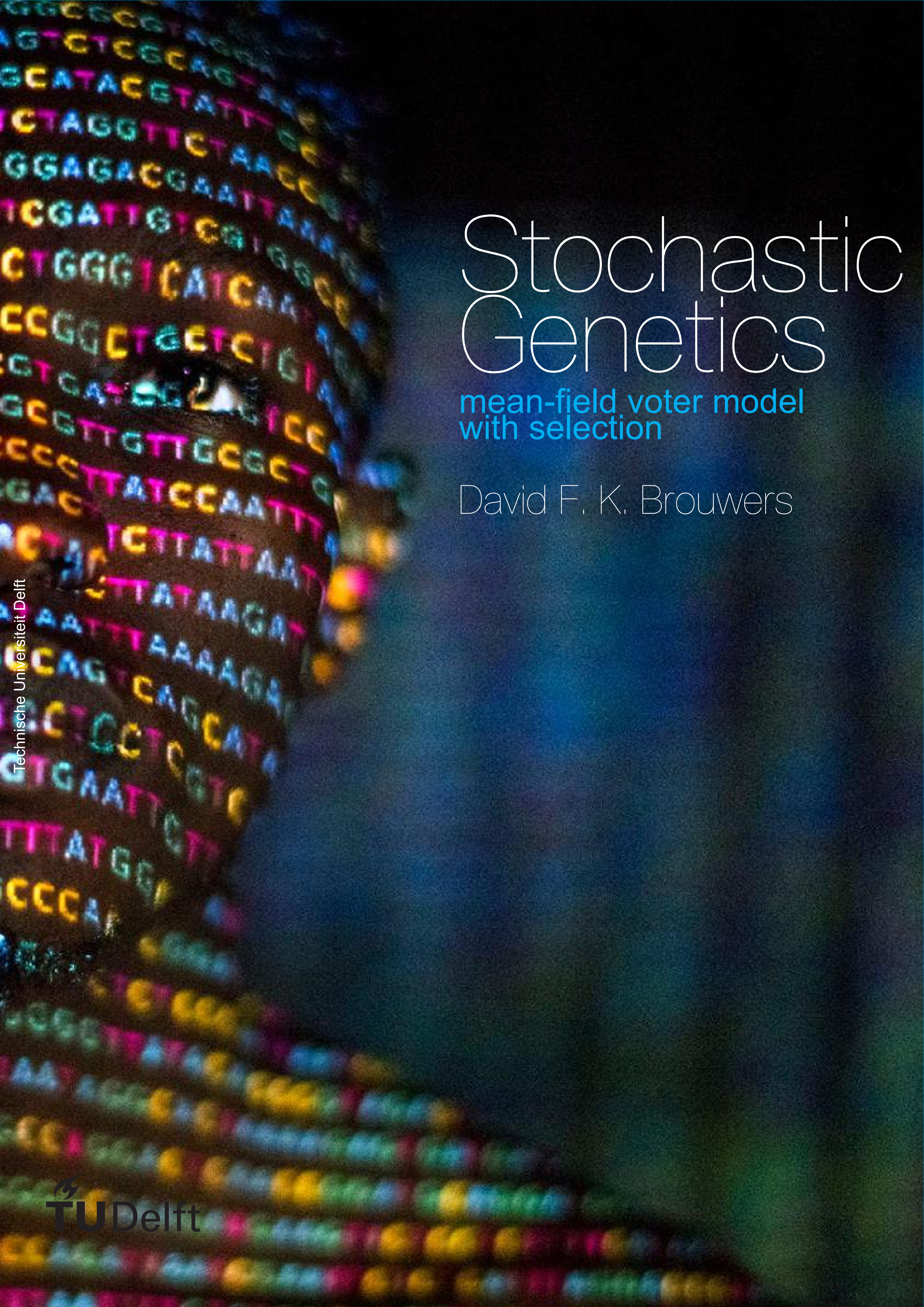

Stochastic Genetics

mean-field voter model with selection

David F. K. Brouwers

Technische Universiteit Delft

TU Delft

Stochastic Genetics

mean-field voter model with selection

by

David F. K. Brouwers

to obtain the degree of Bachelor of Science in Applied Mathematics
at the Delft University of Technology,
to be defended publicly on Tuesday June 28, 2021 at 14:00.

Student number:	4704339	
Project duration:	April 19, 2021 – June 28, 2021	
Thesis committee:	Dr. R. Kraaij,	TU Delft, supervisor
	Dr. C. Kraaikamp	TU Delft, thesis committee
	Dr. ir. R. van der Toorn	TU Delft, thesis committee

Abstract

In this bachelor thesis we use a stochastic model to aspire to explain biodiversity patterns in different ecosystems with selection advantage. The stochastic model we use is an extension of the mean-field voter model [2] where we include a selection factor. In the model individuals with two different types of alleles in two different ecosystems are considered. The model is a stochastic Markov process that describes interactions of individuals with each-other over time. This means that the ratio of individuals with certain alleles stochastically drifts over time. The main goal of this bachelor thesis is investigate whether is it possible that individuals with two different types of alleles can coexist (a stable equilibrium) in two populations. We do this by taking the limit of this Markov process such that we can show convergence to ordinary differential equations. By studying these differential equations we obtain results: vector fields with equilibrium points. We conclude that, under certain conditions and with a selection advantage, coexistence of individuals with two different alleles in two different ecosystems is possible (see Section 4.4) in the form of a stable equilibrium. Furthermore we claim that the typical time to absorption, reaching an absorbing state where all the individuals have the same allele, of the two-dimensional mean-field voter model with selection scales exponentially with the system size.

Contents

1

Introduction

1.1. Introduction to the mean-field voter model

The whole bachelor thesis is based on the mean-field voter model (MFVM) as proposed in [2]. In this Section we give a description of this exact model. Later in this thesis (Chapter 3 and 4) we will make some variations, in the form of selection, on the basic model.

The state of the mean-field voter model is described by a vector of spins $\eta = (\eta_1, \eta_2, \cdots, \eta_N) \in \{0,1\}^N$. Each $\eta_i \in \{0,1\}$ represents, in this case, an individual that can have two different types of alleles. We say that, if the spin of individual i is $\eta_i = 1$ then this individual has allele type A and if the spin of individual i is $\eta_i = 0$, this individual has allele type B.

Next we fix, for each individual, the ecosystem, or population, they are located in. This is done by labeling each individual $\eta_i \in \eta$ with label $h_i \in \{0,1\}$ with $i \in \{0,1,\cdots,N\}$. Each h_i is identically distributed taking values 0 and 1 with probability, respectively, $1-q$ and q with $q \in (0,1)$. When $h_i = 1$ or $h_i = 0$ we say that individual i is in ecosystem 1 or 0 respectively. We assume, without loss of generality, that $q \geq \frac{1}{2}$.

We also have parameter $\rho \in [0,1]$. This parameter represents the chance of adapting the spin for individual $i \in \{1,\cdots,N\}$ with respect to the populations. The population of individual i and of an interacting individual $j \in \{1,\cdots,N\}$ are both considered. The interaction of individuals and the change of spin of each individual goes as follows [2]: each individual i has an independent random waiting time to change its spin. Each individual i, after waiting in accordance with an exponential distribution with mean 1, chooses an individual j with uniform probability. We now have two options:

- if $\eta_j = h_i$ then individual i updates its spin η_i to η_j
- if $\eta_j \neq h_i$ then individual i updates its spin η_i to η_j with probability ρ and keeps its spin with probability $1-\rho$.

We denote the change of spin of individual i to the spin of individual j as $\eta^{j\to i}$. One can see that individual i has a bigger chance to adapt the spin of another individual j that is in the same population h_i. If $\rho = 1$, we obtain the one-dimensional voter model; the fact that each individual i is assigned to population h_i does not play any role. Namely, each individual i updates its spin to the spin of η_j regardless of the fact that $h_i = 0$ or 1. Chapter 3 elaborates on the one-dimensional MFVM. If $\rho < 1$ then the MFVM becomes two-dimensional (See Chapter 4).

One can see the process $\{\eta_N(t)\}_{t\geq 0}$ as a Markov process (see Definition 2.1.0.1 in Section 2.1) with infinitesimal generator, simply generator, (see Definition 2.1.0.5 in Section 2.1):

$$L_N f(\eta) = \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{h_i=\eta_j}[f(\eta^{j\to i}) - f(\eta)] + \rho \mathbb{1}_{h_i\neq\eta_j}[f(\eta^{j\to i}) - f(\eta)]. \tag{1.1}$$

The derivation of the generator above can be found in Section 2.2.

One of the main goals in [2] was to show that the typical time to absorption of the mean-field voter model in two dimensions was different than the typical time to absorption of the mean-field voter model in one dimension. With the term 'time to absorption' we mean the time it takes to reach a state in which all the individuals have the same allele (A or B). For the stochastic proces $\{\eta_N(t)\}_{t\geq 0}$ of the mean-field voter model in one dimension they proved that:

$$T_N = \inf\{t : \{\eta_N(t)\} \in \{0,1\}\} = \mathcal{O}(N). \tag{1.2}$$

This states that the typical time to absorption for the mean-field voter model in one dimension has order $\mathcal{O}(N)$. This means that it takes linear time to reach an absorbing state with system size N. One of the main findings of [2] was that, for the stochastic proces $\{\eta_N(t)\}_{t\geq 0}$ produced by the mean-field voter model in two dimensions, they proved that:

$$T_N = \inf\{t : \{\eta_N(t)\} \in \{0,1\}\} = \mathcal{O}(e^{NI}) \tag{1.3}$$

for certain $I > 0$. This states that the order of the typical time to absorption in the two-dimensional mean-field voter model is $\mathcal{O}(e^{NI})$. Thus the typical time to reach an absorbing state scales exponentially with the population size N. This indicates that, for large populations, individuals with two alleles can coexist for a long time.

1.2. Selection in the mean-field voter model

We want to extend the model, described in Section 1.1, with a selection factor. Without loss of generality we assume that allele A (spin is equal to 1) has a selection advantage. In this thesis we choose fertility selection [4] [1], or just selection, in the form of

$$(1+S) \tag{1.4}$$

with $S > 0$. The selection factor plays a role when an individual adapts the spin of an individual with allele A; therefore it is called fertility selection. We cannot add the selection factor to the generator of the MFVM yet. That is because the generator has to be rewritten such that it is clear in which case the spin of an individual with allele A is adapted. In Section 3.1 and in Section 4.1 the, respectively, one-dimensional and two-dimensional generators are rewritten and the selection factor is added.

1.3. Outline

In Chapter 2 some mathematical background is provided for the reader. Chapter 3 considers the one-dimensional (one population) case and Chapter 4 considers the two-dimensional (two populations) case. Both Chapters 3 and 4 start off with an in depth mathematical derivation of a system of differential equations from the stochastic Markov process of the MFVM including selection. The derivation of starts by reshaping the generators in a preferable form and adding the selection factor (see Section 3.1 and 4.1). Subsequently, the limit of the generators is considered in Section 3.2 and 4.2. Then, the convergence of this limit to differential equations is treated in Section 3.3 and 4.3. Both of the Chapters end with results regarding coexistence of individuals with different alleles.

