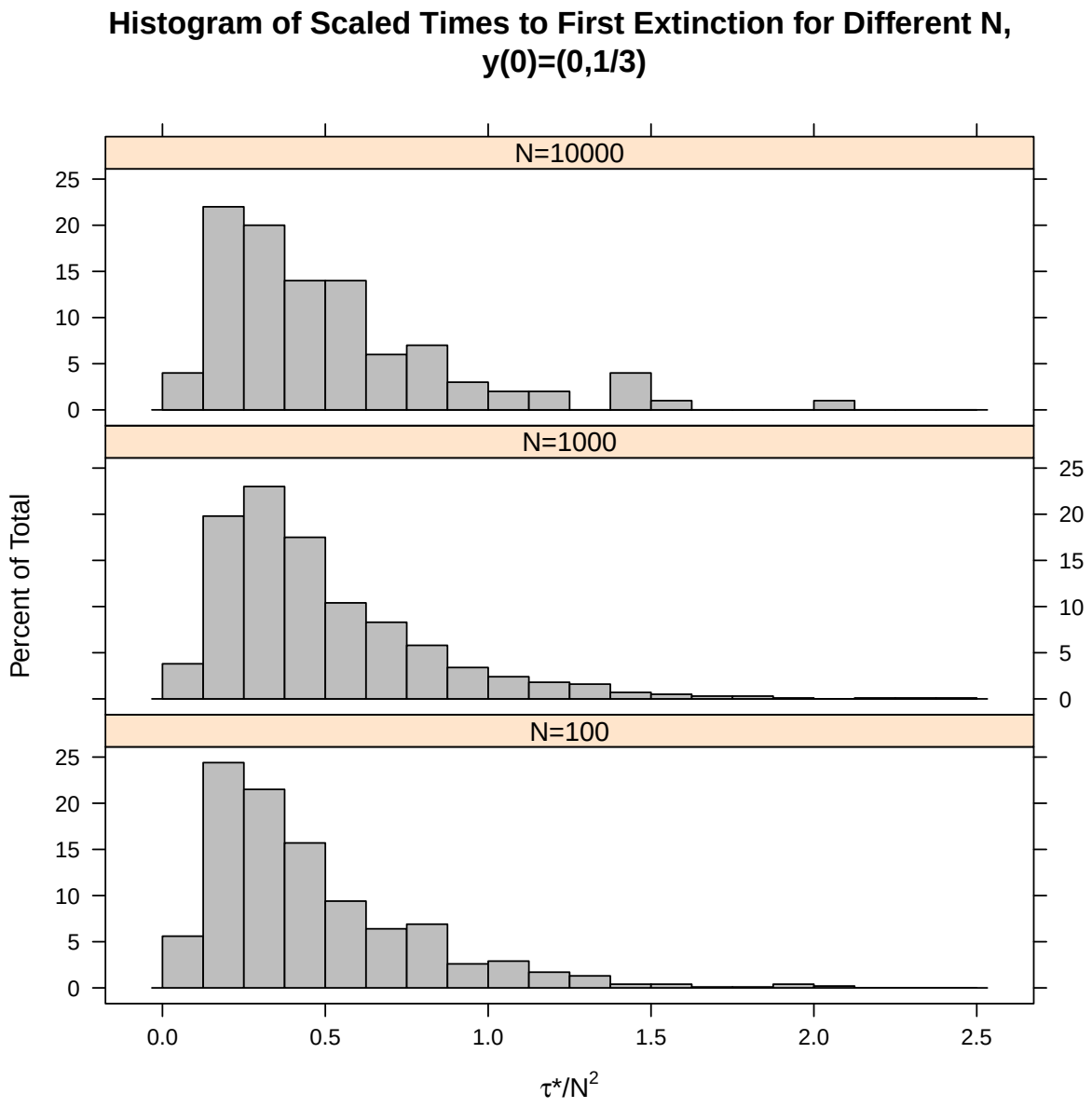

Figure 4.1: Histograms of scaled times to first extinction for different N. These suggest that time is scaled by a factor of N^2.

4.5 Simulations starting on the $D = 0$ line

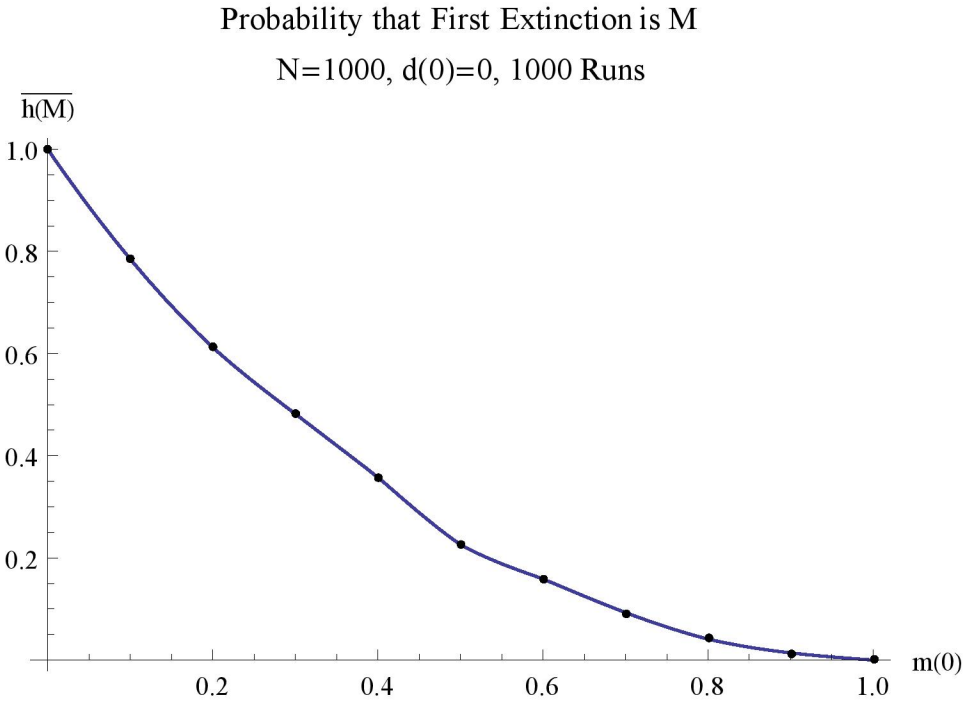

Figure 4.2: Probability that the first extinction is species M

Simulations were run starting on the $D = 0$ stable line with total population $N = 1000$ from initial proportional populations $Y(0)/N = \{(0, \frac{1}{10}), (0, \frac{2}{10}), ..., (0, \frac{9}{10})\}$.

For the probability that M is the first extinction, we can see from figure 4.2 that as $m(0)$ increases, $\overline{h(M)}$ steadily decreases. This is to be expected, since a larger share of the total population should render a population less susceptible to extinction.

For the time until first extinction as shown in figure 4.3, $\overline{\tau^*}$ starts at 0 at $m(0) = 0$, increases with $m(0)$ to a humped peak around $m(0) = 0.4$, then slowly decreases back to 0 at $m(0) = 1$. It is to be expected that the further $m(0)$ is from any absorbing boundary, the larger $\overline{\tau^*}$ will be.

It should be noted that regardless of $x(0)$, $\tilde{\tau}^*$ is very long tailed, a fact captured by the histogram in figure 4.4 as well as by the fact that the median is smaller than the mean.

