

THE FREQUENCY-DEPENDENT WRIGHT-FISHER MODEL: DIFFUSIVE AND NON-DIFFUSIVE APPROXIMATIONS.

FABIO A. C. C. CHALUB AND MAX O. SOUZA

ABSTRACT. We study a class of processes that are akin to the Wright-Fisher model, with transition probabilities weighted in terms of the frequency-dependent fitness of the population types. Following an inverse numerical analysis approach, we obtain a family of partial differential equations (PDE) for the evolution of the probability density, and which will be an approximation of the discrete process in the joint large population and weak selection limit. The equations in this family can be purely diffusive, purely hyperbolic or of convection-diffusion type, with frequency dependent convection, and the particular outcome will depend on the assumed scalings. The diffusive equations are of the degenerate type; using a duality approach, we also obtain a frequency dependent version of the Kimura equation without any further assumptions. We also show that the convective approximation is related to the replicator dynamics and provide some estimate of how good is the convective approximation. In particular, we show that the mode, but not the expected value, of the probability distribution is modeled by the replicator dynamics. Some numerical simulations that illustrate the results are also presented. Wright-Fisher process and diffusion approximations and continuous limits and replicator equation [2000]92D15 and 92D25 and 35K57 and 35K67 and 35L65

1. INTRODUCTION

Evolution is naturally a multiscale phenomenon (Keller, 1999; Metz, 2011). The choice of right scale to describe a particular problem has as much art as science. For some populations (e.g, with non overlapping generations) a discrete time provides adequate description; for different examples, this is excessively simplifying. Large populations can be described as infinite (in order to use differential equations, for example), but this imposes limitations in the time validity of the model (Chalub and Souza, 2009b). On the other hand, some finite population effects, like for example, the bottleneck effect, will be missing in any description relying in infinite populations (Hartle and Clark, 2007).

In this vein, diffusion approximations, frequently used for large populations and long time scales, enjoy a long tradition in population genetics. This tradition dates back as early as the work by Feller (1951) and references there in. In particular, diffusion approximations were implicitly used in the pioneering works of Wright (1938, 1937) and Fisher (1922, 1930). These efforts have been further developed in a number of directions as, for instance, in the studies on multispecies models in Sato (1976, 1983). also the review in Sato (1978). This led to a notable progress in the relevant mathematics, as for instance reported in Strock and Varadhan (1997). This in turn led to a large use of diffusion theory in the field, which can be verified in contemporary introductions to the subject (see Ewens, 2004).

There is also a more general approach, called the Kramers-Moyal expansion, where the kernel of the master equation of the stochastic process is fully expanded in a series. The diffusion approximation consists in the Kramers-Moyal expansion truncated at order 2. Although it is commonly claimed that the we need the full expansion to obtain a continuous approximation of discrete processes, we will show that, in fact, the order 2 expansion provides an approximation valid in all time scales. See (Van Kampen, 2001) for a discussion about this and other techniques for continuous approximations of discrete processes.

A complimentary approach to the study of evolution, based on evolutionary game theory, has also been developed (cf. Smith, 1982) with conclusions that are not always compatible with results

from diffusion theory. As an example, diffusion models without mutation lead to the fixation of a homogeneous population, while frequency dependent models associated to the replicator dynamics¹ may lead to stable mixed populations. For an introduction to evolutionary game theory and replicator dynamics, we refer the reader to Hofbauer and Sigmund (1998) and Weibull (1995).

Consistent interaction among these two modelling schools have been attempted by a number of authors, with different degrees of success (see Traulsen et al., 2005; Lessard and Ladret, 2007; Lessard, 2005; McKane and Waxman, 2007; Waxman, 2011). We will show that both descriptions — the one based on the diffusion approximation and the one based on the replicator dynamics — are both correct as models for the evolutionary dynamics of a given trait, but in different time scales. As byproducts, we will provide a generalization of the Kimura equation valid for an arbitrary number of types and general fitnesses; we will also prove that the replicator equation is a model valid locally in time and that the solution of the replicator equation indicates the most probable state (mode) to find a population, not the expected value of the trait.

1.1. Consistent approximations. In order to be able to study somewhat more general models, we follow the approach used by the authors in Chalub and Souza (2009b). In particular, we are interested not only in diffusion approximations, but in approximations that can be consistent with the dynamics of the corresponding discrete process.

We begin with a definition:

Definition 1 (Dynamically Consistent Approximation). *We shall say that a simplified model \mathcal{M}_0 is a dynamically consistent approximation (DCA or, in short, an approximation) of the family of detailed models \mathcal{M}_γ , $\gamma > 0$, in a sense χ (e.g., point wise, L^2 , L^∞ , etc) if the following holds true:*

- (1) *Consider a certain family of initial conditions h_γ^I such that $\lim_{\gamma \rightarrow 0} h_\gamma^I = h_0^I$ in the sense χ ;*
- (2) *Evolve through the model \mathcal{M}_γ the initial condition h_γ^I and through the model \mathcal{M}_0 the initial condition h_0^I until the time $t_{\max} \in (0, \infty]$ obtaining $h_\gamma(t)$ and $h_0(t)$ respectively, for $t < t_{\max}$;*

If, for every time $t < t_{\max}$, we have that $\lim_{\gamma \rightarrow 0} h_\gamma(t) = h_0(t)$, in the sense χ , then we say that the model \mathcal{M}_γ converge in the limit $\gamma \rightarrow 0$ until time t_{\max} , in the sense χ , to the model \mathcal{M}_0 .

If $t_{\max} \neq \infty$, we say that the model \mathcal{M}_0 approximates the initial dynamics (or locally in time) of the detailed model \mathcal{M}_{γ_0} , $\gamma_0 \ll 1$; if $t_{\max} = \infty$, we then say that \mathcal{M}_0 approximates the detailed model \mathcal{M}_{γ_0} , $\gamma_0 \ll 1$ for all time scales (or globally in time).

Some examples of the relation between detailed and simplified models are listed in Table 1.

In general, some extra assumptions are frequently required. If, for example, there are more than one small parameter in the detailed model, it is natural to assume a relationship among them, called *scaling*, as, in general, the limit model will depend on how these parameters approaches zero. Other assumptions may also be necessary, as it will be discussed in the next paragraph. The process of taking the limit of a family of models, considering a given scaling, will be called “the thermodynamical limit”; by extension, we shall also call the limit model the *thermodynamical limit*. In this work, depending on the precise choice of the scaling, the limit equation can be of drift-type (a partial differential equation fully equivalent to the replicator equation or system), of purely diffusion type, or, in a delicate balance, of drift-diffusion type.

In what follows, an important and natural assumption that must be introduced in order that we have as approximation in the DCA sense is the so-called *weak selection principle*, to be precisely stated in equation (10). Generally speaking, we assume that the *fitness* of a given individual (i.e., the probability of finding descendants of this individual in the next generation) decreases to 1 when the time separation between two successive generations Δt approaches zero. This is a natural assumption when we consider that two successive generations collapses into a single one. However, in most of the literature, the weak selection principle is assumed in the limit of $N \rightarrow \infty$, where N is the population size. Although they are equivalent (as we shall assume a certain scaling relation between N and Δt), we consider our approach more natural.