2

Mathematical Background

Before we take a look at the mean-field voter model with selection itself, it is needful to take a look at some definitions and theorems about Markov processes as well as differential equations. These definitions and theorems help the reader understand the following Chapters more easily.

2.1. Definitions Markov processes

The main property of a Markov process is that it is a random process without memory: its future depends on its present, not on its past.

Definition 2.1.0.1 *A random process whose future probabilities are determined by its most recent values. A stochastic process $x(t)$ is called Markov if for every n and $t_1 < t_2 < \cdots < t_n$ we have [6]:*

$$P(X(t_n) \leq x_n | x(t_{n-1}), \cdots, x(t_1)) = P(X(t_n) \leq x_n | x(t_{n-1})). \tag{2.1}$$

This is equivalent to:

$$P(X(t_n) \leq x_n | x(t) \text{ for all } t \leq t_{n-1}) = P(X(t_n) \leq x_n | x(t_{n-1})). \tag{2.2}$$

Definition 2.1.0.2 *The state space of a Markov process is the set of al possible configurations (states) of the Markov process.*

Let $x(t)_{t \geq 0}$ be a continuous time Markov chain on a state space S. Let $x(0) = i$ be given.

Definition 2.1.0.3 *The transition probability is the probability to go from state i at time $t = 0$ to state j at time t:*

$$P_t(ij) = P(x(t) = j | x(0) = i). \tag{2.3}$$

Definition 2.1.0.4 *A time-homogeneous (continuous) Markov chain on a discrete state space is encoded by its Q-matrix, which describes the transition rate from one state to another [1]:*

$$q_{ij} = \left[\frac{d}{dt} P(x(t) = j | x(0) = i) \right] \Bigg|_{t=0}. \tag{2.4}$$

The infinitesimal generator, or simply generator, of a random process contains information about the process. As we aim to study the convergence of Markov processes that have varying state spaces, we change our point of view to the evolution of expectation values. The infinitesimal generator codes the derivative of this evolution and is defined as follows:

Definition 2.1.0.5 *The infinitesimal generator L of a random process $x(t)$ for certain f is:*

$$(Lf)(i) = \left[\frac{d}{dt} \mathbb{E}[f(x(t) | x(0) = i)] \right] \Bigg|_{t=0}. \tag{2.5}$$

Here we let $f \in C_b^2$.

In the next Section (Section 2.2.1) we relate Q-matrices (see Definition 2.1.0.4) to generators.

2.2. Generator general Markov Process and mean-field voter model

Generators play an important role in this thesis. Section 2.2.1 links the generator of a general Markov process to Q-matrices (see Definition 2.1.0.4). In Section 2.2.2 we obtain the generator of the mean-field voter model as in [2] (see Equation (1.1)) by adjusting the generator of a general Markov process.

2.2.1. Generator of a general Markov process

Let $x(t)$ be a continuous time Markov process with $t \geq 0$ on a countable state space S. We let $f \in C_b^2$ and the Markov process has initial state $x(0) = i$. Using Definition 2.1.0.5 in Section 2.1 we first define the expectation of this process:

$$\mathbb{E}[f(x(t))|x(0) = i] = \sum_{j \in S} P_t(ij)f(j) \tag{2.6}$$

Now we take the derivative and evaluate at $t = 0$ to obtain the following:

$$\frac{d}{dt}\mathbb{E}[f(x(t))|x(0) = i]|_{t=0} = \frac{\mathbb{E}[f(x(t))|x(0) = i] - \mathbb{E}[f(x(0))|x(0) = i]}{t} \tag{2.7}$$

$$= \lim_{t \to 0} \frac{\sum_{j \in S} P_t(ij)f(j)}{t} - \frac{f(i)}{t} \tag{2.8}$$

$$= \lim_{t \to 0} \frac{\sum_{j \in S} P_t(ij)(f(j) - f(i))}{t}. \tag{2.9}$$

The reader can see that if $i = j$, it follows that $f(j) - f(i) = 0$. Furthermore we have that $P_0(ij) = 0$ for all i, j. We continue with Equation (2.9) by subtracting $P_0(ij)$ since it equals 0:

$$= \lim_{t \to 0} \sum_{j \neq i} \frac{(P_t(ij) - P_0(ij))(f(j) - f(i))}{t} \tag{2.10}$$

$$= \sum_{j \neq i} \frac{\delta}{\delta t} P_t(ij)|_{t=0}(f(j) - f(i)) \tag{2.11}$$

$$= \sum_{j \neq i} q_{ij}(f(j) - f(i)) \tag{2.12}$$

$$= \sum_{j=1}^{N} q_{ij}(f(j) - f(i)). \tag{2.13}$$

We see that our derivative in Equation (2.7) can be written as a function of transition rates. Thus our generator of a certain Markov process $x(t)$, on a countable state space S with $|S| = N$ with certain transition rates q_{ij} and initial state i as in Definition 2.1.0.5, can be defined as an operator that acts on a function $f : S \to \mathbb{R}$ where $f \in C_b^2$:

$$(L_N f)(i) = \sum_{j=1}^{N} q_{ij}(f(j) - f(i)). \tag{2.14}$$

2.2.2. Generator of the mean-field voter model without selection

Now we are going to derive the generator as in [2] that applies to our mean-field voter model by adjusting the general generator of a Markov process (see Equation (2.14)). We let $\eta = (\eta_1, \cdots, \eta_N)$ be the initial configuration of spins where each $\eta_j \in \{0, 1\}$ and $j \in \{1, \cdots, N\}$. Let $\eta^{j \to i}$ be the configuration after that individual i changes its spin to that of individual j following the the model described in Section 1.1. We can now see $\{\eta_N(t)\}_{t \geq 0}$ as the evolution of the stochastic process of the change of spins over time. In Equation (2.14) i is the initial condition and can be replaced by initial configuration η; in the same equation state j can be replaced by $\eta^{j \to i}$.

The general Markov generator (2.14) describes alle options to go from initial state i to state j. Here i is fixed and we sum over j. In the generator of the mean-field voter model we still want to describe every combinatorial option to go from initial state η to state $\eta^{j\to i}$. In the case of the MFVM i, j are not states, but indices of individuals. We now still fix the initial state η and work out all options of the adapted configuration $\eta^{j\to i}$. This is done by summing over all individuals i and al individuals j; this is how all the options of the adapted configuration $\eta^{j\to i}$ are considered. The transition rate q_{ij} is now redefined as r_{ij}: the rate of flipping the spin of individual i considering individual j. The general Markov generator is now changed into the generator of the mean-field voter model:

$$(L_N f)(\eta) = \sum_{i=1}^{N}\sum_{j=1}^{N} r_{ij}(f(\eta^{j\to i}) - f(\eta)). \tag{2.15}$$

Now we only have to find the right expression for r_{ij}. In the article of Borile et al. [2] they defined the following: assuming $\rho = 1$, the rate that the spin of individual i flips regardless of one specific j is equal to 1. This can be written as follows:

$$\sum_{j=1}^{N} r_{ij} = 1. \tag{2.16}$$

Now since an individual j is uniformly chosen, as described in Section 1.1, one concludes that $r_{ij} = \frac{1}{N}$. Our generator of the MFVM now becomes:

$$L_N f(\eta) = \sum_{i=1}^{N}\frac{1}{N}\sum_{j=1}^{N}[f(\eta^{j\to i}) - f(\eta)]. \tag{2.17}$$

When $\rho < 1$, the population individuals are located in does matter for the stochastic process; this is the case where two populations are considered. We can distinguish two different cases (see Section 1.1) regarding the population of each individual:

- when $h_i = \eta_j$, individual i adapts the spin of j with probability 1 such that the transition rate becomes
$$r_{ij} = \frac{1}{N}; \tag{2.18}$$
- when $h_i \neq \eta_j$, individual i adapts the spin of j with probability ρ such that the transition rate becomes:
$$r_{ij} = \rho\frac{1}{N}. \tag{2.19}$$

Now we use the two equations above to rewrite the generator of the MFVM as in the article [2]:

$$L_N f(\eta) = \sum_{i=1}^{N}\frac{1}{N}\sum_{j=1}^{N}\mathbb{1}_{h_i=\eta_j}[f(\eta^{j\to i}) - f(\eta)] + \rho\mathbb{1}_{h_i\neq\eta_j}[f(\eta^{j\to i}) - f(\eta)] \tag{2.20}$$

2.3. Differential equations

Differential equations play an important role in this research. In this thesis we eventually want to show convergence of a stochastic process to a solution of a differential equation. We do this by convergence of operators which are, in this case, generators. Therefore we want to link generators to differential equations. Therefore some definitions and theorems are given in this Section.

2.3.1. Definitions differential equations

In section 4.2 we take the limit of the two-dimensional generator of the MVFM with selection. We do this with help of Taylor's Theorem (Theorem 4.2.0.1). Taylors's theorem uses a Hessian matrix.

Definition 2.3.1.1 *A Hessian matrix is a square matrix whose elements are second-order partial derivatives of a given function. If one considers the two-dimensional case we have* $f : \mathbb{R}^2 \to \mathbb{R}$ *is a function* *with continuous second-order partial derivatives the Hessian Matrix is [5]:*

$$\begin{pmatrix} \frac{d^2 f}{dxdx} & \frac{d^2 f}{dxdy} \\ \frac{d^2 f}{dydx} & \frac{d^2 f}{dydy} \end{pmatrix}. \tag{2.21}$$

2.3.2. Convergence of a stochastic process to a differential equation

Definition 2.3.2.1 *A stochastic process $\{m_n(t)\}_{t\geq 0}$ of real valued random-variables is said to converge in distribution, or converge weakly, to $x(t)$, if $\forall t$:*

$$\lim_{N\to\infty} \mathbb{F}_{t,n}(\omega) = \mathbb{F}_t(\omega) \tag{2.22}$$

$\forall \omega \in \mathbb{R}$, at which $\mathbb{F}_t(\omega)$ is continuous. Here $\mathbb{F}_{t,n}$ and $\mathbb{F}_t$ are the cumulative distribution functions of random variables $m_n(t)$ and $x(t)$ respectively, where $m_n(t) \to x(t)$.

Theorem 2.3.2.1 *[3] For a stochastic process $\{m_n(t)\}_{t\geq 0}$ the following are equivalent:*

(i) $\forall t : m_n(t) \to x(t)$ *(weak convergence);*

(ii) $\forall f \in C_b^2, \forall t : \mathbb{E}[f(m_n(t))] \to \mathbb{E}[f(x(t))]$.

Theorem 2.3.2.2 *The following are equivalent:*

(i) $x(t)$ *solves*

$$\begin{cases} \frac{dx}{dt} = F(x) \\ x(0) = x_0. \end{cases} \tag{2.23}$$

(ii) For all $f \in C_b^2$:

$$\begin{cases} \frac{d}{dt} f(x(t)) = \langle \nabla f(x), F(x) \rangle \\ x(0) = x_0. \end{cases} \tag{2.24}$$

When one combines Theorem 2.3.2.1, 2.3.2.2 and a general theorem about convergence of Markov processes (can be found in the book of S. N. Ethier and T. G. Kurtz [3] Section 4.8), we can formulate Theorem 2.3.2.3:

Theorem 2.3.2.3 *Suppose that A_n is the generator of stochastic process $\{m_n(t)\}_{t\geq 0}$ and $Af(x) = \langle \nabla f(x), F(x) \rangle$. Furthermore we suppose that*

1. $m_n(0) \to x_0$ *weakly;*
2. $\lim_{n\to\infty} \sup_x |A_n f(x) - Af(x)| = 0, \forall f \in C_b^2$.