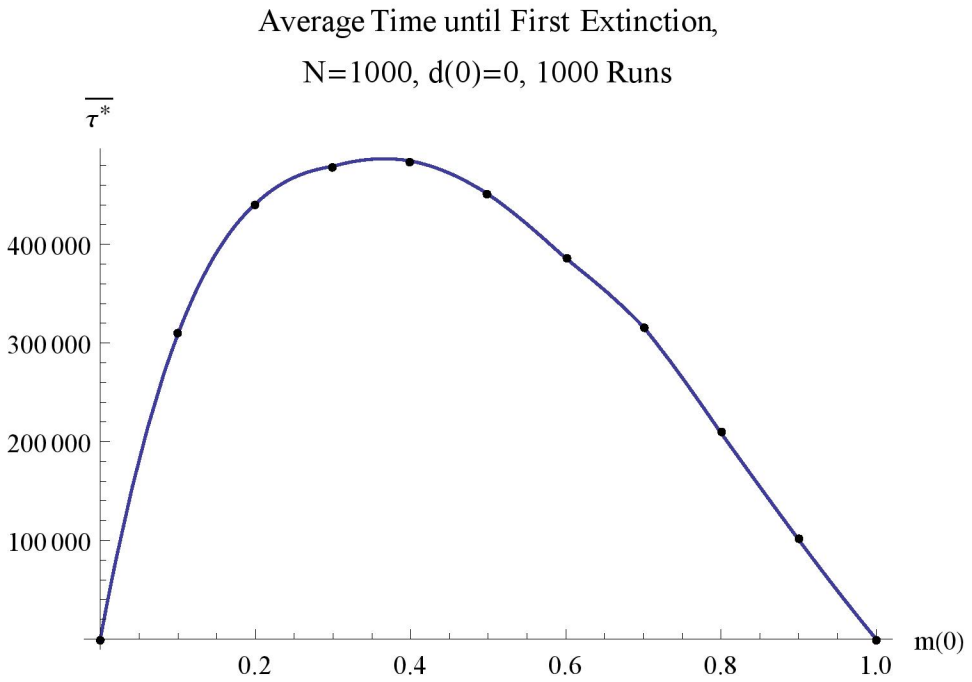

Figure 4.3: Average time until first extinction

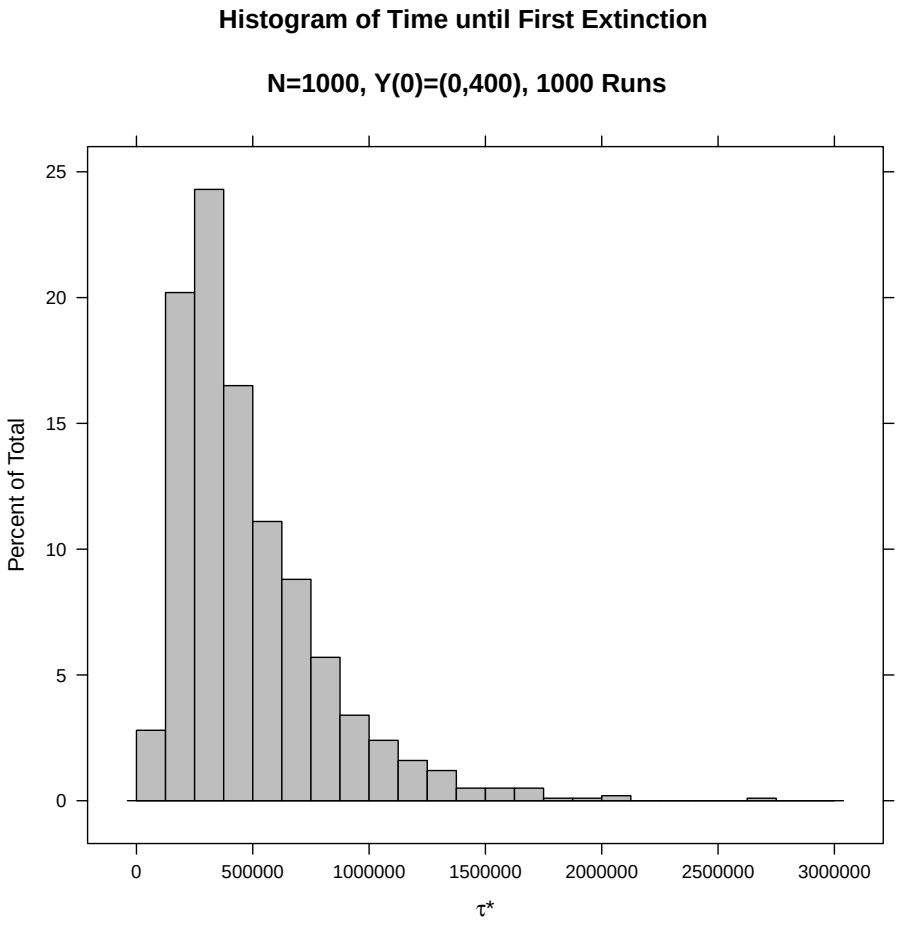

Figure 4.4: Histogram of $\tilde{\tau}^*$, the time until the first extinction occurs

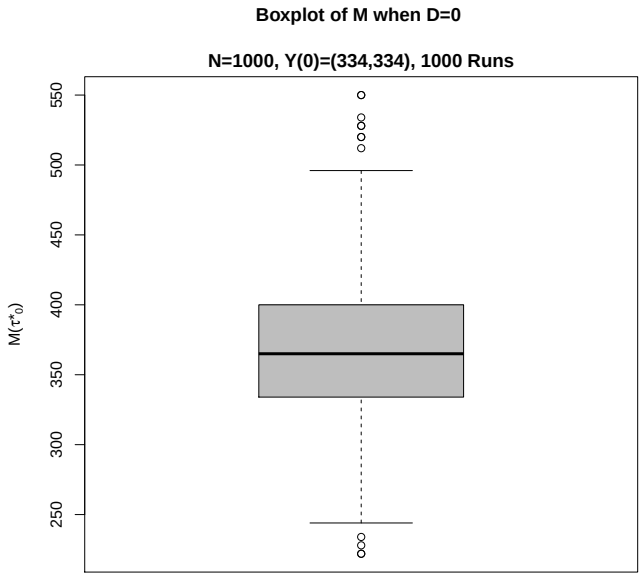

Figure 4.5: Boxplot of M when line $D = 0$ is first reached. Note the upward drift from the initial value $M = 334$.

4.6 Simulations with $D(0) \neq 0$

Define $T^*_{D=0}$ as the time until the discrete Markov Chain first reaches $D = 0$. Define $\tau^*_0 = \tau^* \wedge T^*_{D=0}$, that is as the time until first extinction or $D = 0$, whichever comes first.

All of the simulations in this section were run until $\tilde{\tau}^*_0$. 1000 runs were performed for each initial point $Y(0)$.

Boundary at $\tilde{\tau}^*_0$	(334,334)	(480,480)	(490,490)	(500,50)	(500,20)	(500,10)	(970,10)	(985,5)
$P\{M(\tilde{\tau}^*_0) = 0\}$	0	0	0	0	0.005	0.09	0	0.025
$P\{H(\tilde{\tau}^*_0) = 0\}$	0	0.005	0.05	0	0	0	0	0
$P\{D(\tilde{\tau}^*_0) = 0\}$	1	0.995	0.95	1	0.995	0.91	1	0.975

Table 4.3: Probabilities of first boundary hit for different initial populations

Even with a small $H(0)$ or $M(0)$, $D = 0$ is almost always hit before the first extinction. In the simulations run with $H(0) = 20$ or $M(0) = 20$, $P\{D(\tilde{\tau}^*_0) = 0\} > 0.99$. Therefore in the interior of the triangle away from $D(t) = 0$, the drift seems to override the noise.

It would seem reasonable to assume that for large N, unless $c(0)$, $h(0)$ or $m(0) \approx 0$, then $P\{D(\tilde{\tau}_0^*) = 0\} = 1$. In other words, for large N any initial point not extremely close to the absorbing boundaries will end up reaching the line $D = 0$.

At $Y(0) = (334, 334)$ near the center of the half-triangle, $P\{D(\tilde{\tau}_0^*) = 0\} = 1$. The box plot in figure 4.5 shows the value of M when $D = 0$ was first reached, that is $M(\tilde{\tau}_0^*)$. $\overline{M(\tilde{\tau}_0^*)} = 368.2$, indicating the presence of a slight drift towards greater M during the time that the population approached $D = 0$. And while half the $M(\tilde{\tau}_0^*)$ values fall quite close to the median, the tails are quite long.

Again in simulations with 1000 runs starting with $Y(0) = (334, 334)$, $\tilde{T}_{D=0}^*$ was very small compared to the $\tilde{\tau}^*$ calculated previously at the same N, as can be seen in the histogram in 4.6. $\overline{T_{D=0}^*} = 4803$ turns, is about $\frac{1}{100}$th of $\overline{\tau^*}$. In other words, the model moves very quickly (relative to the time of first extinction) to the stable line $D = 0$.

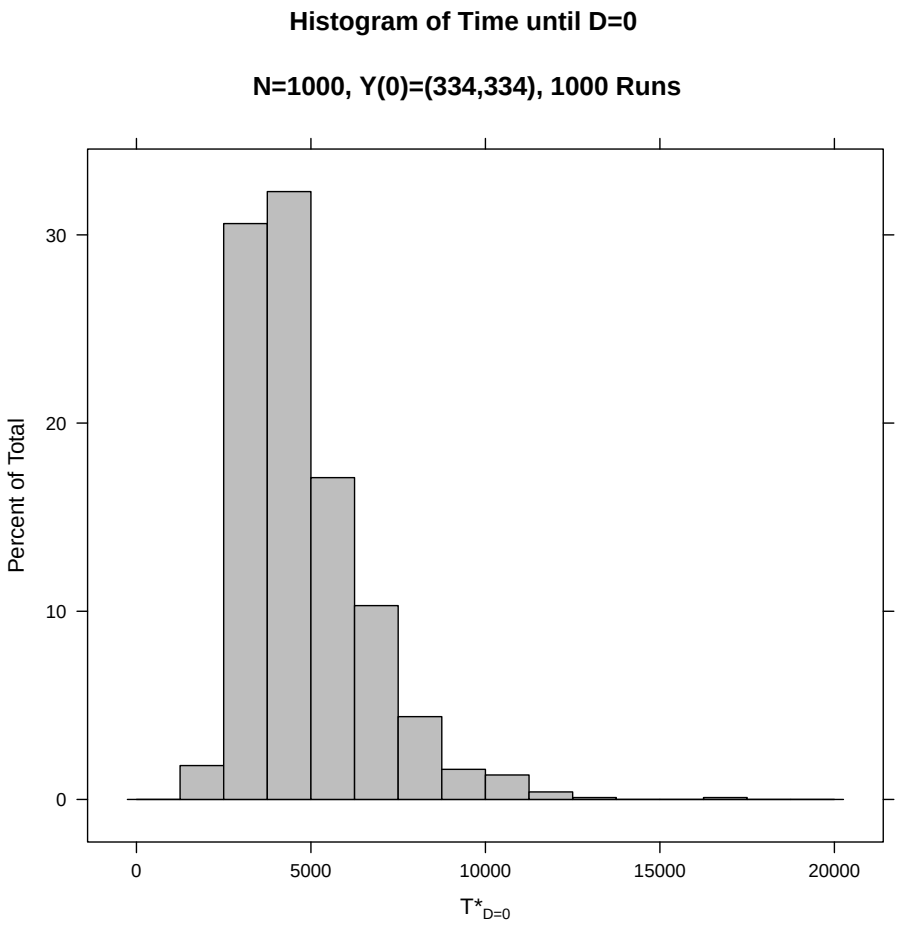

Figure 4.6: Histogram of the time until line $D = 0$ is first reached. Note that the values are very small compared to the time until first extinction as shown in figure 4.4

4.7 Proportion of Time Spent close to D=0 before First Extinction

With initial population $X(0) = (333, 333, 334)$ and 100 runs until time of first extinction $\tilde{\tau}^*$, the simulation spends almost half its time within $-20 \leq D \leq 20$, and 89% within $-50 \leq D \leq 50$. In other words, once the population has reached parity between herbivores and carnivores, it deviates very little from that parity until the first exctinction occurs.