¹In this work, we will use the expressions “replicator dynamics”, “replicator equation” and “replicator system” indistinctly.

Detailed model	Meaning of parameter γ	Simplified model
Kinetic models	mean free path	hydrodynamical models
Othmer-Dumbar-Alt model	mean free path	Keller-Segel model
Quantum Mechanics	rescaled Planck constant	Classical Mechanics
Relativistic mechanics	(rescaled light velocity) ⁻¹	Non-relativistic Mechanics
Moran process	inverse of population size	replicator-diffusion equation
Moran process	inverse of population size	replicator equation

TABLE 1. Detailed and reduced models. The last two lines state that both the replicator equation and the replicator diffusion equation approximates the Moran process, but in different time scales. References to these works are (Bardos et al., 1991, 1993; Cercignani, 2002; Hillen and Othmer, 2000; Othmer and Hillen, 2002; Chalub et al., 2004; Stevens, 2000; Hepp, 1974; Cirincione and Chernoff, 1981; Bjorken and Drell, 1964; Chalub and Souza, 2009a,b).

1.2. Inverse numerical analysis. We shall show that the evolution given by the Wright-Fisher process can be approximated for all time scales by a certain differential equation (the simplified model), to be introduced, that we call the replicator-diffusion equation. The Wright-Fisher process is an evolutionary process for an asexual population of N individuals, constant in size, divided in n different types, that evolves according to a specific rule, with fixed time separation between generations of $\Delta t > 0$ (the detailed model in the discussion above, where γ is the inverse of the population size — or, as we shall see, equivalently, the intergeneration time). If we are interested only in the first time scale of the Wright-Fisher process, we shall assume different scalings and obtain as simplified limit the replicator equation, a first order ordinary differential system.

In short, given a certain scaling and the weak selection principle, we find a certain partial-differential equation of drift-diffusion type with degenerated coefficients as the thermodynamical limit of the Wright-Fisher process (to be rigorously introduced and analyzed for finite populations in section 3).

The derivation is made in four steps:

- (1) The discrete process is studied and shown that its solution is convergent to a linear superposition of homogeneous states for large times. We shall then say, in the language of numerical analysis, that the discrete process is *stable*.
- (2) An asymptotic expansion of the process in powers of $1/N$ is made, and we pick the leading order term. This is then applied to a sufficient smooth function in a dense class. The outcome will be a differential operator. Again, following the language of numerical analysis, we shall say then that the discrete process is *consistent* with the differential operator.
- (3) We show that the discrete process has n linearly independent linear conservation laws and we construct conservation laws at the continuous level that are natural extensions of the conservation laws at the discrete level when $N \rightarrow \infty$.
- (4) The differential operator together with the conservation laws is shown to define a well-posed problem.

We are now in a position to invoke the celebrated Lax-Richtmyer equivalence theorem (Richtmyer and Morton, 1967) which asserts, for linear problems, that a discrete process that is consistent converges, if and only if, it is stable.

Remark 1. *The process of obtaining a equation for which the discrete process is consistent will be termed a continuation of the discrete process. Notice that, for a given discrete process, we can have a number of different continuations that will be obtained in different scalings and valid in different domains.*

Remark 2. *The inverse numerical analysis perspective offers a natural approach to introduce a number of features in modelling, that are usually included by means of ad-hoc constructions.*

1.3. Outline. In this paper we extend the study for the two-types Moran process in Chalub and Souza (2009a,b) to the Wright-Fisher process with an arbitrary number of types. In Section 2,

we introduce the basic notation and make an extended abstract of our main results. In section 3, we will study in detail the discrete process (the finite population Wright-Fisher process). In section 4, we find the continuation of the Wright-Fisher process. In particular, we derive the replicator-diffusion equation; in section 5 we study the replicator-diffusion equation, showing the main properties of its solutions; in particular, we sketch the proof of convergence of the solution of the discrete model to the solution of the continuous model. Following Chalub and Souza (2009a,b), we will present the more technical proofs elsewhere. We also show that the probability distribution associated with all types in the population concentrates along the evolutionary stable states. Additionally, in subsection 5.2, we obtain the backward equation as the proper dual of the replicator-diffusion equation, providing a consistent generalization of the Kimura equation for the n types and arbitrary fitness functions. In section 6, we show that the initial dynamics of the replicator-diffusion equation is given by the replicator equation (or system), thus showing that in the limit of short times and strong selection² the Wright-Fisher process will be well approximate by the trajectories of the replicator equation. In particular, we say that the replicator equation is a dynamically consistent approximation of the Wright-Fisher process, but valid only locally in time (while the replicator-diffusion equation is a DCA valid globally in time). It is important to note that, for intermediate times, the replicator equation will approximate the mode of the discrete evolution³, but not the expected value of a given trait, as will be numerically exemplified in section 7. Conclusions are presented in section 8.

2. PRELIMINARIES AND MAIN RESULTS

We begin by introducing the space of states for the evolution:

Definition 2. Let $\mathbb{R}_+ = [0, \infty)$. We define the $n - 1$ dimensional simplex

$$S^{n-1} := \left\{ \mathbf{x} \in \mathbf{R}_+^n \mid |\mathbf{x}| := \sum_{i=1}^n x_i = 1 \right\}.$$

We also define the set of vertexes of the simplex $\Delta S^{n-1} := \{\mathbf{x} \in S^{n-1} \mid \exists i, x_i = 1\}$, its interior $\text{int } S^{n-1} := \{\mathbf{x} \in S^{n-1} \mid \forall i, x_i > 0\}$ and its boundary $\partial S^{n-1} = S^{n-1} \setminus \text{int } S^{n-1}$. The state of the population is a vector $\mathbf{x} \in S^{n-1}$. The elements of ΔS^{n-1} are denoted \mathbf{e}_i , $i = 1, \dots, n$ and called “homogeneous states”. A vector $\mathbf{x} \in S^{n-1} \setminus \Delta S^{n-1}$ is a “mixed state”.

In what follows, we let $p(\mathbf{x}, t)$ to be the probability density of finding the population at state $\mathbf{x} \in S^{n-1}$ at time $t \geq 0$.

Definition 3. We call the fitness of a given type a sufficiently smooth function $\psi^{(i)} : S^{n-1} \rightarrow \mathbb{R}_+$, and the average fitness in a given population is given by $\bar{\psi}(\mathbf{x}) := \sum_{i=1}^n x_i \psi^{(i)}(\mathbf{x})$. Note that we consider the fitness function to not depend explicitly on time.

In this work, we introduce a family of detailed models described by a parabolic equation of drift-diffusion type, with degenerated coefficients (DiBenedetto, 1993; Carrol and Schowalter, 1976), defined in the simplex S^{n-1} , called *the replicator-diffusion equation*, namely:

$$(1) \quad \begin{cases} \partial_t p = \mathcal{L}_{n-1, x} p := \frac{\kappa}{2} \sum_{i,j=1}^{n-1} \partial_{ij}^2 (D_{ij} p) - \sum_{i=1}^{n-1} \partial_i (\Omega_i p) , \\ D_{ij} := x_i \delta_{ij} - x_i x_j , \\ \Omega_i := x_i (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) , \end{cases}$$

with $i, j = 1, \dots, n - 1$, $\kappa > 0$, and where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise is the Kronecker delta. The above equation has a solution in the classical sense (i.e., everywhere differentiable). Furthermore, in the classical sense, it is a well posed problem, without any boundary conditions. However, this classical solution is not the correct limit of the discrete process. In order to find the correct limit, equation (1) is to be supplemented with n conservation laws. From now on, whenever we refer to the replicator-diffusion equation (1), we are implicitly assuming these conservation laws.