Then $m_n(t)$ converges in distribution to $x(t)$ $\forall t$, where $x(t)$ solves:

$$\begin{cases} \frac{dx}{dt} = F(x) \\ x(0) = x_0. \end{cases} \tag{2.25}$$

3

One-dimensional mean-field voter model with selection

In this Chapter we build up to a one-dimensional differential equation that represents the MFVM with selection. With this differential equation we investigate the coexistence of individuals with two different alleles in the same population where one allele has a selection advantage. In Section 3.1 we rewrite our generator of the MFVM (Section 1.1) such that we can add a selection factor and take a limit. The limit of the generator is worked out in Section 3.2. The convergence of this limit to a differential equation can be found in Section 3.3. Eventually, in Section 3.4, we provide results.

3.1. Generator of the one-dimensional mean-field voter model with selection

In this Section we will rewrite our generator. This is necessary to take a limit (see Section 3.2).

We define the following:

$$x := m_N(\eta) = \frac{1}{N}\sum_{i=1}^{N}(\eta_i). \tag{3.1}$$

We are going to rewrite the generator with the function $f = g \circ m_N$. Here $g \in C_b^2 : [0,1] \to \mathbb{R}$ since $m_N : \eta \to [0,1]$. If we are working with $m_N(\eta)$ we can take a limit $N \to \infty$. $m_N(\eta)$ intuitively is a mean; if we take the mean of the the configuration of spins of the stochastic process, the state space is in $[0,1]$. In this case, the limit cannot go to infinity.
We now are going to show that $\{m_N(\eta)\}_{t\geq 0}$ is indeed Markovian. We fix $m_N(\eta)$ at a certain time t. Let $m_N(\eta^{j\to i})$ be the mean of the vector of spins η after individual i has adapted the spin of individual j at time $t_2 > t$. It is clear that

$$m_N(\eta^{j\to i}) = \begin{cases} m_N(\eta) + \frac{1}{N}, & \text{if } \eta_i = 0, \eta_j = 1; \\ m_N(\eta), & \text{if } \eta_i = \eta_j; \\ m_N(\eta) - \frac{1}{N}, & \text{if } \eta_i = 1, \eta_j = 0. \end{cases} \tag{3.2}$$

As one can see from Equation (3.2) $m_N(\eta^{j\to i})$ only depends on $m_N(\eta)$. We conclude that $m_N(\eta^{j\to i})$ is Markovian. We are going to rewrite the generator as follows with $f = g \circ m_N$ such that we can eventually obtain:

$$L_N f(\eta) = L_N(g \circ m_N)(\eta) = A_N g(m_N(\eta)). \tag{3.3}$$

for some generator A_N.

We start off by taking the original generator of the MFVM (1.1) and set $\rho = 1$ to obtain a one-dimensional model:

$$L_N f(\eta) = \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} [f(\eta^{j\to i}) - f(\eta)]. \tag{3.4}$$

Now we will rewrite the generator (3.4) into two compartments (double sums) such that we can add fertility selection:

- The first sum represents every combinatorial combination such that, after flipping the spin of individual i, the new configuration of all spins $\eta^{j\to i}$ has one more individual with allele A (and one less with B). In this case we add a fertility selection factor $(1+S)$ with $S \geq 0, S \in \mathbb{R}$ (see Section 1.2).
- The second sum represents every combinatorial combination such that, after flipping the spin of individual i, the new configuration of all spins $\eta^{j\to i}$ has one less individual with allele A (and one more with B).

We do not have to consider the case when $\eta_i = \eta_j$ since then the term that is inside the sum in Equation (3.4) is 0. At last we rewrite f as $f(\eta) = g \circ m_N(\eta)$. Writing these steps out we obtain:

$$\begin{aligned} L_N f(\eta) &= \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} [f(\eta^{j\to i}) - f(\eta)] && (3.5) \\ &= (1+S)\Bigg[\sum_{i|\eta_i=0} \frac{1}{N} \sum_{j|\eta_j=1} [f(\eta^{j\to i}) - f(\eta)]\Bigg] + \Bigg[\sum_{i|\eta_i=1} \frac{1}{N} \sum_{j|\eta_j=0} [f(\eta^{j\to i}) - f(\eta)]\Bigg] && (3.6) \\ &= (1+S)\Bigg[\sum_{i|\eta_i=0} \frac{1}{N} \sum_{j|\eta_j=1} [g(m_N(\eta^{j\to i})) - g(m_N)(\eta)]\Bigg] + \Bigg[\sum_{i|\eta_i=1} \frac{1}{N} \sum_{j|\eta_j=0} [g(m_N(\eta^{j\to i})) - g(m_N)(\eta)]\Bigg]. && (3.7) \end{aligned}$$

With Equation (3.2) we rewrite the generator (Equation (3.7)) as follows:

$$\begin{aligned} L_N f(\eta) &= (1+S)R_+(N)\left[g(m_N(\eta) + \frac{1}{N}) - g(m_n(\eta))\right] + R_-(N)\left[g(m_N(\eta) - \frac{1}{N}) - g(m_N(\eta))\right] && (3.8) \\ &= A_N g(m_N(\eta)). && (3.9) \end{aligned}$$

Here A_N is the generator on a random walk on $\{0, \frac{1}{N}, \frac{2}{N}, \cdots, 1\}$ with jump rates $R_+(N)$ and $R_-(N)$ (we define those jump rates a little bit later on). Note that we also take into account the fertility selection $(1+S)$. For our own convenience we substitute $m_N(\eta)$ for x (Equation (3.1)). The generator now looks like this:

$$\begin{aligned} L_N f(\eta) &= (1+S)R_+(N)\left[g(x + \frac{1}{N}) - g(x)\right] + R_-(N)\left[g(x - \frac{1}{N}) - g(x)\right] && (3.10) \\ &= A_N g(x). && (3.11) \end{aligned}$$

The final step in obtaining the final the generator is to calculate the jump rates. We fully calculate the jump rate $R_+(N)$ and give the result of $R_-(N)$ since the calculations are very similar. The jump rate $R_+(N)$ is $\frac{1}{N}$ times every combinatorial option of the pair i, j such that $\eta_i = 0, \eta_j = 1$. This can be written as:

$$R_+(N) = \frac{1}{N}\#\{i,j|\eta_i = 0; \eta_j = 1\} \tag{3.12}$$

This can be written as two independent events: one considering i and the other one considering j:

$$R_+(N) = \frac{1}{N}\#\{i|\eta_i = 0\} \cdot \#\{j|\eta_j = 1\}. \tag{3.13}$$

Our jump rate (3.13) is worked out as follows (using Equation (3.1)):

$$R_+(N) = \frac{1}{N}\sum_{j=1}^{N}\eta_j \sum_{i=1}^{N} 1 - \eta_i = \frac{1}{N}[Nm_N][N(1-m_N)] = Nx(1-x). \tag{3.14}$$

Using the same calculation techniques for $R_-(N)$ we obtain:

$$R_-(N) = Nx(1-x). \tag{3.15}$$

Writing out the jump rates our final version of the one-dimensional generator of the mean-field voter model with selection is:

$$A_N g(x) = (1+S)[Nx(1-x)]\left[g(x+\frac{1}{N}) - g(x)\right] + [Nx(1-x)]\left[g(x-\frac{1}{N}) - g(x)\right]. \tag{3.16}$$

3.2. Limit of the generator of the one-dimensional mean-field voter model with selection

We have found the expression of the generator of the one-dimensional mean-field voter model with selection as in Equation (3.16). Therefore we are now able to take the limit:

$$\begin{aligned}\lim_{N\to\infty} A_N g(x) = {} & \lim_{N\to\infty}(1+S)[Nx(1-x)]\left[g(x+\frac{1}{N}) - g(x)\right] \\ & + \lim_{N\to\infty}[Nx(1-x)]\left[g(x-\frac{1}{N}) - g(x)\right].\end{aligned} \tag{3.17}$$

We need Taylor's theorem to, eventually, obtain the limit of the generator A_N.

Theorem 3.2.0.1 (Taylor's theorem with remainder) *Let $f \in C_b^2[a,b]$. That is, f is a continuous function and its continuous derivative f' is defined on a closed interval $[a,b]$ and for all $x \in (a,b)$ the second derivative $f^{(2)}(x)$ exists: There exists $c \in (a,b)$ such that*

$$f(b) - f(a) = f'(b-a) + R_2 \tag{3.18}$$

where:

$$R_2 = \frac{f^{(2)}(c)}{2!}(b-a)^2. \tag{3.19}$$

Before we are going to use Taylor's theorem we have to check if our function g satisfies the conditions as given above. Note that for any individual $\eta_i \in \eta$ we have that:

$$0 \le \eta_i \le 1 \iff 0 \le \sum_{i=1}^{N}\eta_i \le N \iff 0 \le \frac{1}{N}\sum_{i=1}^{N}\eta_i \le 1. \tag{3.20}$$

So therefore $0 \le m_N(\eta) \le 1$; this functions is continuous on $[0,1]$. Now, since we have chosen, $g \in C_b^2$ it is easy to see that

$$g : [0,1] \to \mathbb{R} \tag{3.21}$$

is also continuous on the defined interval and the first and second derivatives exist. So indeed we are allowed to use Taylor's theorem (see Theorem 3.2.0.1).