Area	% of time spent	Cumulative %
$\lvert D\rvert \leq 10$	26.35%	26.35%
$10 < \lvert D\rvert \leq 20$	22.41%	48.76%
$20 < \lvert D\rvert \leq 30$	18.27%	67.03%
$30 < \lvert D\rvert \leq 40$	13.33%	80.36%
$40 < \lvert D\rvert \leq 50$	8.86%	89.22%
$50 < \lvert D\rvert \leq 100$	10.65%	99.86%
$100 < \lvert D\rvert$	0.14%	100.00%

Table 4.4: Proportion of time spent in different bands around $D = 0$

Chapter 5

ODE Convergence

5.1 Generator of the rescaled Markov Process

Recalling definition 2.1.5, the generator of a Markov chain is

$$(Lf)(x) = \sum_y \beta(x, y)(f(y) - f(x)) \tag{5.1}$$

Take the Markov process $N^{-1}Y(Nt)$ where $Y(t) = (D(t), M(t))$; we rescaled the population count from absolute numbers to proportions, and multiplied time by the same factor of N. According to the previous definition, the generator for this process is

$$(Lf)(y) = N * \sum_i \beta((d, m), (d + \Delta d_i, m + \Delta m_i))(f(y + (\Delta d_i, \Delta m_i)) - f(y)) \tag{5.2}$$

By approximating $f(y + (\Delta d_i, \Delta m_i))$ by its Taylor expansion, we get

$$\begin{aligned}(\tilde{L}f)(y) = N * \sum_i \beta((d, m), (d + \Delta d_i, m + \Delta m_i)) * (\partial_d f(d, m)\Delta d_i + \partial_m f(d, m)\Delta m_i \\ + \frac{1}{2}\partial_{dd} f(d, m)(\Delta d_i)^2 + \frac{1}{2}\partial_{mm} f(d, m)(\Delta m_i)^2 + \partial_{dm} f(d, m)(\Delta d_i)(\Delta m_i))\end{aligned} \tag{5.3}$$

As in section 3.5, the jump rates equal their corresponding probabilities for calculating

the simultaneous birth-death changes in population. Substituting in expectation we get

$$(\tilde{L}f)(y) = N\Big(\partial_d f(d,m)E[\Delta d \mid y] + \partial_m f(d,m)E[\Delta m \mid y] \\ + \frac{1}{2}\partial_{dd} f(d,m)E[(\Delta d)^2 \mid y] + \frac{1}{2}\partial_{mm} f(d,m)E[(\Delta m)^2 \mid y] + \partial_{dm} f(d,m)E[(\Delta d)(\Delta m) \mid y]\Big) \tag{5.4}$$

Recalling theorem 2.2.1, a 2-dimensional diffusion process has generator

$$(Lf)(x,y) = \partial_x f(x,y)b_x + \partial_y f(x,y)b_y + \frac{1}{2}\partial_{xx} f(x,y)a_{xx} + \frac{1}{2}\partial_{yy} f(x,y)a_{yy} + \partial_{xy} f(x,y)a_{xy} \tag{5.5}$$

where b_i are values from the drift coefficient vector and a_{ij} are values from the diffusion coefficient matrix, which is nonnegative definite.

Therefore the Markov process $N^{-1}Y(Nt) = (d(Nt), m(Nt)$ approximates to a diffusion process with drift coefficients equal to its first moments and diffusion coefficients equal to its second moments multiplied by a factor of N^{-1}.

When we substitute in the appropriate values of d and m, we get

$$\begin{aligned}(\tilde{L}f)(d,m) =& \partial_d f(d,m)\frac{-d(1-m-d^2)}{(1+d)(1-d)} + \partial_m f(d,m)\frac{d^2 m}{(1+d)(1-d)} \\ &+ \frac{1}{2N}\partial_{dd} f(d,m)\frac{2-2m-2d^2-md^2}{(1+d)(1-d)} + \frac{1}{2N}\partial_{mm} f(d,m)\frac{m(2-2m-d^2)}{(1+d)(1-d)} \\ &+ \frac{1}{N}\partial_{dm} f(d,m)\frac{-dm(1+m)}{(1+d)(1-d)}\end{aligned} \tag{5.6}$$

By approximating with a Taylor expansion, the approach we used to obtain this generator is similar to the Kramers-Moyal Expansion. [7]

5.2 Deterministic Process

We are interested in creating an approximation of this model for large N. Since the first order of our generator is an order of N larger than the second order, our generator is approximately

a first order generator for large N over most of the population space. [10] Theorem 2.2.2 formalizes this idea.

In our case, for $d > o(N^{-1})$, $\lim_{N\to\infty} \tilde{L}f(y) = \partial_d f(y) b_d + \partial_m f(y) b_m$ defines a deterministic process.

However, when d is of the same order as N^{-1}, i.e. for $o(d) = o(N^{-1})$, we have

$$\tilde{L}f(y(t)) \approx \frac{1}{N}\Big(-(1-m)\partial_d f(y) + (1-m)\partial_{dd} f(y) + m(1-m)\partial_{mm} f(y)\Big) + o(N^{-2}) \quad (5.7)$$

which is clearly not a first-order generator.

5.3 ODE solution

The deterministic process obtained from the first order generator defines a system of ordinary differential equations (ODEs). These ODEs, which can be thought of as the expected simultaneous birth-death change in population, define a velocity vector (which includes both direction and the magnitude) of any point in the interior of the triangle. Here is the ODE system:

$$\begin{aligned} \frac{\partial d}{\partial t} &= \frac{-d(1-m-d^2)}{(1+d)(1-d)} \\ \frac{\partial m}{\partial t} &= \frac{d^2 m}{(1+d)(1-d)} \end{aligned} \quad (5.8)$$

We used "∂" instead of "d" to denote derivatives to avoid confusion with variable d. Notice that neither d nor m explicitly depend on the independent variable t, making the ODEs autonomous. This allows us to represent them as the derivative of d with respect to m and vice-versa to find a set of solutions independent of time.

$$\begin{aligned} \frac{\partial d}{\partial m} &= \frac{-1+m+d^2}{dm} \\ \frac{\partial m}{\partial d} &= \frac{dm}{-1+m+d^2} \end{aligned} \tag{5.9}$$

$$\begin{aligned} d(m) &= \pm\sqrt{1-2m+Cm^2} \\ m(d) &= \frac{-1+d^2}{-1+\sqrt{1+C(-1+d^2)}}, -\frac{-1+d^2}{1+\sqrt{1+C(-1+d^2)}} \end{aligned} \tag{5.10}$$

When we use the initial conditions $d = d_0$ and $m = m_0$ to derive constant C and then apply the boundary conditions of the population space, we obtain the following curve as a solution to the ODE. The following equation describes the same curve, defined as either a function of d or m.

$$\begin{aligned} d(m) &= \pm\sqrt{1-2m+\frac{(-1+2m_0+d_0^2)m^2}{m_0^2}} \\ m(d) &= \frac{1-d^2}{1+\sqrt{1+\frac{(-1+2m_0+d_0^2)(-1+d^2)}{m_0^2}}} \end{aligned} \tag{5.11}$$

5.4 Lyapunov Stable Line

Let $y = (d(t), m(t))$ represents the proportional population at time t of the ODE system. Denote the derivative of this function with respect to time by

$$g(y) = \dot{y}(t) \tag{5.12}$$

For our system, $g(y)$ is given by (5.8), and $g(y^*) = 0$ for $y = (0, m)\ \forall m$. Therefore we have a line of fixed points at $d = 0$. Other fixed points are the absorbing points $(-1, 0)$ and $(1, 0)$, since they were defined as having 0 drift.