Our main conclusions are:

²Strong selection in this context is not directly related or opposed to weak selection as introduced before.

³We call “mode” the most probable state in $\text{int } S^{n-1}$, and not in S^{n-1} (see definition 2).

- (1) An analysis of the equation (1) leads to a unique solution of measure type. This will require definitions of appropriate functional spaces.
- (2) This unique solution approximates, in the thermodynamical limit, the evolution of a discrete population by the Wright-Fisher process in all time scales.
- (3) A reduced model, obtained by setting $\kappa = 0$ in (1) (with only one conservation law), is shown to be equivalent to the replicator dynamics. This will show that the replicator dynamics approximates the discrete process only locally in time;
- (4) Furthermore, the solution of the replicator equation models the time evolution of the mode of the probability distribution associated to the discrete process (and not the *expected value* of the same distribution);
- (5) A frequency dependent generalization of the Kimura equation for an arbitrary number of types is obtained by looking at the dual problem for (1).

Before going into the technical details, we explain the last paragraph a little further.

Equation (1) has two natural time scales, one for the natural selection (the mathematical drift and, as we shall see, fully compatible with the replicator equation), the second for the genetic drift (the mathematical diffusion). That is why we call equation (1) together with the conservation laws to be introduced in subsection 4.2, the “replicator-diffusion equation”. More precisely, the solution of the replicator-diffusion equation when $\kappa = 0$ (which is of hyperbolic type) is the leading order term of the solution p_κ of the replicator-diffusion equation for small κ (i.e., large fitness and/or short times). The replicator-diffusion equation with zero diffusion ($\kappa = 0$) happens to be the replicator equation (or system) (Hofbauer and Sigmund, 1998). In an appropriate sense, to be made precise in section 6.3 (theorem 7), we have $p_\kappa \xrightarrow{\kappa \rightarrow 0} p_0$, locally in time⁴.

This theorem cannot be made global in time, for general fitness functions and initial conditions, as the Wright-Fisher process always converge in $t \rightarrow \infty$ to a linear combination of homogeneous states (while it is possible that the solution of the replicator equation converges to a stable mixed state).

This is the mathematical formulation of a known principle in evolutionary biology that states that “given enough time every mutant gene will be fixed or extinct.” (Kimura, 1962). This means that the final state of the replicator-diffusion equation with any $\kappa > 0$ will be a linear combination of Dirac deltas at the vertexes of the simplex S^{n-1} . Actually, for any positive time, the solution of equation (1) with the conservation laws described above is a sum of a classical function in the simplex plus a sum of singular measures over all the subsimplexes on ∂S^{n-1} and, inductively, on their boundaries subsimplexes. In particular, we shall have also Dirac measures supported on the vertexes of the simplex. These measures appears immediately, i.e., for any $t > 0$. The full evolution and the final states of the replicator-diffusion equation will be studied in section 5.

From the practical point of view, we are, however, often interested in transient states (“in the long run, we are all dead”, said John Maynard Keynes), specially because the transient states become more and more important for the discrete evolution as the population size increases. Heuristically, when the population is large the stochastic fluctuations decreases in importance, and therefore, its evolution is deterministic. The associated limit will be given by equation (1), with $\kappa = 0$, i.e., the hyperbolic limit of equation (1). This equation does not develop finite-time singularities. This is one more peculiarity of equation (1): diffusion seems to be a deregularizing effect; the solution of the parabolic replicator-diffusion equation $\kappa > 0$ is less regular than the solution of the hyperbolic null-diffusion limit $\kappa = 0$, contrary to most examples in the literature (John, 1991; Folland, 1995). See, on the other hand, (Murray, 2003), for diffusion driven instability.

The relationships between the three models is summarized in Figure 1.

3. THE DISCRETE MODEL

In this section, we study the discrete model, i.e., the Wright-Fisher model for constant population, arbitrary number of types and arbitrary fitnesses functions. We start, in subsection 3.1 with basic mathematical definitions; in subsection 3.2 we prove that a state is a stationary state

⁴Except for a special range of parameters (Chalub and Souza, 2011a).

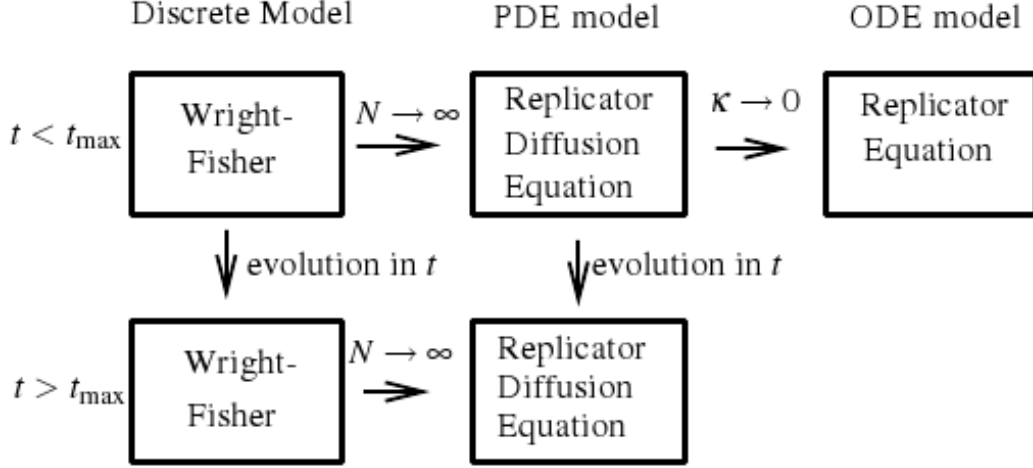


FIGURE 1. The boxes in the figure represents the solutions of three different models: the Wright-Fisher process (finite population N), the replicator-diffusion equation (positive diffusion κ) and the replicator equation. The vertical axis indicates the arrow of time (top-down), and the horizontal axis indicates, first the large population limit, secondly the no-diffusion limit. Note that the no-diffusion limit of the replicator-diffusion equation for large times has no relation with the solutions of the replicator equation at the same time; therefore, the replicator equation should be valid only locally in time. Here t_{\max} indicates the maximal time for the approximation of the model by the replicator ODE, according to the definition 1.

of the discrete evolution if and only if it is homogeneous; therefore, there are exactly n linearly independent stationary state; finally, we prove that the final state is a linear superposition of these independent stationary states, with coefficients that depend on the initial condition and that can be calculated from a set of n linearly independent conservation laws. All these results will be useful in the correct determination of the continuous process, to be done in sections 4 and 5. The Wright-Fisher process was studied, with different level of details in, for example, (Ewens, 2004; Nowak, 2006; Imhof and Nowak, 2006), but, to the best of our knowledge the conservation laws associated to the process were overlooked.