We only analyse the first term of the generator (Equation (3.16))

$$(1+S)[Nx(1-x)]\left[g(x+\frac{1}{N}) - g(x)\right] \tag{3.22}$$

since the calculations of the second term are similar; we only give the result of the second term. We start off by analysing the following part of the first term in Equation (3.22) and using Theorem 3.2.0.1:

$$g(x+\frac{1}{N})-g(x)=g'(x)(x+\frac{1}{N}-x)+\frac{g^{(2)}(c)}{2}\left(x+\frac{1}{N}-x\right)^2=\frac{g'(x)}{N}+\frac{g^{(2)}(c)}{2}\left(\frac{1}{N}\right)^2 \tag{3.23}$$

Now we can rewrite the first term as follows following where $c \in \left(x, x+\frac{1}{N}\right)$:

$$(1+S)[Nx(1-x)]\left[g(x+\frac{1}{N})-g(x)\right]= \tag{3.24}$$

$$(1+S)[x(1-x)]N\left[g(x+\frac{1}{N})-g(x)\right]= \tag{3.25}$$

$$(1+S)[x(1-x)]N\left[\frac{g'(x)}{N}+\frac{g^{(2)}(c)}{2}\left(\frac{1}{N}\right)^2\right]= \tag{3.26}$$

$$(1+S)[x(1-x)]\left[g'(x)+\frac{g^{(2)}(c)}{2}\frac{1}{N}\right]. \tag{3.27}$$

Taking the limit we obtain:

$$\lim_{N\to\infty}(1+S)[x(1-x)]\left[g'(x)+\frac{g^{(2)}(c)}{2}\frac{1}{N}\right]= \tag{3.28}$$

$$(1+S)[x(1-x)]\lim_{N\to\infty}\left[g'(x)+\frac{g^{(2)}(c)}{2}\frac{1}{N}\right]=(1+S)[x(1-x)]g'(x). \tag{3.29}$$

The limit of the second term of Equation (3.16) is, using the same steps as the first term, as follows:

$$\lim_{N\to\infty}[Nx(1-x)]\left[g(x-\frac{1}{N})-g(x)\right]=-[x(1-x)]g'(x). \tag{3.30}$$

Combining the two terms above the limit of the generator of the mean-field voter model with selection becomes:

$$\lim_{N\to\infty}A_N g(x)=(1+S)[x(1-x)]g'(x)-[x(1-x)]g'(x)=S[x(1-x)]g'(x). \tag{3.31}$$

We conclude that $Ag(x)=S[x(1-x)]g'(x)$ where $Ag(x)=\lim_{N\to\infty}A_N g(x)$. Thus our limit of the generator in one dimension is

$$Ag(x)=S[x(1-x)]g'(x). \tag{3.32}$$

3.3. The differential equation of the one-dimensional mean-field voter model with selection

Now that we have shown that $\lim_{N\to\infty}A_N g(x)=Ag(x)$ as in (3.32) we want to show convergence of distribution of this generator to a differential equation. We will achieve this by using Theorem 2.3.2.3. We have $Ag(x(t))=\langle g'(x(t)); S[x(1-x)]\rangle$ and $G(x(t))=S[x(1-x)]$. So therefore, using Theorem 2.3.2.3, we will find that $\{m_N(\eta)\}_{t\geq 0}$ converges in distribution to $x(t)$. Here $x(t)$ solves the following system:

$$\begin{cases}\frac{dx}{dt}=S[x(t)(1-x(t))]\\ x(0)=x_0.\end{cases} \tag{3.33}$$

Using our knowledge about differential equations we see that

$$x(t)=\frac{e^{St+C}}{1+e^{St+C}}, \tag{3.34}$$

with C a constant in $\mathbb{R}$, is the general solution of the our system (3.33) which represents the one-dimensional mean-field voter model with selection. Here C can be expressed in terms of x_0 as follows: $c=\ln\left(\frac{x_0}{1-x_0}\right)$.

3.4. Results of the one-dimensional mean-field voter model with selection

In this section we are going to look into the differential equation as in (3.33) and what effects the parameter S has on this system. Before we change the paramater S we are determining the equilibrium points of this system. The equilibrium points are found by setting the derivative equal to 0. Note that still $x(t)$ is the limit of $m_N(\eta)$ which represents the scaled amount of individuals with allele A. Thus we have

$$\frac{dx}{dt} = 0 \iff S[x(t)(1 - x(t))] = 0. \tag{3.35}$$

It is easy to see that $x = 0$ and $x = 1$ are the solutions. These are the absorbing states of the system. This can be explained: when $x = 0$ every individual has allele B so therefore they cannot reproduce an individual with allele A; when $x = 1$ every individual has allele A so similarly it stays this way. Since we have that

$$\frac{dx}{dt} > 0 \tag{3.36}$$

for every initial condition $x_0 \in (1, 0)$ we can conclude that our solution converges to 1. Only for $x_0 = 0$ we have that our solution stays at $x = 0$. This can be explained by the fact that individuals with allele A has a selection advantage over individuals with allele B. Based on Equation (3.34) (Also see Figure 3.1) we are now able to state the following conjecture: For the stochastic process $\{\eta_N(t)\}_{t \geq 0}$ of the one-dimensional mean field voter model

$$T_N = \inf\{t : \{\eta_N(t)\} \in \{0, 1\}, s > 0\} = \mathcal{O}(1). \tag{3.37}$$

In Equation (3.37) we claim that the typical time of absorption T_N of the stochastic process, in one dimension, hitting a state $\{0, 1\}$, a population with only individuals of the same allele, has order $\mathcal{O}(1)$. This differs from the order of hitting an absorbing state in one dimension without selection; this is $\mathcal{O}(N)$ (see Equation (1.2) in Section 1.1). This can be explained by the selection advantage of allele A; the solution always drifts, as long as there are individuals with allele A, to absorbing state $x = 1$ due to the selection factor S.

To obtain a better insight of this differential equation we provide a few graphs of the solution curve in the next subsections.

3.4.1. Solution curve of the one-dimensional mean-field voter model with selection $S = 0.1$

In this section we will show two graphs of the solution of the differential equation (see Equation (3.33)) with both $S = 0.1$ but different initial conditions. We plot the solution (see (3.34)) of the one-dimensional MFVM with selection with $S = 0.1$ and initial condition $x_0 = 0.8$ to obtain Figure 3.1. We see that in Figure 3.1 at $t = 0$ we have $x(0) = 0.8$ which means that 80 percent of the population has allele A. We see that, very quickly, the solution converges to 1. In Figure 3.2 we start with 10 percent of the individuals with allele A but still we converge to an equilibrium point with 100 percent of the individuals with allele A ($x = 1$).

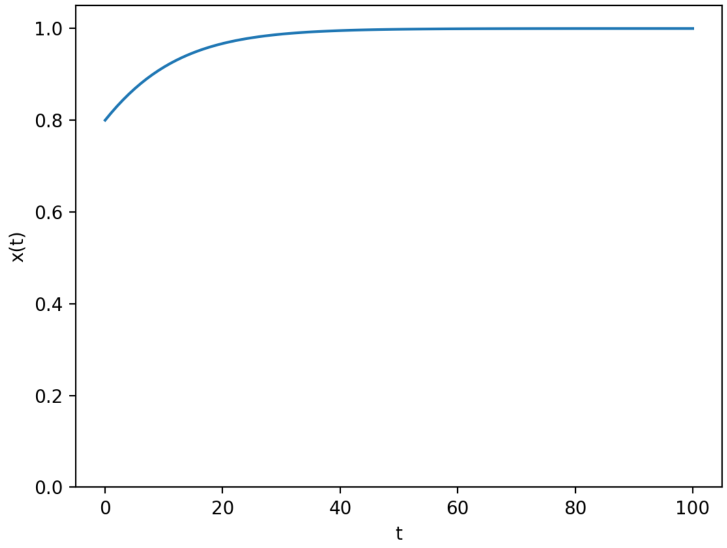

Figure 3.1: Solution curve $x(t)$ (Equation (3.34)) with initial value $x_0 = 0.8$ and $S = 0.1$

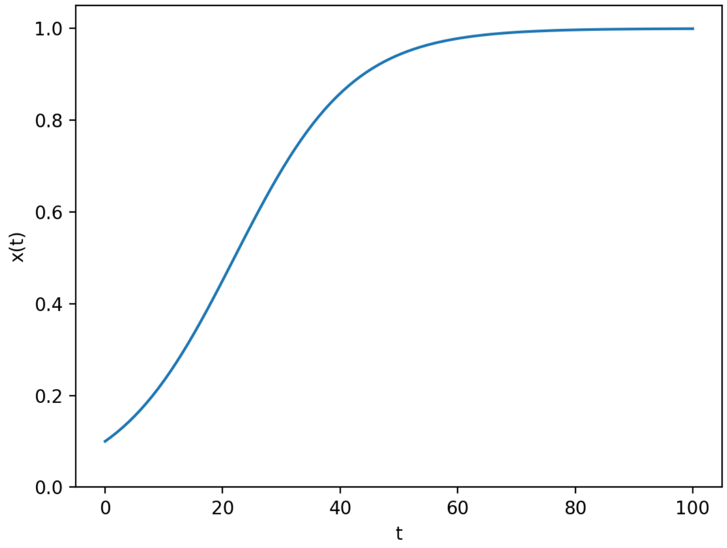

Figure 3.2: Solution curve $x(t)$ (Equation (3.34)) with initial value $x_0 = 0.1$ and $S = 0.1$

3.4.2. Solution curve of the one-dimensional mean-field voter model with selection $S = 0.001$

If the selection takes place with a slower pace it takes longer to reach the absorbing state $x = 1$. We now take $S = 0.001$ and use the same initial conditions as in section 3.4.1: $x_0 = 0.8$ and $x_0 = 0.1$. We obtain the figures 3.3 and 3.4. We see the same behaviour as in section 3.4.1 but only now the time to absorption in Figure 3.3 and Figure 3.4 is longer than Figure 3.1 and 3.2. This is due to the fact that selection factor S is lower. We are also going to provide one more solution curve in Figure 3.5 with an initial condition very close to 0: $x_0 = 0.001$. The selection factor S is still equal to 0.001. In Figure 3.5 we see that, in the beginning, the process goes slow. Then, in the middle, the process is at its fastest. At the end the process slows down again.

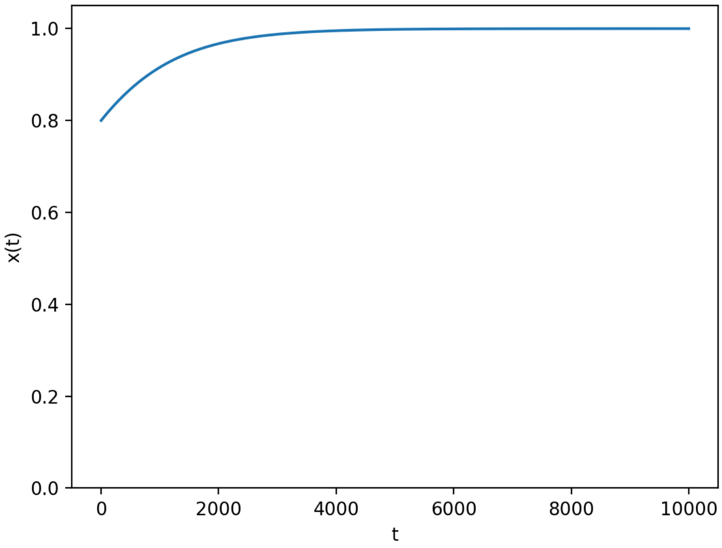

Figure 3.3: Solution curve $x(t)$ (Equation (3.34)) with initial value $x_0 = 0.8$ and $S = 0.001$

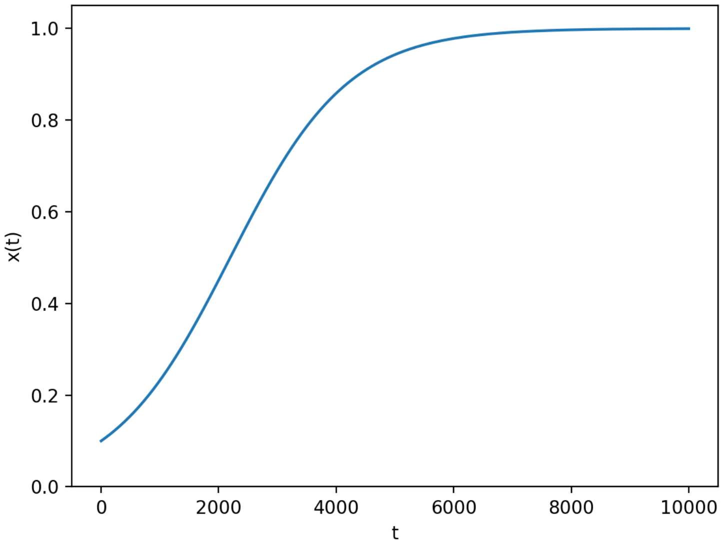

Figure 3.4: Solution curve $x(t)$ (Equation (3.34)) with initial value $x_0 = 0.1$ and $S = 0.001$