Proposition 5.4.1. *Let $W = \{(0, m), m \in [0, 1]\}$, then W is a manifold of fixed points, and*

$V(d, m) = d^2$ is a Lyapunov function of manifold W for system (5.8) *over the whole population space excluding absorbing points* $(-1, 0)$ *and* $(1, 0)$*. Further, W is an asymptotically stable manifold.*

Proof: Since W consists of all the points y^* in the space such that $d = 0$, clearly $V(y^*) = 0$. Further, since the population is always real-valued, $V(d, m) > 0$ for all other values in the population space. Finally, notice that the ODE $\partial d/\partial t$ is negative for $d > 0$, implying that a positive d decreases over time; similarly, $\partial d/\partial t$ is positive for $d < 0$, implying implying that a negative d increases over time. So d^2 will decrease over time over the whole population space (excluding absorbing points $(-1, 0)$ and $(1, 0)$) except at $d = 0$. Therefore, according to definition 2.3.2, $V(d, m) = d^2$ is Lyapunov, which implies that W is an asymptotically stable manifold according to theorem 2.3.1.

5.5 Another rescaling

Our first rescaling of the Markov process $N^{-1}Y(Nt) = y(Nt)$ approximated to a diffusion which degenerated to an ODE over most of the population space except at $d = 0$. When close to $d = 0$, we obtain a diffusion process with coefficients of the order N^{-1}. To better study what happens around the asymptotically stable line, we need to further speed up time. This approach is corroborated by the simulations, in which the time until the process hits the stable line is very small compared to the time until extinction.

Therefore we take a second rescaling of the Markov process $N^{-1}Y(N^2t)$. This gives us a

new generator

$$\begin{aligned}(Lf)(y) =& N * (\tilde{L}f)(y) \\ =& N * \partial_d f(d,m)\frac{-d(1-m-d^2)}{(1+d)(1-d)} + N * \partial_m f(d,m)\frac{d^2 m}{(1+d)(1-d)} \\ &+ \frac{1}{2}\partial_{dd} f(d,m)\frac{2-2m-2d^2-md^2}{(1+d)(1-d)} + \frac{1}{2}\partial_{mm} f(d,m)\frac{m(2-2m-d^2)}{(1+d)(1-d)} \\ &+ \partial_{dm} f(d,m)\frac{-dm(1+m)}{(1+d)(1-d)}\end{aligned} \tag{5.13}$$

5.6 Expected distance from the ODE solution

We will now use a theorem to show that the maximum expected distance between the diffusion and the ODE curve goes to 0 with N.

Theorem 5.6.1. *Let $Y_t = (D_t, M_t)$ be the diffusion in (5.13) and Y_t^0 the solution of (5.8), both starting from (D_0, M_0). Let $0 < \epsilon < 1$. There is a constant γ dependent on ϵ such that if N is large then for all (D_0, M_0) such that $|D_0| < 1 - \epsilon$, $0 < M_0 < |1 - D_0|$, we have*

$$\mathbb{E}(\sup_{0 \le t \le (\gamma \log N)/N} |Y_t - Y_t^0|^2) \le N^{-1/2} \tag{5.14}$$

Proof: We use the proof [6] of Durrett and Popovic, with only 2 slight modifications. In the drift and diffusion coefficients, we replace d by $(-1+\epsilon/2) \vee d \wedge (1-\epsilon/2)$ to obtain Lipschitz continuity of the coefficients. The second modification involves a different bound for the maximal inequality: since $a_{ij} \le 8/\epsilon^2 \; \forall \; i, j$ we get

$$\mathbb{E}(\sup_{0 \le s \le t} |Y_s - Y_s^0|^2) \le 4\mathbb{E}|Y_t - Y_t^0|^2 \le \frac{32t}{\epsilon^2 N} \tag{5.15}$$

Neither of these differences materially change the rest of the calculations, so the rest of the proof remains the same as the original.

Remark: Theorem 5.6.1 implies that the expected distance between the diffusion in (5.13) and the ODE described in (5.8) goes to zero in expectation for large N until the

diffusion reaches the stable line $d = 0$. This justifies our treatment of the model as an ODE for large N over most of the population space.

Chapter 6

Convergence to 1D Diffusion

6.1 Convergence Theorem

In this chapter we will show that the 2-dimensional diffusion in (5.13) stays close to the asymptotically stable line $d = 0$, and that it converges to a 1-dimensional diffusion process along that line. We will present the theorem at the end of this section and spend the rest of the chapter proving it and calculating the coefficients of this new 1-dimensional diffusion. The steps are similar to those taken in a paper by Durrett and Popovic to prove Theorem 2 [6].

We had previously represented the model as a system of ordinary differential equations in (5.8) over most of the population space. Further we had shown that line $d = 0$ is asymptotically stable over the whole population space excluding sole survivor points $(-1, 0)$ and $(1, 0)$, which implies that all other points are on an ODE trajectory which leads to $d = 0$. Let $\Phi(d, m)$ be the map that takes a point $y = (d, m)$ to the destination of the ODE trajectory on the asymptotically stable line $d = 0$. Let $m^*(d, m)$ be the m-coordinate of point (d, m) when it follows the ODE to the stable line $d = 0$. Then $\Phi(d, m) = (0, m^*)$. Applying the projection to the semimartingale Y_t gives $\Phi(D_t, M_t) = (0, M_t^*)$.

Theorem 6.1.1. *Consider the diffusion* $Y_t = (D_t, M_t)$ *in* (5.13). *Let* $\tau = \inf\{t : M_t =$

0 *or* $D_t = \pm 1\}$ *be the time of first extinction. Let* $0 < \delta < 1/2$. *Suppose* $|D_0| \leq N^{-\delta}$. *Then if* N *is large, with high probability we have* $|D_t| \leq 2N^{-\delta}$ *for all* $t \leq \tau$. *Also, as* $N \to \infty$ *the process* M_t^* *converges in distribution to a diffusion process on the aymptotically stable line* $d = 0$.

6.2 Ito's formula for our 2-dimensional diffusion

To apply Ito's formula, we write the 2-dimensional diffusion in (5.13) as a stochastic differential equation, following theorem 2.2.3. $Y_t = (D_t, M_t)$ is a vector semimartingale, $B_s = (B_s^d, B_s^m)$ is a 2-dimensional Brownian motion whose components B_s^d and B_s^m are independent and

$$
\begin{aligned}
Y_t &= Y_0 + \int_0^t \sigma(s) dB_s + \int_0^t b(s) ds \\
\begin{pmatrix} D_t \\ M_t \end{pmatrix} &= \begin{pmatrix} D_0 \\ M_0 \end{pmatrix} + \int_0^t \begin{pmatrix} \sigma_{dd} & \sigma_{dm} \\ \sigma_{md} & \sigma_{mm} \end{pmatrix} \times d \begin{pmatrix} B_s^d \\ B_s^m \end{pmatrix} + \int_0^t \begin{pmatrix} b_d \\ b_m \end{pmatrix} ds
\end{aligned} \tag{6.1}
$$

Here we will use capital D_s and M_s to denote that they are semimartingales of proportional variables d and m, and not directly related to absolute population values $D(t) = d(t) * N$ or $M(t) = m(t) * N$.

From (5.13) we know that the drift and diffusion coefficients of the 2-dimensional diffusion

are:

$$
\begin{aligned}
b_d &= N * \frac{-d(1-m-d^2)}{(1+d)(1-d)} \\
b_m &= N * \frac{d^2 m}{(1+d)(1-d)} \\
a_{dd} &= \frac{2-2m-2d^2-md^2}{(1+d)(1-d)} \\
a_{mm} &= \frac{m(2-2m-d^2)}{(1+d)(1-d)} \\
a_{dm} &= a_{md} = \frac{-dm(1+m)}{(1+d)(1-d)}
\end{aligned}
\tag{6.2}
$$

Ito's formula uses σ_{ij}, which can be obtained from the diffusion coefficient since $a = \sigma \times \sigma^T$, but since σ_{ij} values are only used as intermediaries and the coefficients of the 1 dimensional diffusion can be represented using a_{ij}, they are not shown here.

From theorem 2.2.4, we get that Ito's formula for a 2-dimensional semimartingale is

$$
\begin{aligned}
F(Y_t) =& F(Y_0) + \int_0^t \partial_d F(Y_s) dD_s + \int_0^t \partial_m F(Y_s) dM_s \\
&+ \frac{1}{2}\int_0^t \partial_d^2 F(Y_s) d\langle D, D\rangle_s + \frac{1}{2}\int_0^t \partial_m^2 F(Y_s) d\langle M, M\rangle_s + \int_0^t \partial_d \partial_m F(Y_s) d\langle D, M\rangle_s
\end{aligned}
\tag{6.3}
$$

Since $\langle Y^i, Y^j\rangle_s = a_{i,j} ds$, we can substitute the quadratic variations for the corresponding values in the diffusion coefficient. When we also substitute dD_s and dM_s from their values in (6.1) we get

$$
\begin{aligned}
F(Y_t) = F(Y_0) &+ \int_0^t \Big(\partial_d F(Y_s) b_d + \partial_m F(Y_s) b_m\Big) ds \\
&+ \int_0^t \Big(\partial_d F(Y_s)\sigma_{dd} + \partial_m F(Y_s)\sigma_{md}\Big) dB_s^d + \int_0^t \Big(\partial_d F(Y_s)\sigma_{dm} + \partial_m F(Y_s)\sigma_{mm}\Big) dB_s^m \\
&+ \frac{1}{2}\int_0^t \Big(\partial_d^2 F(Y_s) a_{dd} + \partial_m^2 F(Y_s) a_{mm} + 2\partial_d \partial_m F(Y_s) a_{dm}\Big) ds
\end{aligned}
\tag{6.4}
$$