The fact that the final state in the Wright-Fisher process, among others, for a finite population is always homogeneous was also a matter of dispute with respect to the validity of the modelling (Vickery, 1988; Smith, 1988). As we will shortly see in this work, this dispute is basically a consequence of the existence of two different time scales hidden in the model: the non-diffusive (drift) and the diffusive one.

3.1. Preliminaries. We consider a fixed size population of N individuals at time t consisting of a fraction $x_i \in \{0, \frac{1}{N}, \frac{2}{N}, \dots, 1\}$ of individuals of type $i = 1, 2, \dots, n$. The population evolves in discrete generations with time-step separation of Δt . We introduce the following notation:

Definition 4. *The state of a population is defined by a vector in the N -discrete $n - 1$ -dimensional simplex*

$$S_N^{n-1} := \left\{ \mathbf{x} = (x_1, \dots, x_n) \mid |\mathbf{x}| := \sum_{i=1}^n x_i = 1, x_i \in \left\{ 0, \frac{1}{N}, \frac{2}{N}, \dots, 1 \right\} \right\} .$$

We also define the set of vertexes of the $n - 1$ -dimensional simplex

$$\Delta S_N^{n-1} := \{ \mathbf{x} \in S_N^{n-1} \mid \exists i, x_i = 1 \} = \{ \mathbf{e}_i \mid i = 1, \dots, n \} .$$

The elements of ΔS_N^{n-1} are called the homogeneous states. To each type we attribute a function, called fitness, $\Psi_{\Delta t}^{(i)} : S_N^{n-1} \rightarrow \mathbb{R}$. It is convenient to assume that $\Psi_{\Delta t}^{(i)}$ is a discretization of a smooth function on the simplex S^{n-1} ; more assumptions on $\Psi_{\Delta t}^{(i)}$ will be introduced in section 4.

A population at time $t + \Delta t$ is obtained from the population at time t choosing N individuals with probability proportional to the fitness. More precisely, we define the average fitness $\bar{\Psi}_{\Delta t}(\mathbf{x}) = \sum_{i=1}^n x_i \Psi_{\Delta t}^{(i)}(\mathbf{x})$ and then the transition probability from a population at state \mathbf{y} to a population at state \mathbf{x} is given by

$$(2) \quad \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x}) = \frac{N!}{(Nx_1)!(Nx_2)! \cdots (Nx_n)!} \prod_{i=1}^n \left(\frac{y_i \Psi_{\Delta t}^{(i)}(\mathbf{y})}{\bar{\Psi}_{\Delta t}(\mathbf{y})} \right)^{Nx_i}.$$

The evolutionary process given by a Markov chain with transition probabilities given by equation (2) is called the (frequency dependent) Wright-Fisher process.

Let $\mathcal{P}(t) = (P(\mathbf{x}, t))_{\mathbf{x} \in S_N^{n-1}}$, with

$$\mathcal{P} \in \Upsilon := \{P : S_N^{n-1} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+ \mid \sum_{\mathbf{x} \in S_N^{n-1}} P(\mathbf{x}, \cdot) = 1\},$$

where $P(\mathbf{x}, t)$ is the probability of finding the population at a given state $\mathbf{x} \in S_N^{n-1}$ at time t . Then, the evolution is given by

$$(3) \quad P(\mathbf{x}, t + \Delta t) = (\mathcal{T}\mathcal{P}(t))(\mathbf{x}) := \sum_{\mathbf{y} \in S_N^{n-1}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x}) P(\mathbf{y}, t).$$

The probability conservation is a consequence of the definition (2) and reads

$$(4) \quad \sum_{\mathbf{x} \in S_N^{n-1}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x}) = 1, \quad \forall \mathbf{y} \in S_N^{n-1}.$$

3.2. Stationary states, final states and conservation laws. We call an homogeneous population a population of a single type, i.e., $P(\mathbf{x}, t) = \hat{P}_{\mathbf{v}}(\mathbf{x})$ for $\mathbf{v} \in \Delta S_N^{n-1}$, where

$$\hat{P}_{\mathbf{x}}(\mathbf{y}) = \begin{cases} 1, & \mathbf{y} = \mathbf{x}, \\ 0, & \mathbf{y} \neq \mathbf{x}. \end{cases}$$

From the fact that it is always possible to have a homogeneous population in a single step, we also have that

$$(5) \quad \sum_{\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x}) < 1, \quad \forall \mathbf{y} \in S_N^{n-1}.$$

Furthermore, we suppose there is no mutation in the evolutionary process, more precisely, for $\mathbf{y} \in S_N^{n-1}$, with $y_i = 0$, for a certain i , then for any vector $\mathbf{x} \in S_N^{n-1}$ such that $x_i \neq 0$, we have that $\Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x}) = 0$. In particular

$$(6) \quad \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} = \mathbf{y} \in \Delta S_N^{n-1}, \\ 0 & \text{if } \mathbf{x} \neq \mathbf{y} \in \Delta S_N^{n-1}. \end{cases}$$

In the remainder of this section, we will consider a function $f : S_N^{n-1} \rightarrow \mathbb{R}_+$ also as a vector in the non-negative cone of a Euclidean space of sufficiently high dimension K , where K is the number of points in S_N^{n-1} ; i.e. $f : \{1, \dots, K\} \rightarrow \mathbb{R}_+$. The order of this set is irrelevant to the sequel. Therefore, the transition operator given by equation (2) can be seen as a transition matrix in \mathbb{R}^K which preserves the non-negative cone \mathbb{R}_+^K .

Lemma 1. *Let $f : S_N^{n-1} \rightarrow \mathbb{R}_+$ be such that there is $\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1}$, such that $f(\mathbf{x}) > 0$. Then,*

$$\sum_{\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1}} (\mathcal{T}f)(\mathbf{x}) < \sum_{\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1}} f(\mathbf{x}).$$

Proof.

$$\begin{aligned}
\sum_{\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1}} (\mathcal{T}f)(\mathbf{x}) &= \sum_{\substack{\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1} \\ \mathbf{y} \in S_N^{n-1} \setminus \Delta S_N^{n-1}}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x})f(\mathbf{y}) + \sum_{\substack{\mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1} \\ \mathbf{y} \in \Delta S_N^{n-1}}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x})f(\mathbf{y}) \\
&= \sum_{\substack{\mathbf{y} \in S_N^{n-1} \setminus \Delta S_N^{n-1} \\ \mathbf{x} \in S_N^{n-1} \setminus \Delta S_N^{n-1}}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x})f(\mathbf{y}) \quad (\text{from equation (6)}) \\
&< \sum_{\mathbf{y} \in S_N^{n-1} \setminus \Delta S_N^{n-1}} f(\mathbf{y}) \quad (\text{from equation (5)}) .
\end{aligned}$$

□

□

This immediately implies the following Theorem:

Theorem 1. *A function f defined in S_N^{n-1} is a fixed state of the operator \mathcal{T} if, and only if, f is a linear combination of homogeneous states. In particular, \mathcal{T} has exactly n linearly independent eigenfunctions associated to the eigenvalue $\lambda = 1$.*