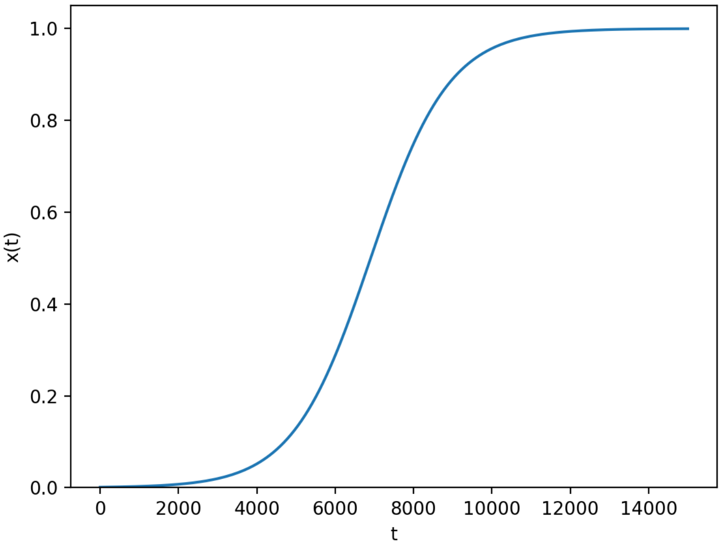

Figure 3.5: Solution curve $x(t)$ (Equation (3.34)) with initial value $x_0 = 0.001$ and $S = 0.001$

4

Two-dimensional mean-field voter model with selection

In this Chapter we build up to a two-dimensional system of differential equations that represents the MFVM with selection. We investigate the coexistence of individuals with two different alleles in two populations where one allele has a selection advantage. We do this the same way as in Chapter 3. First, in Section 4.1, we rewrite our general generator of Section 1.1, considering two populations, such that we can add a selection factor and take a limit. Then the limit of the generator is worked out in Section 4.2. The convergence of this limit to a system of differential equations can be found in Section 4.3. Finally, in Section 4.4, we provide results.

4.1. Generator of two-dimensional mean-field voter model with selection

In this Section a selection factor is added to the MFVM (see 1.1) and the generator of the MFVM, considering two populations, is rewritten. We start of by defining the following:

$$x := m_N^+(\eta) = \frac{1}{N}\sum_{i=1}^{N} \eta_i h_i \tag{4.1}$$

$$y := m_N^-(\eta) = \frac{1}{N}\sum_{i=1}^{N} \eta_i (1 - h_i) \tag{4.2}$$

$$m_N(\eta) = (m_N^+(\eta), m_N^-(\eta)) \tag{4.3}$$

$$q := N^+(\eta) = \frac{1}{N}\sum_{i=1}^{N} h_i \tag{4.4}$$

$$1 - q := N^-(\eta) = \frac{1}{N}\sum_{i=1}^{N} (1 - h_i). \tag{4.5}$$

Here, the first two equations count the number of alleles A in population 1 and 2 respectively. The fourth and fifth equation count the number of individuals in population 1 and 2 respectively. The fact that we are working with averages $m_N^+(\eta)$ and $m_N^-(\eta)$ allows us to take a limit which is done in Section 4.2.

The generator will be rewritten, as in Section 3.1, with the function $f = g \circ m_N$. We let $g \in C_b^2 : [0,1]^2 \to \mathbb{R}$ since $m_N : \eta \to [0,1]^2$. Again, m_N is Markovian; this is proven the same way as we have done in Section 3.1 but now in two dimensions. Using $f = g \circ m_N$, we rewrite our generator (Equation (1.1)) so that we can obtain

$$L_N f(\eta) = L_N(g \circ m_N)(\eta) = A_N g(m_N(\eta)) \tag{4.6}$$

for some generator A_N.

Our original generator of the MFVM, with $\rho < 1$ since we consider the two-dimensional model, is:

$$L_n f(\eta) = \sum_{i=1}^{N} \frac{1}{N} \sum_{j=1}^{N} \mathbb{1}_{h_i=\eta_j}[f(\eta^{j\to i}) - f(\eta)] + \rho \mathbb{1}_{h_i\neq\eta_j}[f(\eta^{j\to i}) - f(\eta)]. \tag{4.7}$$

First we rewrite the generator into four compartments which are double sums: The first compartment represents all combinatorial options of adding an individual with allele A (and losing an individual with allele B) in population 1 where the second compartment represents all combinatorial options of adding an individual with allele A (and losing an individual with allele B) in population 0. The second and fourth do the same respectively but only now subtracting an individual with allele A (and adding an individual with allele B) instead of adding. When an individual with allele A is added to a population we also take into account the fertility selection $S > 0$. This gives the following result:

$$\begin{aligned} L_N f(\eta) = &\sum_{i|\eta_i=0,h_i=1} \frac{1}{N} \sum_{j|\eta_j=1} (1+S)[f(\eta^{j\to i}) - f(\eta)] + \sum_{i|\eta_i=0,h_i=0} \frac{1}{N} \sum_{j|\eta_j=1} \rho(1+S)[f(\eta^{j\to i}) - f(\eta)] + \\ &\sum_{i|\eta_i=1,h_i=1} \frac{1}{N} \sum_{j|\eta_j=0} \rho[f(\eta^{j\to i}) - f(\eta)] + \sum_{i|\eta_i=1,h_i=0} \frac{1}{N} \sum_{j|\eta_j=0} [f(\eta^{j\to i}) - f(\eta)]. \end{aligned} \tag{4.8}$$

The second double sum gets a factor $\rho(1+S)$ since there is an individual with allele A added ($\eta_i = 0, \eta_j = 1$) which gives us the factor $1+S$ and also $h_i \neq \eta_j$ such that we get the factor ρ. With the same arguments the factors of each double sum are determined. For the next step we rewrite our function f as $f = g \circ m_N$ where $g : (0,1)^2 \to \mathbb{R}, g \in C_b^2$. We have that $m_N(\eta) = (m_N^+(\eta), m_N^-(\eta))$ where $m_N : \eta \to (0,1)^2$. We are now able to rewrite the generator in the following way:

$$\begin{aligned} L_N f(\eta) = &\sum_{i|\eta_i=0,h_i=1} \frac{1}{N} \sum_{j|\eta_j=1} (1+S)[g(m_N(\eta^{j\to i})) - g(m_N(\eta))] + \\ &\sum_{i|\eta_i=0,h_i=0} \frac{1}{N} \sum_{j|\eta_j=1} \rho(1+S)[g(m_N(\eta^{j\to i})) - g(m_N(\eta))] + \\ &\sum_{i|\eta_i=1,h_i=1} \frac{1}{N} \sum_{j|\eta_j=0} \rho[g(m_N(\eta^{j\to i})) - g(m_N(\eta))] + \\ &\sum_{i|\eta_i=1,h_i=0} \frac{1}{N} \sum_{j|\eta_j=0} [g(m_N(\eta^{j\to i})) - g(m_N(\eta))]. \end{aligned} \tag{4.9}$$

To help us find generator A_N we investigate, just as in Section 3.1, the following:

$$m_N(\eta^{j\to i}) = \begin{cases} m_N(\eta) + (\frac{1}{N}, 0) & \text{if } \eta_i = 0, \eta_j = 1, h_i = 1, \\ m_N(\eta) + (0, \frac{1}{N}) & \text{if } \eta_i = 0, \eta_j = 1, h_i = 0, \\ m_N(\eta) & \text{if } \eta_i = \eta_j, \\ m_N(\eta) - (\frac{1}{N}, 0) & \text{if } \eta_i = 1, \eta_j = 0, h_i = 1, \\ m_N(\eta) - (0, \frac{1}{N}) & \text{if } \eta_i = 1, \eta_j = 0, h_i = 0. \end{cases} \tag{4.10}$$

With the equations above we can now rewrite the generator (4.9) further:

$$
\begin{aligned}
L_N f(\eta) = R_+(N)^{h+}(1+S)&\left[g(m_N^+(\eta)+\frac{1}{N}, m_N^-(\eta)) - g(m_N^+(\eta), m_N^-(\eta))\right] + \\
R_+(N)^{h-}\rho(1+S)&\left[g(m_N^+(\eta), m_N^-(\eta)+\frac{1}{N}) - g(m_N^+(\eta), m_N^-(\eta))\right] + \\
R_-(N)^{h+}\rho&\left[g(m_N^+(\eta)-\frac{1}{N}, m_N^-(\eta)) - g(m_N^+(\eta), m_N^-(\eta))\right] + \\
R_-(N)^{h-}&\left[g(m_N^+(\eta), m_N^-(\eta)-\frac{1}{N}) - g(m_N^+(\eta), m_N^-(\eta))\right] \\
&= A_N g(m_N)(\eta).
\end{aligned} \tag{4.11}
$$

Here A_N is the generator that describes a two-dimensional random walk on $\{0, \frac{1}{N}, \cdots, 1\}^2$ with jump rates: $R_+(N)^{h+}, R_+(N)^{h-}, R_-(N)^{h+}, R_-(N)^{h-}$ and additional factors regarding fertility selection $(1+S)$ and the population ρ. For our own convenience we use x, y as we defined in Equation (4.1) and Equation (4.2). Substituting this x, y notation the generator looks like this:

$$
\begin{aligned}
A_N g(x,y)(\eta) = R_+(N)^{h+}(1+S)&\left[g(x+\frac{1}{N}, y) - g(x,y)\right] + \\
R_+(N)^{h-}\rho(1+S)&\left[g(x, y+\frac{1}{N}) - g(x,y)\right] + \\
R_-(N)^{h+}\rho&\left[g(x-\frac{1}{N}, y) - g(x,y)\right] + \\
R_-(N)^{h-}&\left[g(x, y-\frac{1}{N}) - g(x,y)\right].
\end{aligned} \tag{4.12}
$$

The last step in determining the two-dimensional generator of the mean-field voter model with selection is to calculate the jump rates as mentioned above. We are going to fully work out the first jump rate $R_+(N)^{h+}$ as the calculations for the other three jump rates are very similar and therefore not needful. We only provide the results of the three other jump rates at the end of this Section. The jump rate $R_+(N)^{h+}$ is $\frac{1}{N}$ times every combinatorial option of the pair i, j such that $\eta_i = 0, h_i = 1; \eta_j = 1$. Written out mathematically we obtain:

$$R_+(N)^{h+} = \frac{1}{N}\#\{i, j | \eta_i = 0, h_i = 1; \eta_j = 1\}. \tag{4.13}$$

Note that in the upcoming calculation of the jump rates we are using the 4 Equations defined in the beginning of this Section: (4.1), (4.2), (4.4) and (4.5). We first calculate condition of (4.13) depending on j:

$$\#\{j | \eta_j = 1\} = \sum_{j=1}^{N} \eta_j = \sum_{j=1}^{N} \eta_j (h_j + (1-h_j)) = N(m_N^+ + m_N^-) = N(x+y). \tag{4.14}$$