6.3 Applying Ito's formula to the projection map

To apply Ito's formula to the projection map, we will need to obtain m^* from $m(0)$ in 5.11 and then calculate its derivatives.

$$m^* = \frac{1}{1+\sqrt{1-\frac{(-1+2m+d^2)}{m^2}}} = \frac{m}{m+\sqrt{(1-m)^2-d^2}} \tag{6.5}$$

$$\begin{aligned}
\frac{\partial m^*}{\partial d} =& \frac{dm}{\sqrt{(1-m)^2-d^2}\left(\sqrt{(1-m)^2-d^2}+m\right)^2} \\
\frac{\partial^2 m^*}{\partial d^2} =& \frac{m\left(m^2\left(\sqrt{(m-1)^2-d^2}-2\right)-2m\sqrt{(m-1)^2-d^2}+\left(2d^2+1\right)\sqrt{(m-1)^2-d^2}+m^3+m\right)}{\left((m-1)^2-d^2\right)^{3/2}\left(\sqrt{(m-1)^2-d^2}+m\right)^3} \\
\frac{\partial m^*}{\partial m} =& \frac{-d^2-m+1}{\sqrt{(m-1)^2-d^2}\left(\sqrt{(m-1)^2-d^2}+m\right)^2} \\
\frac{\partial^2 m^*}{\partial m^2} =& -\frac{2d^4+d^2\left(-3m\sqrt{(m-1)^2-d^2}+2\sqrt{(m-1)^2-d^2}-3m^2+6m-4\right)}{\left((m-1)^2-d^2\right)^{3/2}\left(\sqrt{(m-1)^2-d^2}+m\right)^3} \\
& -\frac{2(m-1)^2\left(\sqrt{(m-1)^2-d^2}+m-1\right)}{\left((m-1)^2-d^2\right)^{3/2}\left(\sqrt{(m-1)^2-d^2}+m\right)^3} \\
\frac{\partial^2 m^*}{\partial m \partial d} =& -\frac{dm\left(m\left(2\sqrt{(m-1)^2-d^2}-3\right)-\sqrt{(m-1)^2-d^2}-d^2+1\right)}{\left((m-1)^2-d^2\right)^{3/2}\left(\sqrt{(m-1)^2-d^2}+m\right)^3} \\
& -\frac{d\left(d^2-1\right)\left(\sqrt{(m-1)^2-d^2}+2m^3\right)}{\left((m-1)^2-d^2\right)^{3/2}\left(\sqrt{(m-1)^2-d^2}+m\right)^3}
\end{aligned} \tag{6.6}$$

Recall that the projection of the semimartingale Y_t gives $\Phi(D_t, M_t) = (0, M_t^*)$. If we take

function $m_s^* = m^*(D_s, M_s)$, then by applying Ito's formula to the ODE we get the following

$$\begin{aligned} M_t^* = M_0^* &+ \int_0^t \Big(\frac{\partial m_s^*}{\partial d} b_d + \frac{\partial m_s^*}{\partial m} b_m\Big) ds \\ &+ \int_0^t \Big(\frac{\partial m_s^*}{\partial d} \sigma_{dd} + \frac{\partial m_s^*}{\partial m} \sigma_{md}\Big) dB_s^d + \int_0^t \Big(\frac{\partial m_s^*}{\partial d} \sigma_{dm} + \frac{\partial m_s^*}{\partial m} \sigma_{mm}\Big) dB_s^m \\ &+ \frac{1}{2} \int_0^t \Big(\frac{\partial^2 m_s^*}{\partial d^2} a_{dd} + \frac{\partial^2 m_s^*}{\partial m^2} a_{mm} + 2\frac{\partial^2 m_s^*}{\partial d \partial m} a_{dm}\Big) ds \end{aligned} \tag{6.7}$$

In the previous equation, the constant N is only present from the drift coefficients b_d and b_m. The constant N disappears since

$$\begin{aligned} &\frac{\partial m_s^*}{\partial d} b_d + \frac{\partial m_s^*}{\partial m} b_m = \\ &\qquad \frac{dm}{\sqrt{(1-m)^2 - d^2}\left(\sqrt{(1-m)^2 - d^2} + m\right)^2} * N * \frac{-d(1-m-d^2)}{(1+d)(1-d)} \\ &+ \frac{-d^2 - m + 1}{\sqrt{(m-1)^2 - d^2}\left(\sqrt{(m-1)^2 - d^2} + m\right)^2} * N * \frac{d^2 m}{(1+d)(1-d)} = 0 \end{aligned} \tag{6.8}$$

which leaves us with the following equation, with no factors of N:

$$\begin{aligned} M_t^* = M_0^* &+ \int_0^t \Big(\frac{\partial m_s^*}{\partial d} \sigma_{dd} + \frac{\partial m_s^*}{\partial m} \sigma_{md}\Big) dB_s^d + \int_0^t \Big(\frac{\partial m_s^*}{\partial d} \sigma_{dm} + \frac{\partial m_s^*}{\partial m} \sigma_{mm}\Big) dB_s^m \\ &+ \frac{1}{2} \int_0^t \Big(\frac{\partial^2 m_s^*}{\partial d^2} a_{dd} + \frac{\partial^2 m_s^*}{\partial m^2} a_{mm} + 2\frac{\partial^2 m_s^*}{\partial d \partial m} a_{dm}\Big) ds \end{aligned} \tag{6.9}$$

6.4 1-dimensional diffusion and coefficients

Here we will show that equation (6.9) represents a 1D diffusion. The sum of 2 integrals of Brownian motion is also an integral of Brownian motion, therefore equation (6.9) is that of a 1D semimartingale.

From equation (6.9), we can conclude that the drift coefficient is

$$b^* = \frac{1}{2}\left(\frac{\partial^2 m_0^*}{\partial d^2} a_{dd} + \frac{\partial^2 m_0^*}{\partial m^2} a_{mm} + 2\frac{\partial^2 m_0^*}{\partial d \partial m} a_{dm}\right) \tag{6.10}$$

To calculate the diffusion coefficient, we can take the quadratic variation of the semi-martingale. Since B_s^1 and B_s^2 are independent, this gives us

$$\begin{aligned}\langle M_t^* \rangle &= \int_0^t \left(\frac{\partial m_s^*}{\partial d}\sigma_{dd} + \frac{\partial m_s^*}{\partial m}\sigma_{md} \right)^2 ds + \int_0^t \left(\frac{\partial m_s^*}{\partial d}\sigma_{dm} + \frac{\partial m_s^*}{\partial m}\sigma_{mm} \right)^2 ds \\ &= \int_0^t \left(\frac{\partial m_s^*}{\partial d} \right)^2 (\sigma_{dd}^2 + \sigma_{dm}^2) + \left(\frac{\partial m_s^*}{\partial m} \right)^2 (\sigma_{md}^2 + \sigma_{mm}^2) ds \\ &+ \int_0^t 2 \frac{\partial m_s^*}{\partial d} \frac{\partial m_s^*}{\partial m} (\sigma_{dd}\sigma_{md} + \sigma_{dm}\sigma_{mm}) ds\end{aligned} \tag{6.11}$$

Recall that $a = \sigma\sigma^T$, which implies $a_{ij}(x) = \sum_k \sigma_{ik}(x)\sigma_{jk}(x)$. Substituting back a_{ij} values gives

$$a^* = \left(\frac{\partial m_0^*}{\partial d} \right)^2 a_{dd} + \left(\frac{\partial m_0^*}{\partial m} \right)^2 a_{mm} + 2 \left(\frac{\partial m_0^*}{\partial d} \right) \left(\frac{\partial m_0^*}{\partial m} \right) a_{dm} \tag{6.12}$$

Plugging in the values from the derivatives of the ODE and the generating function, the coefficients give:

$$b^* = \frac{-m\left(-2d^4 - 2(m-1)(-2 + 3\sqrt{-d^2 + (m-1)^2} + 3m) + d^2(6 - 4m + 3\sqrt{-d^2 + (m-1)^2}m + 3m^2)\right)}{2(-1+d^2)\sqrt{-d^2 + (m-1)^2}(\sqrt{-d^2 + (m-1)^2} + m)^3} \tag{6.13}$$

$$a^* = -\frac{m\left(-2 + d^2 + 2m\right)}{\left(\sqrt{-d^2 + (-1+m)^2} + m\right)^4} \tag{6.14}$$

6.5 Convergence Near the Stable Line

The Lyapunov function $V(d, m) = d^2$ introduced in section 5.4 can also be used to help study the distance from the asyptotically stable line $d = 0$. Let $\delta > 0$. For all (d, m) in the

neighborhood $|d| \leq N^{-\delta}$ of the fixed points the change in direction of the strong drift is

$$\nabla V \cdot F = 0 + 2d * b_d = \frac{-2d^2(1-m-d^2)}{(1+d)(1-d)} = \frac{-2(1-m-d^2)}{(1+d)(1-d)} * V \leq -\beta V \tag{6.15}$$

where ∇V is the gradient of V, F is the strong drift, and

$$\beta = \inf_{d,m:|d|\leq N^{-\delta}<0.5} \left\{ \frac{2(1-m-d^2)}{(1+d)(1-d)} \right\} > 0 \tag{6.16}$$

We wish to show that the Lyapunov function converges to 0 as the population tends to infinity for all $0 < t < \tau^N$, where $\tau^N = \inf\{t \geq 0 : |D_t| > N^{-\delta}\}$.