Proof. From equation (6) it is clear that all homogeneous states are stationary states. From Lemma 1 it is clear that no non-homogeneous state can be a stationary state. It is immediate that the algebraic and geometric multiplicities of $\lambda = 1$ are equal. Suppose, by contradiction, that there is a generalized eigenvector f of \mathcal{T} . Therefore, there is an homogeneous state $\mathbf{v} \in \Delta S_N^{n-1}$ such that

$$\begin{aligned}
\sum_{\mathbf{y} \in S_N^{n-1}} f(\mathbf{y}) &= \sum_{\mathbf{x}, \mathbf{y} \in S_N^{n-1}} \Theta_{N,\Delta t}(\mathbf{y} \rightarrow \mathbf{x})f(\mathbf{y}) = \sum_{\mathbf{x} \in S_N^{n-1}} (\mathcal{T}f)(\mathbf{x}) \\
&= \sum_{\mathbf{x} \in S_N^{n-1}} [f(\mathbf{x}) + \hat{P}_{\mathbf{v}}(\mathbf{x})] = \sum_{\mathbf{x} \in S_N^{n-1}} f(\mathbf{x}) + 1 .
\end{aligned}$$

□

□

A convenient eigenbase can be obtained by considering one vector for each homogeneous state. More precisely, consider the states given by $\hat{P}_{\mathbf{x}}$. Immediately

$$\langle \hat{P}_{\mathbf{x}}, \hat{P}_{\mathbf{y}} \rangle = \begin{cases} 1, & \mathbf{x} = \mathbf{y}, \\ 0, & \mathbf{x} \neq \mathbf{y}, \end{cases}$$

where

$$\langle v, w \rangle := \sum_{\mathbf{x} \in S_N^{n-1}} v(\mathbf{x})w(\mathbf{x}) .$$

From equation (4) it is clear that all eigenvalues λ are such that $|\lambda| \leq 1$. The set $\{\hat{P}_{\mathbf{e}_1}, \hat{P}_{\mathbf{e}_2}, \dots, \hat{P}_{\mathbf{e}_n}\}$ is a complete set of eigenvectors associated to $\lambda = 1$ and from lemma 1 all other eigenvectors are such that its associated eigenvalue is $|\lambda| < 1$. This implies the existence of

$$\mathcal{T}^\infty f = \lim_{k \rightarrow \infty} \mathcal{T}^k f, \quad f(\mathbf{x}) \geq 0, \quad \forall \mathbf{x} \in S_N^{n-1} .$$

Therefore, we state:

Corollary 1. *For all non-negative initial condition P^I , the final result is a linear combination of homogeneous states,*

$$P^\infty := \lim_{t \rightarrow \infty} P(\cdot, t) = \sum_{i=1}^n c_i [P^I] \hat{P}_{\mathbf{e}_i} .$$

Definition 5. *We define a linear conservation law as one given by a linear functional over the functions of S_N^{n-1} . A set of linear conservation laws is linearly independent, if the only linear combinations providing a trivial conservation law is the trivial one.*

Now, consider the adjoint evolution operator \mathcal{T}^\dagger . There are n linearly independent eigenvectors associated to the eigenvalue $\lambda = 1$. To each of them, we obtain a linear conservation law for the evolution given by \mathcal{T} . In effect, let w be one of these eigenvectors.

$$(7) \quad \langle \mathcal{P}(t + \Delta t), w \rangle = \langle \mathcal{T}\mathcal{P}(t), w \rangle = \langle \mathcal{P}(t), \mathcal{T}^\dagger w \rangle = \langle \mathcal{P}(t), w \rangle .$$

As there are n linearly independent eigenvectors, there are n linearly independent conservation laws associated to the direct evolution given by equation (3).

Now, we consider a particular set of eigenvectors, associating conservation laws and the fixation probabilities of the n different types. Given any initial condition $P \in \mathcal{Y}$, we define $F_P^{(i)}$ as the fixation probability of the type i in a population initially in the state P . From the fact that

$$\mathcal{T}^\infty P = \sum_{i=1}^n F_P^{(i)} \hat{P}_{\mathbf{e}_i}$$

we find

$$(8) \quad F_P^{(i)} = (\mathcal{T}^\infty P)(\mathbf{e}_i) = \langle \mathcal{T}^\infty P, \hat{P}_{\mathbf{e}_i} \rangle = \langle P, (\mathcal{T}^\dagger)^\infty \hat{P}_{\mathbf{e}_i} \rangle .$$

In particular

$$\sum_{\mathbf{x} \in S_N^{n-1}} F_{\hat{P}_{\mathbf{x}}}^{(i)} \hat{P}_{\mathbf{x}} = \sum_{\mathbf{x} \in S_N^{n-1}} \langle \hat{P}_{\mathbf{x}}, (\mathcal{T}^\dagger)^\infty \hat{P}_{\mathbf{e}_i} \rangle \hat{P}_{\mathbf{x}} .$$

Finally,

$$(\mathcal{T}^\dagger)^\infty \hat{P}_{\mathbf{e}_i} = \sum_{\mathbf{x} \in S_N^{n-1}} F_{\hat{P}_{\mathbf{x}}}^{(i)} \hat{P}_{\mathbf{x}} .$$

Therefore, by defining

$$(9) \quad \mathbf{F}^{(i)} = \sum_{\mathbf{x} \in S_N^{n-1}} F_{\hat{P}_{\mathbf{x}}}^{(i)} \hat{P}_{\mathbf{x}} , \quad i = 1, \dots, n ,$$

we have n eigenvectors of \mathcal{T}^\dagger , one for each $i = 1, \dots, n$. In fact

$$\mathcal{T}^\dagger \mathbf{F}^{(i)} = \mathcal{T}^\dagger (\mathcal{T}^\dagger)^\infty \hat{P}_{\mathbf{e}_i} = (\mathcal{T}^\dagger)^\infty \hat{P}_{\mathbf{e}_i} = \mathbf{F}^{(i)} .$$

In particular,

$$\mathbf{F}^{(i)}(\mathbf{e}_j) = \langle (\mathcal{T}^\dagger)^\infty \hat{P}_{\mathbf{e}_i}, \hat{P}_{\mathbf{e}_j} \rangle = \langle \hat{P}_{\mathbf{e}_i}, \mathcal{T}^\infty \hat{P}_{\mathbf{e}_j} \rangle = \langle \hat{P}_{\mathbf{e}_i}, \hat{P}_{\mathbf{e}_j} \rangle = \delta_{ij} .$$

The set $\{\mathbf{F}^{(1)}, \dots, \mathbf{F}^{(n)}\}$ provides n linearly independent conservation laws; however none of them is the conservation of probability, the most natural conservation law associated to a Markov process. Indeed, for any $\mathbf{x} \in S_N^{n-1}$

$$\sum_{i=1}^n \mathbf{F}^{(i)}(\mathbf{x}) = \sum_{i=1}^n F_{\hat{P}_{\mathbf{x}}}^{(i)} = 1 .$$

In order to have the conservation of probability as one of the conservation laws of our system, we define \mathbf{G} such that $\mathbf{G}(\mathbf{x}) = K^{-1}$ where the real positive constant K is such that $\sum_{\mathbf{x} \in S_N^{n-1}} K^{-1} = 1$ and prove that

Lemma 2. *The set $\{\mathbf{F}^{(i)}\}_{i=1, \dots, n-1} \cup \mathbf{G}$ is linearly independent.*

Proof. Assume that $\alpha_0 \mathbf{G} + \sum_{i=1}^{n-1} \alpha_i \mathbf{F}^{(i)} = 0$, i.e., for every $\mathbf{x} \in S_N^{n-1}$,

$$\alpha_0 \mathbf{G}(\mathbf{x}) + \sum_{i=1}^{n-1} \alpha_i \mathbf{F}^{(i)}(\mathbf{x}) = 0 .$$

For $\mathbf{x} = \mathbf{e}_n$ we find $\alpha_0 = 0$ and for $\mathbf{x} = \mathbf{e}_i$ we find $\alpha_i = -K^{-1} \alpha_0 = 0$, $i = 1, \dots, n-1$. \square \square

Lemma 2 and equation (9) provide n distinguished linearly independent conservation laws. These laws will be instrumental to obtain the correct notion of solution in the continuous model.