Now we calculate the other condition of (4.13) depending on i:

$$\#\{i | \eta_i = 0, h_i = 1\} = \sum_{i=1}^{N} h_i(1-\eta_i) = \sum_{i=1}^{N} h_i - \sum_{i=1}^{N} h_i \eta_i = N(N^+) - N(m_N^+) = Nq - Nx = N(q-x). \tag{4.15}$$

Multiplicating the two events above, since they are independent, the jump rate $R_+(N)^{h+}$ becomes:

$$R_+(N)^{h+} = \frac{1}{N} N(x+y) N(q-x) = N(q-x)(x+y). \tag{4.16}$$

Using the same way of calculating the first jump rate, the other three jump rates are as follows:

$$R_+(N)^{h-} = N(1-q+y)(x+y) \tag{4.17}$$

$$R_-(N)^{h+} = Nx(1-(x+y)) \tag{4.18}$$

$$R_-(N)^{h-} = Ny(1-(x+y)). \tag{4.19}$$

Our final expression of the generator of the two-dimensional MFVM with selection is:

$$\begin{aligned} A_N g(x,y)(\eta) = (1+S)N(q-x)(x+y)\left[g(x+\frac{1}{N},y)-g(x,y)\right]+ \\ \rho(1+S)N(1-q+y)(x+y)\left[g(x,y+\frac{1}{N})-g(x,y)\right]+ \\ \rho Nx(1-(x+y))\left[g(x-\frac{1}{N},y)-g(x,y)\right]+ \\ Ny(1-(x+y))\left[g(x,y-\frac{1}{N})-g(x,y)\right]. \end{aligned} \tag{4.20}$$

4.2. Limit of the generator of the two-dimensional mean-field voter model with selection

Now that we have found the generator of two-dimensional MFVM with selection we want to obtain the limit of this mathematical object just as in Section 3.2. We want to find $Ag(x,y)$:

$$\lim_{N\to\infty} A_N g(x,y) = Ag(x,y). \tag{4.21}$$

with $A_N g(x,y)$ as in Equation (4.20). Again, using the same tools as in Section 3.2, we need Taylor's theorem. Only now we are working with two dimensions:

Theorem 4.2.0.1 (Taylor's theorem in two dimensions) *Suppose that $S \subset \mathbb{R}^n$ is an open set and that $f : S \to \mathbb{R}$ is a function of class C^2 on S. Then, for $\mathbf{a}$ and $\mathbf{b} \in S$, such that the line segment connecting $\mathbf{a}$ and $\mathbf{b}$ is contained in S, there exists $\mathbf{c} \in (\mathbf{a}, \mathbf{b})$ such that:*

$$f(\mathbf{b}) = f(\mathbf{a}) + \nabla f(\mathbf{b}-\mathbf{a}) + \frac{1}{2}(\mathbf{b}-\mathbf{a})^T H(\mathbf{c})(\mathbf{b}-\mathbf{a}). \tag{4.22}$$

Here H is the Hessian matrix as in Definition 2.3.1.1.

We are going to compute, with the use of the two-dimensional theorem of Taylor, the first term of the generator A_N as in Equation (4.20):

$$(1+S)N(q-x)(x+y)\left[g(x+\frac{1}{N},y)-g(x,y)\right]. \tag{4.23}$$

The other three terms are computed similarly and therefore we only provide the results later in this Section. We apply Theorem 4.2.0.1 to $\mathbf{a} = (x,y), \mathbf{b} = \left(x+\frac{1}{N}, y\right)$ and $c = (c_1, c_2)$. We observe the following:

$$g(x+\frac{1}{N},y) = g(x,y) + \nabla g(x,y)\begin{pmatrix} x+\frac{1}{N}-x \\ y-y \end{pmatrix} + \begin{pmatrix} x+\frac{1}{N}-x & y-y \end{pmatrix} H(g(\mathbf{c}))\begin{pmatrix} x+\frac{1}{N}-x \\ y-y \end{pmatrix} \tag{4.24}$$

$$= g(x,y) + \nabla g(x,y)\begin{pmatrix} \frac{1}{N} \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{N} & 0 \end{pmatrix} H(g(\mathbf{c}))\begin{pmatrix} \frac{1}{N} \\ 0 \end{pmatrix} \tag{4.25}$$

$$= g(x,y) + \begin{pmatrix} d_x g(x,y) & d_y g(x,y) \end{pmatrix}\begin{pmatrix} \frac{1}{N} \\ 0 \end{pmatrix} + \begin{pmatrix} \frac{1}{N} & 0 \end{pmatrix}\begin{pmatrix} d_{xx}g(\mathbf{c}) & d_{xy}g(\mathbf{c}) \\ d_{yx}g(\mathbf{c}) & d_{yy}g(\mathbf{c}) \end{pmatrix}\begin{pmatrix} \frac{1}{N} \\ 0 \end{pmatrix} \tag{4.26}$$

$$= g(x,y) + \frac{1}{N}d_x g(x,y) + \left(\frac{1}{N}\right)^2 d_{xx}g(\mathbf{c}). \tag{4.27}$$

Shifting $g(x,y)$ to the left side of the equation we obtain:

$$g(x+\frac{1}{N},y) - g(x,y) = \frac{1}{N}d_x g(x,y) + \left(\frac{1}{N}\right)^2 d_{xx}g(\mathbf{c}). \tag{4.28}$$

Now we are able to take the limit of the first term (4.23) (S is not dependent on N):

$$\lim_{N\to\infty}(1+S)N(q-x)(x+y)\left[g(x+\frac{1}{N},y)-g(x,y)\right]= \tag{4.29}$$

$$\lim_{N\to\infty}(1+S)N(q-x)(x+y)\left[\frac{1}{N}d_x g(x,y)+\left(\frac{1}{N}\right)^2 d_{xx}g(\mathbf{c})\right]= \tag{4.30}$$

$$\lim_{N\to\infty}(1+S)(q-x)(x+y)\left[d_x g(x,y)+\left(\frac{1}{N}\right)d_{xx}g(\mathbf{c})\right]= \tag{4.31}$$

$$(1+S)(q-x)(x+y)d_x g(x,y). \tag{4.32}$$

For the three other terms of the generator as in (4.20) we have the following results:

$$\lim_{N\to\infty}\rho(1+S)N(1-q+y)(x+y)\left[g(x,y+\frac{1}{N})-g(x,y)\right]=\rho(1+S)(1-q-y)(x+y)d_y g(x,y) \tag{4.33}$$

$$\lim_{N\to\infty}\rho N x(1-(x+y))\left[g(x-\frac{1}{N},y)-g(x,y)\right]=\rho x(1-(x+y))\cdot -d_x g(x,y) \tag{4.34}$$

$$\lim_{N\to\infty}N y(1-(x+y))\left[g(x,y-\frac{1}{N})-g(x,y)\right]=y(1-(x+y)\cdot -d_y g(x,y). \tag{4.35}$$

Combining the the limits of the different terms of the generator, the limit of the generator itself becomes:

$$\begin{aligned}\lim_{N\to\infty}A_N g(x,y) = {} & (1+S)(q-x)(x+y)d_x g(x,y)+\\ & \rho(1+S)(1-q-y)(x+y)d_y g(x,y)+\\ & \rho x(1-(x+y))\cdot -d_x g(x,y)+\\ & y(1-(x+y))\cdot -d_y g(x,y)\\ & = Ag(x,y).\end{aligned} \tag{4.36}$$

Using some simple algebra we rewrite the generator above:

$$Ag(x,y)=V^+(x,y)\frac{dg(x,y)}{dx}+V^-(x,y)\frac{dg(x,y)}{dy}. \tag{4.37}$$

Where $V^+(x,y)$ and $V^-(x,y)$ are defined as follows:

$$V^+(x,y)=[-\rho x(1-x-y)+(1+S)(q-x)(x+y)] \tag{4.38}$$

$$V^-(x,y)=[-y(1-x-y)+\rho(1+S)(1-q-y)(x+y)]. \tag{4.39}$$

The limit of the two-dimensional generator of the MFVM with selection as in Equation (4.37) is our final expression.

4.3. The system of differential equations of the two-dimensional mean-field voter model with selection

In Section 4.2 we have achieved the desired result: $\lim_{N\to\infty}A_N g(x,y)=Ag(x,y)$ with $Ag(x,y)$ as in Equation (4.37). This Section shows how to show convergence of this generator to a system of differential equations. Just as in Section 3.3 we will use Theorem 2.3.2.3 (convergence of generators to differential equations, see Section 2.3) to obtain our system of differential equations.

Note that still $(x,y):=(m_N^+(\eta),m_N^-(\eta))$ is just notation. Since our generator

$$Ag(x(t),y(t))=\langle\nabla g(x(t),y(t)),(V^+\quad V^-)\rangle \tag{4.40}$$

and $G(x(t), y(t)) = (V^+, V^-)$ we can see that, using Theorem 2.3.2.3, our stochastic process $\{m_N^+(\eta), m_N^-(\eta)\}_{t\geq 0}$ converges in distribution to the solution $(x(t), y(t))$ of the system of differential equations:

$$\begin{cases} \frac{d}{dt}x = V^+(x,y) \\ \frac{d}{dt}y = V^-(x,y) \\ (x(0), y(0)) = (x_0, y_0). \end{cases} \tag{4.41}$$

Where V^+ and V^- are (also see Section 4.2):

$$V^+(x,y) = [-\rho x(1-x-y) + (1+S)(q-x)(x+y)] \tag{4.42}$$

$$V^-(x,y) = [-y(1-x-y) + \rho(1+S)(1-q-y)(x+y)]\,. \tag{4.43}$$

4.4. Results of the two-dimensional mean-field voter model with selection

To obtain insight in the system of differential equations we are going to look at vector field plots. This will show how the solution will behave and if there are any equilibrium points. Eventually we want to find out if it is possible that the two different alleles (A and B) can coexist in two different populations if there is selection.

For the whole Section 4.4 we have that the vector fields are displayed with a sequential colourmap: the colour yellow represents strong vectors and the colour black represents weak vectors; in-between we have the colour red.

Note that x, y still represent $m_N^+(\eta), m_N^-(\eta)$ respectively which are the scaled number of alleles A in population 1 and 0 respectively. We have that:

$$0 \leq \eta_i \leq 1 \iff 0 \leq \sum_{i=1}^{N} \eta_i \leq N \iff 0 \leq \frac{1}{N}\sum_{i=1}^{N} \eta_i \leq 1 \iff 0 \leq x \leq 1. \tag{4.44}$$

$$0 \leq 1-\eta_i \leq 1 \iff 0 \leq \sum_{i=1}^{N} 1-\eta_i \leq N \iff 0 \leq \frac{1}{N}\sum_{i=1}^{N} 1-\eta_i \leq 1 \iff 0 \leq y \leq 1. \tag{4.45}$$

We also have the fact that $x + y$ can never be greater than 1 since, in that case, there would be more individuals with allele A than there are individuals. This is, of course, impossible. Therefore we can conclude that our stochastic process lies in the plane $\{(a,b) \in [0,1]^2 \,:\, a + b \leq 1\}$. However we can condition the boundaries of the plane even more because of the population for each individual depending on q; as $N \to \infty$ we have that, by the law of large numbers [2]:

$$x = \frac{1}{N}\sum_{i=1}^{N} \eta_i h_i \leq \frac{1}{N}\sum_{i=1}^{N} h_i \to q \tag{4.46}$$

$$y = \frac{1}{N}\sum_{i=1}^{N} \eta_i(1-h_i) \leq \frac{1}{N}\sum_{i=1}^{N} 1-h_i \to 1-q. \tag{4.47}$$

Therefore the limit points of the sequence (x, y) belong to the plane$[0,q] \times [0,1-q]$.