For that, we will apply Ito's formula to $e^{N\beta(t\wedge\tau^N)}V(D_{t\wedge\tau^N}, M_{t\wedge\tau^N})$. By taking steps similar to those in section 6.3 but with the inclusion of the product rule, we get

$$\begin{aligned} e^{N\beta(t\wedge\tau^N)}V(D_{t\wedge\tau^N}, M_{t\wedge\tau^N}) &= V(D_0, M_0) \\ &+ \int_0^{t\wedge\tau^N} e^{N\beta s}(N\beta V + V_d b_d + V_m b_m + \frac{1}{2}V_{dd}a_{dd} + \frac{1}{2}V_{mm}a_{mm} + V_{dm}a_{dm})\mathrm{d}s \\ &+ \int_0^{t\wedge\tau^N} e^{N\beta s}(V_d\sigma_{dd} + V_m\sigma_{md})dB_s^d + \int_0^{t\wedge\tau^N} e^{N\beta s}(V_d\sigma_{dm} + V_m\sigma_{mm})dB_s^m \end{aligned} \tag{6.17}$$

By substititing values, which is made easy by the fact that all m-derivatives of V equal 0, and dividing both sides by $e^{N\beta t}$ we get

$$\begin{aligned} V(D_{t\wedge\tau^N}, M_{t\wedge\tau^N}) &= e^{-N\beta(t\wedge\tau^N)}V(D_0, M_0) \\ &+ e^{-N\beta(t\wedge\tau^N)} \int_0^{t\wedge\tau^N} N * e^{N\beta s}(\beta V + \nabla V \cdot F)\mathrm{d}s \\ &+ e^{-N\beta(t\wedge\tau^N)} \int_0^{t\wedge\tau^N} e^{N\beta s} * a_{dd}\mathrm{d}s \\ &+ e^{-N\beta(t\wedge\tau^N)} \int_0^{t\wedge\tau^N} e^{N\beta s} * 2D_s * \sigma_{dd}dB_s^d \\ &+ e^{-N\beta(t\wedge\tau^N)} \int_0^{t\wedge\tau^N} e^{N\beta s} * 2D_s * \sigma_{dm}dB_s^m \end{aligned} \tag{6.18}$$

It is clear that the constant in (6.18) converges to 0. Now we will show the same for each

integral of equation (6.18).

Convergence of Second Integral

Since $D_{t\wedge\tau^N} \leq N^{-\delta}$, we have

$$a_{dd} = \frac{2 - 2M_s - 2D_s^2 - M_s D_s^2}{(1 + D_s)(1 - D_s)} \leq \frac{2}{1 - N^{-2\delta}} = \frac{2N^{2\delta}}{N^{2\delta} - 1} \tag{6.19}$$

So the limit of the second integral of (6.18) is less than

$$\lim_{N\to\infty} \frac{2N^{2\delta}}{N^{2\delta} - 1} * e^{-N\beta t} \int_0^{t\wedge\tau^N} e^{N\beta s} \mathrm{d}s = 0 \tag{6.20}$$

Values of σ_{ij}

For the third and fourth integral, we can apply theorem 2.4.1 on the square root of 2×2 matrices to the identity $a = \sigma \times \sigma^T$ to deal with the σ_{ij} values. σ_{ij} values are real and of the same sign as their corresponding a_{ij} values.

Let $q = \sqrt{a_{dd}a_{mm} - a_{dm}a_{md}} > 0$ and $r = \sqrt{a_{dd} + a_{mm} + 2q}$ and recall that both $a_{dd}, a_{mm} > 0$. Then the following inequalities hold:

$$\begin{aligned} \sigma_{dd} &= \frac{a_{dd} + q}{r} < \sqrt{a_{dd}} \\ \sigma_{dm} &= \frac{a_{dm}}{r}, |\sigma_{dm}| < \frac{a_{dm}}{\sqrt{a_{dd}}} \end{aligned} \tag{6.21}$$

Convergence of Third Integral

For the third integral of (6.18), we have a sequence of random times τ^N and $T \geq 0$ such that the following inequalities hold for all sequence values:

$$I_3 = e^{-N\beta(t\wedge\tau^N)} \int_0^{t\wedge\tau^N} e^{N\beta s} * 2D_s * \sigma_{dd} dB_s^d < \sup_{0\leq t\leq T\wedge\tau^N} e^{-N\beta t} | \int_0^t e^{N\beta s} * 2D_s * \sigma_{dd} dB_s^d | \tag{6.22}$$

We can then use the inequalities found in equations 6.19 and 6.21 to assert

$$I_3 < \sup_{0\le t\le T\wedge\tau^N} e^{-N\beta t} * 2N^{-\delta} * \sqrt{\frac{2N^{2\delta}}{N^{2\delta}-1}}|\int_0^t e^{N\beta s} dB_s^d| \tag{6.23}$$

Next we will use the theorem of maximal inequality applied to Brownian motion to set an upper limit to the probability that the integral surpasses a certain value.

Theorem 6.5.1. *Theorem: Maximal inequality (Brownian Motion).*

If $\int_0^t \mathbb{E}[Y_s^2]\mathrm{d}s < \infty$, *then*

$$P\{\max_{0\le s\le t}|\int_0^t Y_s \mathrm{d}B_s| \ge a\} \le \frac{\mathbb{E}[(\int_0^t Y_s \mathrm{d}B_s)^2]}{a^2} = \frac{\int_0^t \mathbb{E}[Y_s^2]\mathrm{d}s}{a^2} \tag{6.24}$$

Applying the theorem of maximal inequality to (6.23) we get

$$\begin{aligned} P\Big\{\sup_{0\le t\le T\wedge\tau^N} \frac{\sqrt{8}e^{-N\beta t}}{\sqrt{N^{2\delta}-1}}|\int_0^t e^{N\beta s} dB_s^d| \ge a\Big\} &\le \frac{8e^{-2N\beta t}}{a^2(N^{2\delta}-1)}\int_0^t e^{2N\beta s}\mathrm{d}s \\ &\le \frac{4}{N\beta a^2(N^{2\delta}-1)} \end{aligned} \tag{6.25}$$

If we set $a^2 = N^{-1}$ and take the limit as N goes to infinity, we get

$$\lim_{N\to\infty} P\Big\{\sup_{0\le t\le T\wedge\tau^N} \frac{\sqrt{8}e^{-N\beta t}}{\sqrt{N^{2\delta}-1}}|\int_0^t e^{N\beta s} dB_s^d| \ge N^{-\frac{1}{2}}\Big\} = 0 \tag{6.26}$$

This shows us that the third integral of (6.18) converges in probability to 0 by using the following definition.

Definition 6.5.1. Sequence $X_n \to X$ in probability iff $\forall\epsilon > 0$

$$\lim_{n\to\infty} P\{|X_n - X| > \epsilon\} = 0 \tag{6.27}$$

Convergence of Fourth Integral

The fourth integral of (6.18) is identical to the third, except with σ_{dm} instead of σ_{dd}. If we demonstrate that $|\sigma_{dm}| \leq \sigma_{dd} \ \ \forall (d, m)$ in the population space, then the convergence in probability of the third integral (done previously) implies the convergence in probability of the fourth.

From equation 6.21, showing $|a_{dm}| \leq |a_{dd}|$ would imply $|\sigma_{dm}| \leq |\sigma_{dd}|$. Recall from chapter 3 that the population boundaries are $0 \leq m \leq 1$, $|d| \leq 1$ and $m + |d| \leq 1$. Without loss of generality, assume $d \geq 0$ (the following can easily be shown to hold for $d < 0$). Then

$$a_{dd} - |a_{dm}| = \frac{2 - 2m - d(2d + m(1 + d + m))}{(1 + d)(1 - d)} \tag{6.28}$$

By substituting $m + d \leq 1$ we get

$$a_{dd} - |a_{dm}| \geq \frac{2 - 2m - 2d}{(1 + d)(1 - d)} \geq 0 \tag{6.29}$$

Therefore $|\sigma_{dm}| \leq \sigma_{dd} \ \ \forall d, m$ in the population space and the fourth integral converges.

Convergence of the Lyapunov Function

Finally, for the first integral, recall from equation (6.15) that $\nabla V \cdot F + \beta V \leq 0$, so the integral is negative. Since the Lyapunov function V is non-negative and all the other components of (6.18) converge in probability to 0, then so must the first integral. This implies that $V(D_{t \wedge \tau^N}, M_{t \wedge \tau^N}) \xrightarrow{p} 0$, which implies

$$D_{t \wedge \tau^N} \xrightarrow{p} 0 \tag{6.30}$$

We used Ito's formula to obtain M_t, which only implies convergence in law. So we have

$$(D_{t \wedge \tau^N}, M_{t \wedge \tau^N}) \xrightarrow{d} (0, M_{t \wedge \tau^N}) \tag{6.31}$$

6.6 Convergence to the 1D Diffusion

Finally we wish to show that M_t^* converges as N goes to infinity to a 1-dimensional diffusion. In section 6.4 we showed that for each N, the 1-dimensional drift and diffusion coefficients are given by $b^*(D_t, M_t)$ and $a^*(D_t, M_t)$ as defined in (6.13) and (6.14). Since in the neighborhood of the asymptotically stable line the coefficients b^* and a^* are bounded, the sequence of processes is tight, and the convergence in law $(D_{t\wedge\tau^N}, M_{t\wedge\tau^N}) \xrightarrow{d} (0, M^*_{t\wedge\tau^N})$ previously shown combined with Theorem 5.4 in Kurtz and Protter (1991) [11] implies that for all $0 \leq t \leq \tau^N$ we have convergence of M_t^* to the one-dimensional diffusion process.