4. CONTINUATIONS OF THE DISCRETE MODEL

The aim of this section is to obtain a differential equation that approximates the discrete evolution, when the population is large ($N \rightarrow \infty$) and there is no time-separation between successive generations ($\Delta t \rightarrow 0$). The relevant variables, $\mathbf{x} \in S^{n-1}$ and $t > 0$ will be forced to be continuous. The correct heuristics to understand the procedure of taking the limit is the one described in section 1, the dynamically consistent approximation.

We begin in subsection 4.1 developing three models, based on partial differential equations formally obtained from the Wright-Fisher process, when $N \rightarrow \infty$ and $\Delta t \rightarrow 0$ (see equations (14), (15) and (1'), respectively). There is no "right choice" of the simplified model. As we could expect, models valid locally in time are simpler than models that are valid globally in time. For example, the model given by equation (14) is equivalent to a system of a ordinary differential equations; actually, it is exactly equivalent to the well-know replicator dynamics (see Hofbauer and Sigmund, 1998). On the other hand, the diffusive approximation, given by equation (15), is a parabolic partial differential equation that is much simpler to solve than the full model; in fact, explicit solutions are known using Gegenbauer polynomials (Ewens, 2004). Our focus will be on the global in time approximation, equation (1').

Results known for the Wright-Fisher process, and stated in section 3 will guide the formal derivation, i.e., the choice of the right thermodynamical limit. After the formal step, it is necessary to prove directly all required mathematical properties for the continuous model, without reference to the discrete model.

For the benefit of exposition and for a presentation geared more towards the biological application, we have chosen not to include a number of mathematical details in the present work, which will be presented elsewhere⁵. Nevertheless, see section 5.1 for some results.

We shall enforce that for every conservation law of the discrete process (described in Lemma 2) should correspond a conservation law in the continuous model. This will be done in subsection 4.2 referring to the discrete case. The conservation laws can also be derived directly in the continuous case, restricting the functional space of interest properly; see remark 7. As a by product, the final state of the continuous model shall be a linear superposition of homogeneous states (see corollary 1 and compare it with theorem 2).

4.1. Formal derivation. We start the formal construction of a differential equation that approximates the discrete evolution given by (3) when N is large. First, we define the density of probability

$$p_N := \frac{P}{1/N} .$$

We shall assume that $p_N \rightarrow p$ weakly. Then, with appropriate scalings for $\Delta t(N)$, there will be a continuous equation for the evolution of the density probability p , defined in the simplex S^{n-1} .

Using the Stirling formula $x! \approx \sqrt{2\pi x} x^x e^{-x}$ we write

$$\frac{N!}{(Nx_1)!(Nx_2)! \cdots (Nx_n)!} \approx \frac{(2\pi)^{\frac{1-n}{2}}}{N^{n-1}} \frac{N^{\frac{n-1}{2}}}{(x_1 x_2 \cdots x_n)^{\frac{1}{2}} x_1^{x_1 N} x_2^{x_2 N} \cdots x_n^{x_n N}} .$$

We also assume the so called weak selection principle

$$(10) \quad \Psi_{\Delta t}^{(i)}(\mathbf{y}) = 1 + (\Delta t)^\nu \psi^{(i)}(\mathbf{y}),$$

and then

$$\bar{\Psi}_{\Delta t}(\mathbf{y}) = 1 + (\Delta t)^\nu \bar{\psi}(\mathbf{y}) ,$$

⁵As for instance in (Chalub and Souza, 2011c)

where $\nu > 0$ is a parameter to be defined⁶. Keeping only order 1 terms in $(\Delta t)^\nu$, we find

$$\begin{aligned} \left(\frac{y_i \Psi_{\Delta t}^{(i)}(\mathbf{y})}{\bar{\Psi}_{\Delta t}(\mathbf{y})} \right)^{Nx_i} &\approx \exp \left\{ Nx_i \left[\log y_i + \log \left(1 + (\Delta t)^\nu \psi^{(i)}(\mathbf{y}) \right) - \log \left(1 + (\Delta t)^\nu \bar{\psi}(\mathbf{y}) \right) \right] \right\} \\ &\approx y_i^{Nx_i} \exp \left[x_i \left(\psi^{(i)}(\mathbf{y}) - \bar{\psi}(\mathbf{y}) \right) \right]. \end{aligned}$$

Finally, we have

$$\Theta_{N, \Delta t}(\mathbf{y} \rightarrow \mathbf{x}) \approx \frac{1}{N^{n-1}} \Lambda(\mathbf{y}, \mathbf{x}, N^{-\frac{1}{2}}) (1 + \Xi(\mathbf{y}, \mathbf{x}, N, \Delta t) + o(1)),$$

where

$$\begin{aligned} \Lambda(\mathbf{y}, \mathbf{x}, z) &:= \frac{(2\pi)^{\frac{1-n}{2}} z^{1-n}}{(x_1 x_2 \cdots x_n)^{\frac{1}{2}}} \prod_{i=1}^n \left(\frac{y_i}{x_i} \right)^{\frac{x_i}{z^{\frac{1}{2}}}} \\ \Xi(\mathbf{y}, \mathbf{x}, N, \Delta t) &:= \sum_{i=1}^n \left[N (\Delta t)^\nu x_i \left(\psi^{(i)}(\mathbf{y}) - \bar{\psi}(\mathbf{y}) \right) \right]. \end{aligned}$$

Note that Ξ is associated to the drift generated by the fitness; i.e., in the neutral case ($\psi^{(i)}$ independent of i), we have that $\Xi(\mathbf{y}, \mathbf{x}, N, \Delta t) = 0$.

We introduce the new variables $\tau_i = y_i \sqrt{N}$ and $z = \frac{1}{\sqrt{N}}$.