4.4.1. Vector field of the two-dimensional mean-field voter model with $S = 0$, $\rho = 1$

When start of considering the the most basic form; we do not consider any form of selection ($S = 0$) and we let $\rho = 1$ such that it does not matter for the stochastic process in which population an individual is. This is the one-dimensional case without selection. When $\rho = 1$ it takes little work to show that

$$\frac{dx}{dt} + \frac{dy}{dt} = 0 \tag{4.48}$$

This means there is no movement over the diagonal of the vector field (see Figure 4.1). It is the consequence of the fact that, without selection and with $\rho = 1$, we are considering a symmetric random walk $W_N := N(x+y)$ [2]. In this case any initial condition is attracted along the line (coloured in black) $m := x+y$ to the fixed point manifold $x = y$. One can interpret this as that the solution always converges to a point where there are as many individuals with allele A in population 1 as there are in population 0. If you now look at sum of the two populations $x+y$ as one population, since there is no difference in both, you get back the one-dimensional case. This can be seen in the vector field in Figure 4.1 where, without loss of generality, we choose $q = \frac{1}{2}$. We can see that how closer we are to to equilibrium solution, the slower we progress towards this solution since the vectors are black (see beginning of this Section 4.4).

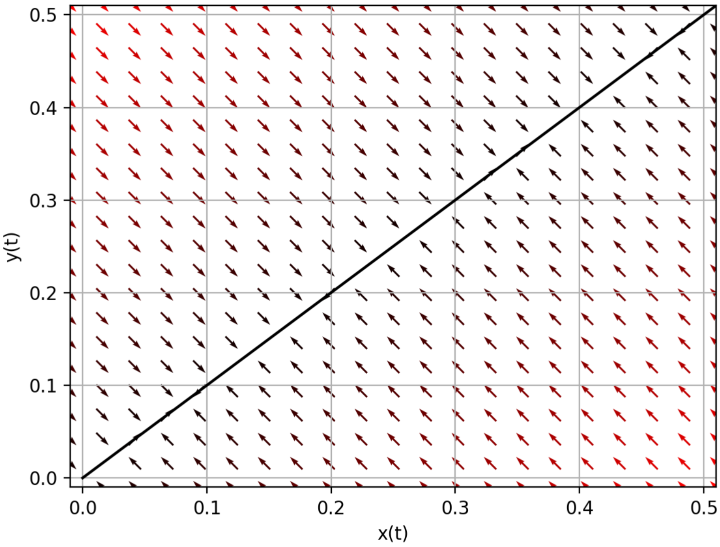

Figure 4.1: Vector field two-dimensional mean-field voter model $S = 0, \rho = 1, q = \frac{1}{2}$. The black line is the equilibrium solution in the form of a line $m := x + y$. The colour black represents weak vectors.

4.4.2. Vector field of the two-dimensional mean-field voter model with $S = 0$, $\rho < 1$

In this Section we will shortly evaluate the main results found by [2] in the case that $\rho < 1$. In Section 4.4.3 we will add selection to our model to obtain new results. However it is very useful to go over the results in the case $S = 0$ and provide some clarification.

If $\rho < 0, S = 0$ the system (4.41) has the following equilibrium points:

1. $(x, y) = (q, 1-q)$ which represents a limiting behaviour where all the spins are equal to 1;
2. $(x, y) = (0, 0)$ which is the case with all spins equal to 0;
3. $(x, y) = \left(\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}, \rho\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}\right)$.

We omit the algebraic calculations of these equilibrium points since they are very long and tedious. One can see that equilibrium three lies inside $[0, q] \times [0, 1-q]$ if and only if the condition

$$\rho < \frac{1-q}{q} \tag{4.49}$$

holds by solving the inequality $\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)} < q$ or the inequality $\rho\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)} < 1-q$. When this condition holds, the solution is admissible. We will now provide an admissible vector field (see Figure 4.2) without selection but with dependency of the population of each individual; a vector field with $\rho < 1, S = 0$

where condition (4.49) holds. We set $q = \frac{1}{2}$ as in Section 4.4.1 and choose $\rho = \frac{1}{2}$. The three equilibria

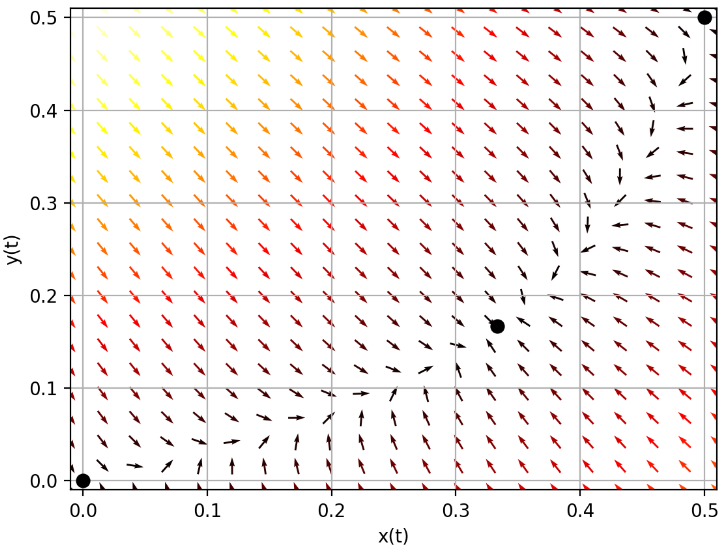

Figure 4.2: Vector field two-dimensional mean-field voter model $S = 0, \rho < \frac{1-q}{q} < 1, q = \frac{1}{2}, \rho = \frac{1}{2}$. Here the equilibrium points $(q, 1-q)$, $(0,0)$ and $\left(\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}, \rho\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}\right)$ are displayed as black dots. The colour black represents weak vectors.

as noted in the beginning of this Section are displayed as black dots. We see that the equilibria $(0,0)$ and $(q, 1-q)$ are unstable equilibria. If the initial conditions differ from those two equilibria the solution will converge to the third equilibrium point $(x,y) = \left(\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}, \rho\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}\right)$. One can see that the convergence of the solution to the equilibrium point $(x,y) = \left(\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}, \rho\frac{q(1+\rho)-\rho}{(1+\rho)(1-\rho)}\right)$ gets slower when the solution gets closer to this point.

The third equilibria is stable and shows the two alleles A and B can coexist in both populations. In the article of Borile et al. [2] they eventually prove that

$$T_N = \inf\{t : \{\eta_N(t)\} \in \{0,1\}\} = \mathcal{O}(e^{NI}). \tag{4.50}$$

with $I > 0$. This proof is out of our scope for this bachelor thesis; the interested reader can read the article of Borile et al. [2]. The result of the proof, shown in Equation (4.50), tells us that the typical time to absorption has order $\mathcal{O}(e^{NI})$. With time to absorption T_N we mean reaching an absorbing state $(q, 1-q)$ or $(0,0)$ which represents that all individuals in both populations have only allele A or only allele B respectively (see Section 1.1). One can conclude that, with N large, it takes a very long time (exponentially growing with N) before all the individuals in both populations have the same allele; two different alleles can coexist.

4.4.3. Vector field of the two-dimensional mean-field voter model with $S > 0$, $\rho < 1$

Now that we have considered the general cases in the previous Section we are now adding selection by letting $S > 0$. Also we have that $\rho < 1$. Since, in this Section, the main results of this thesis are obtained, we recall the system of differential equations of the MFVM with selection

$$\begin{cases} \frac{d}{dt}x = -\rho x(1-x-y) + (1+S)(q-x)(x+y) \\ \frac{d}{dt}y = -y(1-x-y) + \rho(1+S)(1-q-y)(x+y) \end{cases} \tag{4.51}$$

as in Equation (4.41)). The equilibrium points of this system, with $S > 0, \rho < 1$ are:

1. $(x, y) = (q, 1-q)$ which represents a limiting behaviour where all the spins are equal to 1;
2. $(x, y) = (0, 0)$ which is the case with all spins equal to 0;
3. $(x_3, y_3) := (x, y) =$

$$\left(\frac{\rho - (1+S)\rho^2 - (1+S)q + (1+S)\rho^2 q}{(\rho^2-1)((1+S)-\rho)}, \frac{-\rho(\rho - (1+S)\rho^2 - (1+S)q + (1+S)\rho^2 q)}{(\rho^2-1)((1+S)\rho-1)} \right). \quad (4.52)$$

Again, as in Section 4.4.2, the algebraic calculations are excluded since they are not of particular interest. For the third equilibrium point (x_3, y_3) to be located in the admissible square $[0, q] \times [0, 1-q]$ we need to solve

$$x_3 < q \iff \left(\frac{\rho - (1+S)\rho^2 - (1+S)q + (1+S)\rho^2 q}{(\rho^2-1)((1+S)-\rho)} \right) < q \quad (4.53)$$

or

$$y_3 < 1-q \iff \left(\frac{-\rho(\rho - (1+S)\rho^2 - (1+S)q + (1+S)\rho^2 q)}{(\rho^2-1)((1+S)\rho-1)} \right) < 1-q. \quad (4.54)$$

Both equations result in the same condition when solved where $(1+S) > 1, 0 < \rho < 1$ and $\frac{1}{2} < q < 1$:

$$\rho < \frac{(1+S) - q\sqrt{\frac{(1+S)^2+4q^2-4q}{q^2}}}{2q}. \quad (4.55)$$

For our own convenience we define

$$h(S, q) := \frac{(1+S) - q\sqrt{\frac{(1+S)^2+4q^2-4q}{q^2}}}{2q}. \quad (4.56)$$

When the condition in Equation (4.55) holds, the third equilibrium (x_3, y_3) is admissible. Furthermore one can see that if $S = 0$ we obtain the condition as in Equation (4.49) without selection after simplifying the expression in Equation (4.55).

In Figure 4.3 a vector field of the two-dimensional MFVM with selection (see Equation (4.51)), under the mentioned condition, is given. Without loss of generality we set $\rho = \frac{1}{2}, q = \frac{1}{2}$ and $S = 0.1$. With these parameters the condition holds. Now $(0, 0)$ and $(q, 1-q)$ are unstable equilibrium points; we see that the equilibrium point (x_3, y_3) is a stable equilibrium.
When the condition does not hold, the third equilibrium (x_3, y_3) is not admissible. Then we only have two equilibrium points in the vector field; in Figure 4.4 we set $S = 5$ such that the condition does not hold. We keep the rest of the parameters the same as in Figure 4.3. One observes that the equilibrium point $(q, 1-q)$ is now the only stable equilibrium point and all initial conditions are attracted to it.