Chapter 7

Comparison of Simulations and the Diffusion Approximation

The simulation results are plotted in blue, with the actual points shown as black dots. Simulations are based on population $N = 1000$, 9 starting points, 1000 reps. Probabilities and times of extinction for $m = 0$ and $m = 1$ were also included as simulation results. The approximation is plotted in purple, based off 1000 points for which the integrations were calculated, as well as appropriate values for $m = 0$.

7.1 Natural Scale

The definitions in this section are based on Chapter 7 of Durrett's book [5].

Hitting probabilities and occupation times are related to the natural scale ϕ of a diffusion.

$$\phi(x) = \int^x exp\left(\int^y \frac{-2b^*(z)}{a^*(z)}\mathrm{d}z\right)\mathrm{d}y \tag{7.1}$$

$$\frac{-2b^*(z)}{a^*(z)} = -\left(\sqrt{-d^2+(m-1)^2}+m\right)\Big(-2d^4-2(m-1)(-2+3\sqrt{-d^2+(m-1)^2}+3m) + d^2(6-4m+3\sqrt{-d^2+(m-1)^2}m+3m^2)\Big)/\left((-1+d^2)\sqrt{-d^2+(m-1)^2}(-2+d^2+2m)\right) \tag{7.2}$$

So the innermost integral over z to find the natural scale is in fact a 2D integral over the area projected by the ODE onto the line $d = 0$ between $m = y$ and $m = 0$.

This area A corresponds to the area defined by

$$-\sqrt{1-2m+\frac{-1+2y}{y^2}*m^2} \le d \le \sqrt{1-2m+\frac{-1+2y}{y^2}*m^2} \qquad 0 \le m \le y \tag{7.3}$$

Note that when $y = 1$ this corresponds to the whole area of the triangle.

The resulting integral

$$\int_A \frac{-2b^*(d,m)}{a^*(d,m)}\mathrm{d}d\mathrm{d}m \tag{7.4}$$

had to be integrated numerically. Therefore the outermost integral for calculating ϕ also needed to be calculated numerically. A table with 1000 values of $\phi(x)$ was compiled, and $\phi(x)$ was normalized to obtain the probability of m being the first extinction given starting location $(d = 0, m = y)$

$$P_x\left(T_{m=0} < T_{m=1}\right) = \frac{\phi(1)-\phi(x)}{\phi(1)-\phi(0)} = \frac{\phi(1)-\phi(x)}{\phi(1)} \tag{7.5}$$

The natural scale is also used to derive Green's Function, thus the expected time until first extinction.

Define $\tau_x = T_0 \wedge T_1$ starting at $(d = 0, m = x)$. Then

$$E[\tau_x] = \int G(x,y)\mathrm{d}y \tag{7.6}$$

Where $G(x,y)$ is the Green's function for the interval $[0,1]$, thus

$$G(x,y) = \begin{cases} 2\frac{(\phi(1)-\phi(x))(\phi(y)-\phi(0))}{\phi(1)-\phi(0)}\frac{1}{\phi'(x)a^*(x)}, & y \leq x \\ 2\frac{(\phi(x)-\phi(0))(\phi(1)-\phi(y))}{\phi(1)-\phi(0)}\frac{1}{\phi'(x)a^*(x)}, & x < y \end{cases} \tag{7.7}$$

The approximations used in the subsequent graphs are based off 1000 values of $m(0)$ for which the integrations were numerically calculated.

7.2 Probability of first extinction being m over starting value of m

In purple: $h(M)$. In blue: $\overline{h(M)}$

The approximation seems to be a good fit; although it seems to consistently underestimate the probability of M going extinct first, the gap might close if N were larger.

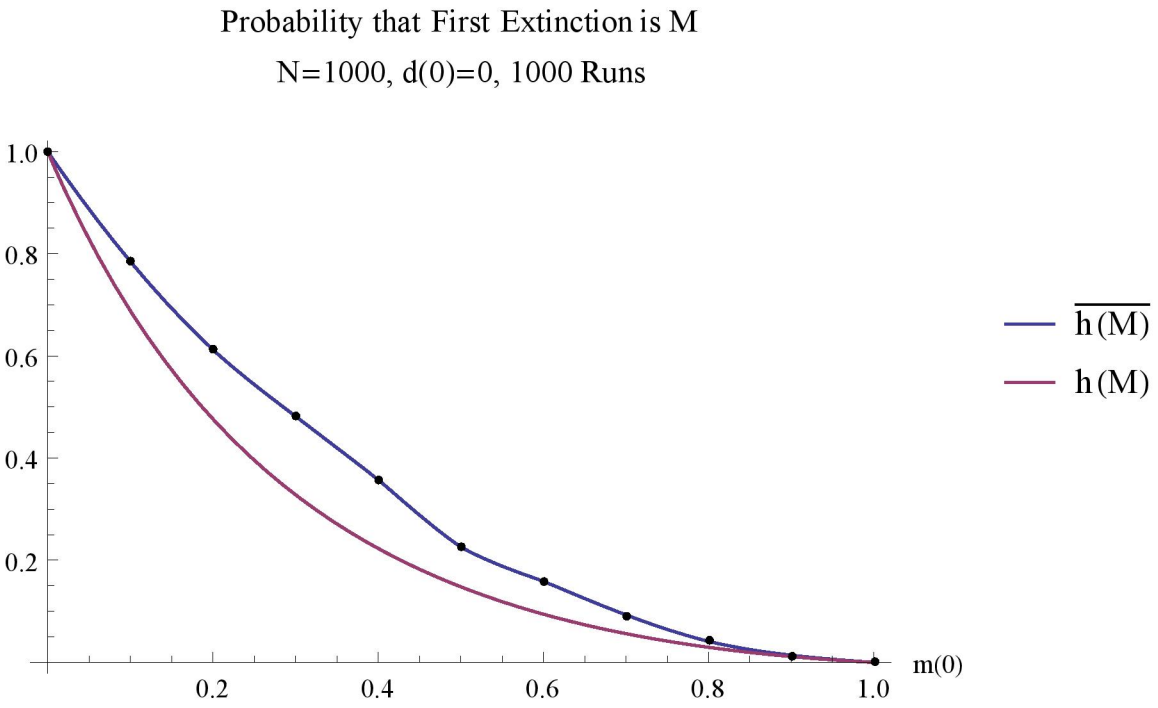

Figure 7.1: Pobability that the first extinction is species M, approximation and simulation

7.3 Expected Time until first extinction over starting value of m

In purple: $\tau * N^2$. In blue: $\overline{\tau^*}$

Since the simulations deal with populations and the model deals with proportions, the approximation has been scaled by multiplying the results by $N^2 = 1000^2$ in the following graph, which is the appropriate scaling from discrete markov chains to diffusions. The approximation has the same shape as the simulations, yet it has significantly higher average times than the simulation. Part of the reason the simulations have a lower average time is probably due to the long tails of the time to first extinction (see section 4.4), which a limited amount of simulations might have a hard time capturing. Other possible sources of difference include the size of N as well as compounded numerical errors to integrate the Green function.

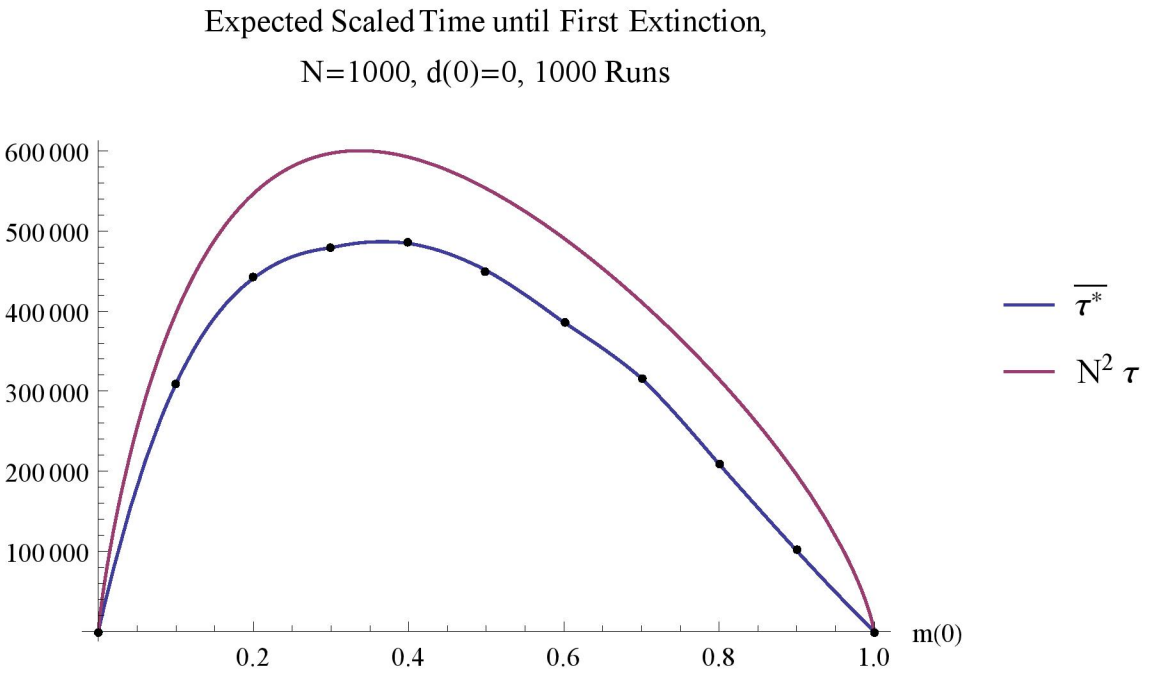

Figure 7.2: Average time until first extinction, approximation and simulation

7.4 Verdict

Based off of the probability of first extinction and the expected time to the first extinction, the approximation seems to conform to the simulation results, and is therefore appropriate to use.