Lemma 3. *For large N (and then small z) the neutral transition probability Λ scales as*

$$\Lambda(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z) \approx \frac{(2\pi)^{\frac{1-n}{2}} z^{1-n}}{(x_1 x_2 \cdots x_n)^{\frac{1}{2}}} \exp \left(-\frac{1}{2} \mathcal{Q}(\boldsymbol{\tau}, \boldsymbol{\tau}) \right),$$

where \mathcal{Q} is a quadratic form with associated eigenvalues $\sigma_1, \dots, \sigma_{n-1}$. These eigenvalues are the eigenvalues of the matrix $\mathbf{F} = (F_{ij})$, $i, j = 1, \dots, n-1$, defined by $F_{ii} = x_i^{-1} + x_n^{-1}$ and $F_{ij} = x_n^{-1}$, for $i \neq j$, i.e., $\sigma_1 \cdots \sigma_{n-1} = (x_1 \cdots x_n)^{-1}$. This implies that

$$\int_{\mathbb{R}^{n-1}} \exp \left(-\frac{1}{2} \mathcal{Q}(\boldsymbol{\tau}, \boldsymbol{\tau}) \right) d\boldsymbol{\tau} = (2\pi)^{\frac{n-1}{2}} \sqrt{x_1 \cdots x_n}.$$

Proof. We first write

$$\begin{aligned} \prod_{i=1}^n \left(\frac{x_i - y_i}{x_i} \right)^{Nx_i} &= \left[\prod_{i=1}^{n-1} \left(\frac{x_i - y_i}{x_i} \right)^{x_i} \left(\frac{x_n + \sum_{i=1}^{n-1} y_i}{x_n} \right)^{x_n} \right]^N \\ &= \exp \left(N \left(\sum_{i=1}^{n-1} x_i \log \left(1 - \frac{y_i}{x_i} \right) + x_n \log \left(1 + \frac{1}{x_n} \sum_{j=1}^{n-1} y_j \right) \right) \right). \end{aligned}$$

The maximum happens when $y_i = 0$, and note that

$$\begin{aligned} \sum_{i=1}^{n-1} x_i \log \left(1 - \frac{y_i}{x_i} \right) + x_n \log \left(1 + \frac{1}{x_n} \sum_{j=1}^{n-1} y_j \right) &\approx -\frac{1}{2} \left[\sum_{i=1}^{n-1} \frac{y_i^2}{x_i} + \frac{1}{x_n} \left(\sum_{j=1}^{n-1} y_j \right)^2 \right] \\ &= -\frac{1}{2} \left[\sum_{i=1}^{n-1} \left(\frac{1}{x_i} + \frac{1}{x_n} \right) y_i^2 + \frac{2}{x_n} \sum_{i < j=1}^{n-1} y_i y_j \right] = -\frac{1}{2} \mathcal{Q}(\mathbf{y}, \mathbf{y}) = -\frac{1}{2N} \mathcal{Q}(\boldsymbol{\tau}, \boldsymbol{\tau}), \end{aligned}$$

□

□

⁶See also the discussion on the weak selection principle and the choice of Δt (and not N) in its expansion in Chalub and Souza (2009b).

Lemma 4. *For large N (and then small z) the neutral transition probability Λ has the following first moments:*

$$\begin{aligned} z^{n-1} \int_{\mathbb{R}^{n-1}} \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} &= \int_{S^{n-1}} \Lambda(\mathbf{x}, \mathbf{x} + \mathbf{y}, z) d\mathbf{y} = 1 , \\ z^n \int_{\mathbb{R}^{n-1}} \tau_i \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} &= 0 , \\ z^{n+1} \int_{\mathbb{R}^{n-1}} \tau_i \tau_j \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} &= O(z^3) + z^2 \times (x_i \delta_{ij} - x_i x_j) . \end{aligned}$$

where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise is the Kronecker delta.

Proof. The first one is a simple consequence of normalization. The second one indicates that the average displacement in the neutral case is identically zero. The last one can be directly computed using Lemma 3. Note that the first two equations are exact, while the third one has order three corrections in z . \square \square

The evolution equation for large N , with $d\mathbf{y} = \frac{1}{N^{n-1}}$ is given by

$$\begin{aligned} p(\mathbf{x}, t + \Delta t) &= \sum_{\{\mathbf{y} | \mathbf{x} - \mathbf{y} \in S^{n-1}\}} \Theta_{N, \Delta t}(\mathbf{x} - \mathbf{y} \rightarrow \mathbf{x}) p(\mathbf{x} - \mathbf{y}, t) \\ &\approx \int_{\{\mathbf{y} | \mathbf{x} - \mathbf{y} \in S^{n-1}\}} [1 + \Xi(\mathbf{x} - \mathbf{y}, \mathbf{x}, z, \Delta t)] \Lambda(\mathbf{x} - \mathbf{y}, \mathbf{x}, z) p(\mathbf{x} - \mathbf{y}, t) d\mathbf{y} + O(N^{-2}) \\ &= \int_{\{\boldsymbol{\tau} | \mathbf{x} - z\boldsymbol{\tau} \in S^{n-1}\}} [1 + \Xi(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z, \Delta t)] \Lambda(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z) p(\mathbf{x} - z\boldsymbol{\tau}, t) z^{n-1} d\boldsymbol{\tau} + O(z^4) \\ &\approx \int_{\mathbb{R}^{n-1}} [1 + \Xi(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z, \Delta t)] \Lambda(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z) p(\mathbf{x} - z\boldsymbol{\tau}, t) z^{n-1} d\boldsymbol{\tau} + O(z^4) . \end{aligned}$$

It is clear that $\{\boldsymbol{\tau} \in S^{n-1} | \mathbf{x} - z\boldsymbol{\tau} \in S^{n-1}\} \approx \mathbb{R}^{n-1}$ when z is sufficiently small. Furthermore, we always can impose that the transition probability for a state not in the simplex S^{n-1} is identically zero. From the previous equation, we multiply by an appropriate test function g and integrate

over the simplex to obtain the following equation⁷:

$$\begin{aligned}
 & \int_{S^{n-1}} p(\mathbf{x}, t + \Delta t) g(\mathbf{x}, t) d\mathbf{x} \\
 & \approx z^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} [1 + \Xi(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z, \Delta t)] \Lambda(\mathbf{x} - z\boldsymbol{\tau}, \mathbf{x}, z) p(\mathbf{x} - z\boldsymbol{\tau}, t) g(\mathbf{x}, t) d\boldsymbol{\tau} d\mathbf{x} \\
 & = z^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} [1 + \Xi(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z, \Delta t)] \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) p(\mathbf{x}, t) g(\mathbf{x} + z\boldsymbol{\tau}, t) d\boldsymbol{\tau} d\mathbf{x} \\
 & \approx z^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} \left[1 + zN (\Delta t)^\nu \sum_{i=1}^n \tau_i \left(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \right] \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) p(\mathbf{x}, t) \\
 & \quad \times \left[g(\mathbf{x}, t) + z \sum_{j=1}^{n-1} \tau_j \partial_{x_j} g(\mathbf{x}, t) + \frac{z^2}{2} \sum_{k,l=1}^{n-1} \tau_k \tau_l \partial_{x_k x_l}^2 g(\mathbf{x}, t) \right] d\boldsymbol{\tau} d\mathbf{x} \\
 & \approx z^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) p(\mathbf{x}, t) g(\mathbf{x}, t) d\boldsymbol{\tau} d\mathbf{x} \\
 & \quad + z^n \iint_{\mathbb{R}^{n-1} \times S^{n-1}} p(\mathbf{x}, t) \left[N (\Delta t)^\nu \sum_{i=1}^n \left(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \tau_i + \sum_{j=1}^{n-1} \tau_j \partial_{x_j} g(\mathbf{x}, t) \right] \\
 & \quad \times \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} d\mathbf{x} \\
 & \quad + z^{n+1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} p(\mathbf{x}, t) \left[\sum_{k,l=1}^{n-1} \frac{\tau_k \tau_l}{2} \partial_{x_k x_l}^2 g(\mathbf{x}, t) \right. \\
 & \quad \left. + \sum_{i=1}^n \sum_{j=1}^{n-1} N (\Delta t)^\nu \partial_{x_j} g(\mathbf{x}, t) (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) \tau_i \tau_j \right] \times \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} d\mathbf{x} .
 \end{aligned}$$