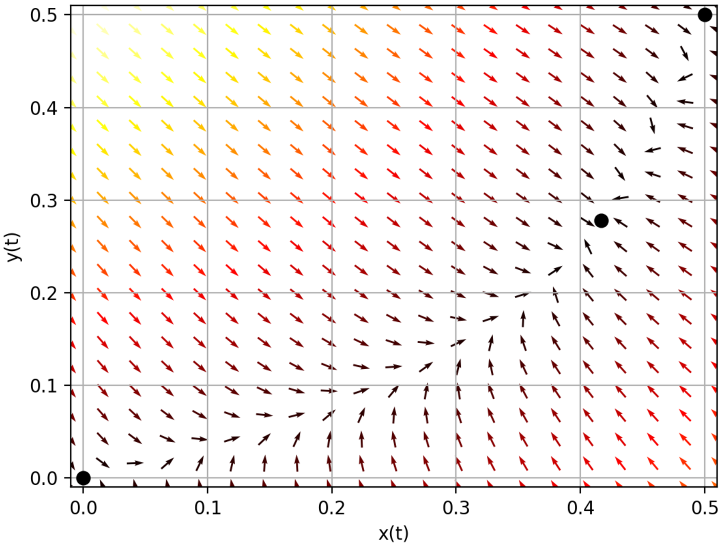

Figure 4.3: Vector field two-dimensional mean-field voter model $S = 0.1, \rho < \frac{1-q}{q} < 1, q = \frac{1}{2}, \rho = \frac{1}{2}$. Here the equilibrium points $(q, 1-q)$, $(0, 0)$ and (x_3, y_3) (see Equation (4.52)) are displayed as black dots. The colour black represents weak vectors.

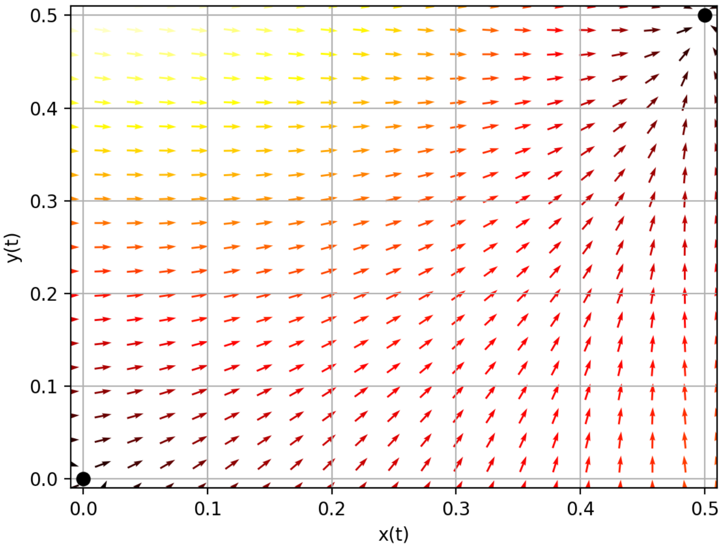

Figure 4.4: Vector field two-dimensional mean-field voter model $S = 5, \rho < \frac{1-q}{q} < 1, q = \frac{1}{2}, \rho = \frac{1}{2}$. Here the equilibrium points $(q, 1-q)$, $(0, 0)$ are displayed as black dots. The colour black represents weak vectors.

We want find out for which values of the parameters ρ, S and q the third equilibrium (x_3, y_3) is admissible. Therefore we examine the surface plot of $h(S, q)$ (Equation (4.56)) in Figure 4.5. In Figure 4.5 we see that the value of $h(S, q)$ is small for S large; when S is small, we see that the value of $h(S, q)$ is small for large values of q. When the value of $h(S, q)$ is small, ρ has to be even smaller to obtain an admissible third stable equilibrium. Therefore, regarding the paramaters and the existence of the third stable equilibrium (x_3, y_3) with selection, we can state the following:

1. smaller ρ with $0 < \rho < 1$ favours equilibrium (x_3, y_3) (clear from Equation (4.55));
2. larger $S > 0$ favours absorption on $(q, 1-q)$;

3. smaller q with $\frac{1}{2} < q < 1$ favours equilibrium (x_3, y_3).

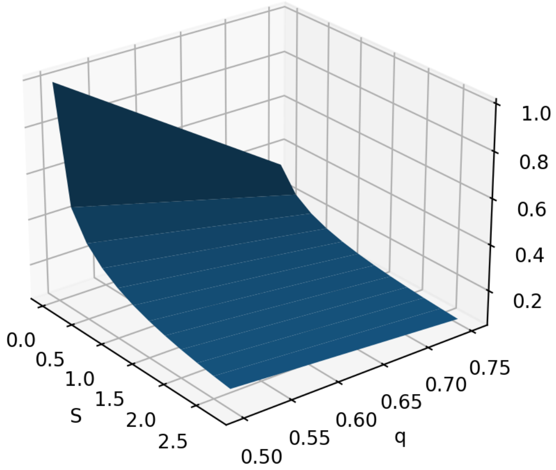

Figure 4.5: Surface plot of the function $h(S,q) = \frac{(1+S) - q\sqrt{\frac{(1+S)^2+4q^2-4q}{q^2}}}{2q}$ (see Equation (4.55)). For large values of $(1+S)$ the function value is small; fixing $(1+S)$, the function value is small for large values of q.

Taking all things into consideration, we see that the third equilibrium of the MFVM with selection is stable (see Figure 4.3). This shows that the two alleles A and B can coexist in both populations even when there is a selection factor. In the article of Borile et al. [2] they proved that the two-dimensional MFVM has order $\mathcal{O}(e^{NI})$ because of the third equilibrium (not $(0,0)$ or $(q, 1-q)$). Therefore we state the following conjecture: For the two-dimensional MFVM with selection $\{\eta_N(t)\}_{t\geq 0}$ the typical time to absorption T_N has order $\mathcal{O}(e^{NI})$ with $I > 0$. That is:

$$. \tag{4.57}$$

Note that this conjecture is not proven; we recommend further research to obtain a proof (see Chapter 6). As explained earlier, time to absorption is reaching an absorbing state $(q, 1-q)$ or $(0,0)$. In an absorbing state all individuals in both populations only have allele A or only allele B respectively (see Section 1.1). With this conjecture one can conclude that, under certain conditions and with selection, with N large, it takes a very long time (exponentially growing with N) before all the individuals in both populations have the same allele; two different alleles can coexist even when one allele has a selection advantage.

5

Conclusion

In this thesis we extended the two-dimensional stochastic mean-field voter model (MFVM), as in the article of Borile et al. [2], with a selection factor. Therefore we were able to investigate wether individuals with two different alleles can coexist in two populations when there is a selection advantage. We came to the conclusion that, under the circumstance

$$\rho < \frac{(1+S) - q\sqrt{\frac{(1+S)^2+4q^2-4q}{q^2}}}{2q}, \tag{5.1}$$

a stable equilibrium (x_3, y_3) arises (see Figure 4.3 and Equation (4.52)) which is not $(0,0)$ or $(q, 1-q)$. This means that individuals with two different alles can coexist in two populations even when one allele has a selection advantage. In the inequality in Equation (5.1) S is selection factor and ρ and q are parameters of the MFVM (see Section 1.1). We concluded, regarding this admissible stable equilibrium, that:

1. smaller ρ with $0 < \rho < 1$ favours equilibrium (x_3, y_3);
2. larger $S > 0$ favours absorption on $(q, 1-q)$;
3. smaller q with $\frac{1}{2} < q < 1$ favours equilibrium (x_3, y_3).

Since we have found a stable equilibrium which is not an absorbing state $(0,0)$ or $(q, 1-q)$, we stated the following conjecture (see Section 4.4.3): For the stochastic process $\{\eta_N(t)\}_{t\geq 0}$, described by the two-dimensional MFVM with selection, the typical time to absorption T_N has order $\mathcal{O}(e^{NI})$ with $I > 0$.

6

Discussion

When writing this thesis, the following questions and ideas came to mind.

- In this research we choose the selection factor $S \in \mathbb{R}$. However S can also be a function of the population size N. In the work of Avena, da Costa and den Hollander [1] they propose $S = S(N) = \frac{\sigma}{N}$ with $\sigma \in \mathbb{R}$. This means that, as the population size gets larger, the selection advantage becomes smaller. One can also choose other monotone decreasing functions. Since S is a function of N, the limit of the generator as in Section 3.2 and 4.2 is going to differ. Therefore the differential equations as in Section 3.3 and 4.3 also are not the same; the vector fields will differ as well. I recommend doing research in which type of selection factor S is most preferable for certain situations. Then one can also, with the same steps as in this thesis, find equilibrium points of the new vector field.
- The goal of this thesis was to investigate wether it was possible that two individuals with different alleles can coexist. However we did not try to match te model to reality; we only found conditions such that the equilibrium was admissible. For further research I recommend to collect data of numerous biological populations. Then one can try fitting the MFVM with selection to the collected data to obtain the best possible values of the parameters.
- During this bachelor thesis we worked with the MFVM in biological context. Is it possible to look at the MFVM in a different setting? For example, as the name suggest, one can investigate the behaviour of voters for a political party. For instance, the American two party system for different states. In this case an individual with spin equal to 1 votes for the Republican Party and an individual with spin equal to 0 votes for the Democratic Party.
  Or maybe we could use the MFVM to describe the spread of a virus, as the Corona virus, in different countries. Then an individual with spin equal to 1 is infected and an individual with spin equal to 0 is not. Again, as in the previous point, data could be collected to fit the model and the model could be adapted. If this succeeds the MFVM could predict biological diversity, the election in the United States or the spread of the Corona virus between countries.
- We stated a conjecture claiming that the typical time to absorption of the two-dimensional MFVM with selection has order $\mathcal{O}(e^{NI})$. We stated this conjecture because of the resemblance to the two-dimensional MFVM without selection. The proof of the model without selection is given in the article of Borile et al. [2]. The resemblance we found is a third stable equilibrium under certain conditions (see Section 4.4.3). However we have not proven the conjecture in this thesis since it is out of our scope. Further research could provide a proof.

Bibliography

[1] Luca Avena, Conrado da Costa, and Frank den Hollander. *Stochastic Models For Genetic Evolution*. 2019.

[2] Claudio Borile, Paolo Dai Pra, Markus Fischer, Marco Formentin, and Amos Maritan. Time to absorption for a heterogeneous neutral competition model. *Journal of Statistical Physics*, 156 (1):119–130, 2014. doi: 10.1007/s10955-014-0989-8. URL `https://doi.org/10.1007/s10955-014-0989-8`.

[3] Stewart N. Ethier and Thomas G. Kurtz. *Markov processes – characterization and convergence*. Wiley Series in Probability and Mathematical Statistics: Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, 1986. ISBN 0-471-08186-8.

[4] Sandra Kluth and Ellen Baake. The moran model with selection: Fixation probabilities, ancestral lines, and an alternative particle representation, 2013.

[5] Fred E. Szabo. About the author. In Fred E. Szabo, editor, *The Linear Algebra Survival Guide*. Academic Press, Boston, 2015. ISBN 978-0-12-409520-5. doi: https://doi.org/10.1016/B978-0-12-409520-5.50006-0. URL `https://www.sciencedirect.com/science/article/pii/B9780124095205500060`.

[6] Eric W Weisstein. Markov process. URL `https://mathworld.wolfram.com/MarkovProcess.html`.