Chapter 8

Discussions

Our work analysed a continuous time Markov chain model to eventually approximate it by a 1 dimensional diffusion. Chapter 3 introduced the continuous time, vector valued Markov process. The probabilities and moments of the simultaneous birth deaths were calculated, and with those both the direction and magnitude of the drift vectors were graphed to better understand the behaviour of the model.

Chapter 4 examined the discrete time Markov chain with the same transition probabilities and expected times to extinction as the continuous version. This discrete time Markov chain was used in simulations with different total populations N and with different initial populations $X(t)$. The simulations mainly measured which species went extinct first and after how much time, as well as if the line $D = 0$ was hit before extinction and the proportion of time spent close to that line.

In Chapter 5 we used the generator of the rescaled Markov process to show that the process approximates a 2-dimensional diffusion for large N. This diffusion degenerated to a deterministic process over most of the population space, with the drift becoming an ordinary differential equation and the variance going to zero. We solved the ODE and showed via the Lyapunov function d^2 that $d = 0$ is an asymptotically stable line of fixed points. Finally we proved that the expected maximum distance between the diffusion and the ODE trajectory

goes to zero as N becomes large.

Chapter 6 centered around proving theorem 6.1.1, which states that the 2-dimensional degenerate diffusion converges to a 1-dimensional diffusion on the asymptotically stable line. To this end, Ito's formula was applied to the ODE projection map to find the 1D diffusion and drift coefficients. Then we used Ito's formula again on the Lyapunov function d^2 to show convergence near the stable line. This allowed us to wrap up our proof of theorem 6.1.1 of the convergence.

Finally Chapter 7 evaluates the 1-dimensional diffusion approximation by comparing some quantities with simulation results. The natural scale was computed numerically from the coefficients (6.13) and (6.14), which allowed us to calculate both the probability that the first extinction is M, as well as the expected time until first extinction. These values, along with their counterparts obtained via simulations in Chapter 3, were graphed for different initial population values on the asymptotically stable line. This allowed for easy comparison between the approximated and simulated values, and showed that at a total population of $N = 1000$ the approximation already seemed to be a good fit on both measures.

It would have been possible to obtain the results of this thesis from a different approach. From the 2-dimensional degenerate diffusion derived at the beginning of Chapter 5, we showed convergence in distribution to the 1-dimensional diffusion along the asymptotically stable line. However, since the 2D diffusion was obtained by approximation via the generator, we only showed that the 1D diffusion approximates the continuous time Markov chain, which of course is less rigorous than a convergence. It remains possible to show that the Markov chain converges in distribution to a diffusion, and a paper by Katzenberger [9] provides a guideline to doing so from a semi-martingale. This convergence would lead to the same result for practical and calculation purposes as the approximation, since the coefficients would be identical.

We used a continuous time Markov chain for our starting model, which lent itself more easily to a diffusion approximation than a discrete time Markov chain. However, the discrete

time version, in which birth-death events happen in regular, discrete intervals rather than as a Poisson process was used for simulations, as has been explained in Chapter 4. The continuous time model could have been programmed for the simulations, but would have been less convenient without providing much in terms of further insights. The discrete time Markov chain could be explored further, and it should be possible to approximate it with a 1-dimensional diffusion with coefficients very similar or identical to those in (6.13) and (6.14). The matrix of one-step transition probabilities of the discrete version is a square matrix $(N+1)(N+2)/2$ to each side. It is mostly composed of zeros, with at most 7 non-zero values per row, however these non-zero probabilities are not simple. Even if it were possible to calculate numerically the limiting probabilities for a given N, it would not provide us with a general rule for large N as our diffusion approximation does.

Our model was presented as having three distinct species with different diets and two seasons. This biological framework is not a restriction; the model can be made to fit many different situations. Instead of three species, the model could instead deal with three distinct phenotypes or genotypes within a single species. In this case, a mutation rate can be added, since it would now be theoretically possible for individuals from one group to mutate to another, and extinction might no longer be permanent. The diets of the population can be replaced with other adaptive traits, as well as survival or mating strategies which depend on the environment. Even the seasons can be generalized to different environmental states.

Though our analysis is confined to our original model, the continuous time Markov chain introduced in Chapter 3, it can serve as a useful guide to better understand other probabilistic population models which feature degenerate diffusions forcing the population onto an asymptotically stable line. Diffusions of interest could be the original models; it is not necessary that they be derived from a Markov chain. And the models need not de inspired by the realm of biology, they might be purely theoretical or arise from the modeling needs of another branch of science.

Another generalization would involve changing the probability of each season occurring.

In our original model, any given birth-death period has an equal probability of occurring during either of the 2 different seasons. However, we could change these probabilities to, say, p for meat season and $(1-p)$ for plant season. Preliminary calculations and simulations show that this produces an asymptotically stable line in the population space along $d = 2p-1$. By using the same approach as in Chapters 5 and 6, it should be possible to approximate this more generalized continuous time Markov chain with a 1-dimensional diffusion. Furthermore, we could subject the model to climate shifts by changing the value p slowly over time, either in a deterministic or probabilistic fashion. If the changes in p are abrupt, the population would suddenly find itself out of equilibrium, and would likely hew closely to an ODE trajectory like in Chapter 5 until it reached the new equilibrium, where it could once again be modeled by a different 1-dimensional diffusion. In a world of increasing climatic shifts and uncertainty, models such as these might serve to better understand what new equilibriums might arise in ecosystems subject to changing environmental pressures.

Chapter 9

Bibliography

[1] Linda J. S. Allen. *An Introduction to Stochastic Processes with Applications to Biology.* Pearson Education, Inc., Upper Saddle River, 2003.

[2] Fred Brauer and Carlos Castillo-Chavez. *Mathematical Models in Population Biology and Epidemiology.* Springer, New York, 2012.

[3] Charles C. Cowden. Game theory, evolutionary stable strategies and the evolution of biological interactions. *Nature Education Knowledge*, 3(10), 2012.

[4] Rick Durrett. Coexistence in stochastic spatial models. *Annals of Applied Probability*, 19 (2):477–496.

[5] Rick Durrett. *Probability Models for DNA Sequence Evolution.* Springer, New York.

[6] Rick Durrett and Lea Popovic. Degenerate diffusions arising from gene duplication models. *The Annals of Applied Probability*, 19(1), 2009.

[7] C. W. Gardiner. *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences.* Springer-Verlag, Berlin.

[8] Geoffrey Grimmett and David Stirzaker. *Probability and Random Processes.* Oxford University Press, Oxford.

[9] G. S. Katzenberger. Solutions of a stochastic differential equation forced onto a manifold by a large drift. *The Annals of Probability*, 19(4), 1991.

[10] Thomas G. Kurtz. *Approximation of Population Processes.* Society for Industrial and Applied Mathematics, Philadelphia.

[11] Thomas G. Kurtz and Philip Protter. Weak limit theorems for stochastic integrals and stochastic differential equations. *The Annals of Probability*, 19(3), 1991.

[12] Alfred J. Lotka. *Elements of Physical Biology.* Williams & Wilkins, Baltimore.

[13] Bernt Oksendal. *Stochastic Differential Equations; An Introduction with Applications.* Springer, Berlin.

[14] Daniel Revuz and Marc Yor. *Continuous Martingales and Brownian Motion.* Springer-Verlag, Berlin.

[15] Sebastian J. Schreiber and Timothy P. Killingback. Spatial heterogeneity promotes coexistence of rock-paper-scissor metacommunities. *Theoretical Population Biology*, 86, 2013.

[16] Adam Shwartz and Alan Weiss. *Large Deviations for Performance Analysis.* Chapman & Hall, London.

[17] Gerald Teschl. *Ordinary Differential Equations and Dynamical Systems.* American Mathematical Society.

[18] Arne Traulsen and Christoph Hauert. Stochastic evolutionary game dynamics. In *Reviews of Nonlinear Dynamics and Complexity, Volume 2.* Wiley.

[19] Vito Volterra. Variazioni e fluttuazioni del numero dindividui in specie animali conviventi. *Memoria della Regia Accademia Nazionale dei Lincei*, 1926.