We analyze each term in the integrand using lemma 4. The first one is simply

$$z^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) p(\mathbf{x}, t) g(\mathbf{x}, t) d\boldsymbol{\tau} d\mathbf{x} = \int_{S^{n-1}} p(\mathbf{x}, t) g(\mathbf{x}, t) d\mathbf{x} .$$

The second one is identically zero. Now, we obtain the last one, dividing it in two parts:

$$\begin{aligned}
 & z^{n+1} \sum_{k,l=1}^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} p(\mathbf{x}, t) \left[\frac{\tau_k \tau_l}{2} \partial_{x_k x_l}^2 g(\mathbf{x}, t) \right] \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} d\mathbf{x} \\
 & \approx \frac{z^2}{2} \int_{S^{n-1}} \sum_{k=1}^{n-1} (x_k (1 - x_k) \partial_{x_k}^2 g(\mathbf{x}, t)) p(\mathbf{x}, t) d\mathbf{x} \\
 & \quad - \frac{z^2}{2} \int_{S^{n-1}} p(\mathbf{x}, t) \sum_{k,l=1, k \neq l}^{n-1} (x_k x_l \partial_{x_k x_l}^2 g(\mathbf{x}, t)) d\mathbf{x} ,
 \end{aligned}$$

⁷To avoid confusion, over this section we shall write explicitly the variable under differentiation, i.e., ∂_{x_i} instead of ∂_i .

and

$$\begin{aligned}
& z^{n+1} \sum_{i=1}^n \sum_{j=1}^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} p(\mathbf{x}, t) \partial_{x_j} g(\mathbf{x}, t) (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) \tau_i \tau_j \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} d\mathbf{x} \\
& \approx z^2 \sum_{j=1}^{n-1} \sum_{i=1}^n \iint_{\mathbb{R}^{n-1} \times S^{n-1}} p(\mathbf{x}, t) \partial_{x_j} g(\mathbf{x}, t) (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) (x_i \delta_{ij} - x_i x_j) d\mathbf{x} \\
& = z^2 \sum_{j=1}^{n-1} \int_{S^{n-1}} p(\mathbf{x}, t) \partial_{x_j} g(\mathbf{x}, t) (\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) d\mathbf{x} \\
& \quad - z^2 \sum_{j=1}^{n-1} \sum_{i=1}^{n-1} \int_{S^{n-1}} p(\mathbf{x}, t) \partial_{x_j} g(\mathbf{x}, t) (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) x_i x_j d\mathbf{x}.
\end{aligned}$$

From the fact that $\sum_{i=1}^n x_i (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) = 0$ we conclude that the last integral is zero and then

$$\begin{aligned}
& N (\Delta t)^\nu z^{n+1} \sum_{i=1}^n \sum_{j=1}^{n-1} \iint_{\mathbb{R}^{n-1} \times S^{n-1}} p(\mathbf{x}, t) \partial_{x_j} g(\mathbf{x}, t) (\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) \tau_i \tau_j \Lambda(\mathbf{x}, \mathbf{x} + z\boldsymbol{\tau}, z) d\boldsymbol{\tau} d\mathbf{x} \\
& = N (\Delta t)^\nu z^2 \int_{S^{n-1}} \sum_{j=1}^{n-1} \left[x_j (\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) \partial_{x_j} g(\mathbf{x}, t) \right] p(\mathbf{x}, t) d\mathbf{x}.
\end{aligned}$$

Hence, on using that $z = 1/\sqrt{N}$, we obtain that

$$\begin{aligned}
& \int_{S^{n-1}} (p(\mathbf{x}, t + \Delta t) - p(\mathbf{x}, t)) g(\mathbf{x}, t) d\mathbf{x} \\
& = \frac{1}{2N} \int_{S^{n-1}} p(\mathbf{x}, t) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t) \right) d\mathbf{x} \\
& \quad + (\Delta t)^\nu \int_{S^{n-1}} p(\mathbf{x}, t) \left[\sum_{j=1}^{n-1} x_j (\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) \partial_j g(\mathbf{x}, t) \right] d\mathbf{x} + O(N^{-2}).
\end{aligned}$$

Remark 3. *If one is willing to assume differentiability of p with respect to t , then one may divide the last equation by Δt , and it is straightforward to show that, in the limit $\Delta t \rightarrow 0$, the left hand side converges to*

$$\int_{S^{n-1}} \partial_t p(\mathbf{x}, t) g(\mathbf{x}, t) d\mathbf{x}.$$

Then, on integrating in t and performing an integration by parts, one arrives at a weak formulation for the $\partial_t p$, which seems to be completely unnecessary in this case. Alternatively, we are left with the options of either obtaining a mixed strong-weak formulation, or by switching to a strong formulation in space, at the expense of losing the conservative property of the weak formulation. For a derivation without any regularity assumption, see appendix A.

On imposing the scaling relationship $N^{-1} = \kappa (\Delta t)^\mu$, we find that

$$\begin{aligned}
& \int_{S^{n-1}} \frac{p(\mathbf{x}, t + \Delta t) - p(\mathbf{x}, t)}{\Delta t} g(\mathbf{x}, t) d\mathbf{x} \\
& = \frac{(\Delta t)^{\mu-1}}{2} \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t) \right) d\mathbf{x} dt \\
& \quad + (\Delta t)^{\nu-1} \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left[\sum_{j=1}^{n-1} x_j (\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x})) \partial_j g(\mathbf{x}, t) \right] d\mathbf{x} dt + O(\Delta t).
\end{aligned}$$

With the above scalings, for $\mu < 1$ or $\nu < 1$, after multiplication by $(\Delta t)^{\min\{\mu, \nu\}-1}$, we obtain stationary approximations. Thus, in order to allow for time evolution, we need $\mu \geq 1$, $\nu \geq 1$. However, if both $\mu, \nu > 1$ we also have stationary solutions given by the initial condition.

The other cases are as follows:

$\mu > 1$, $\nu = 1$: This is the *convective or drift approximation*:

$$(11) \quad \begin{aligned} & - \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \partial_t g(\mathbf{x}, t) \, d\mathbf{x} \, dt - \int_{S^{n-1}} p(\mathbf{x}, t_0) g(\mathbf{x}, t_0) \, d\mathbf{x} \\ & = \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left[\sum_{j=1}^{n-1} x_j \left(\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \partial_j g(\mathbf{x}, t) \right] \, d\mathbf{x}. \end{aligned}$$

$\mu = 1$, $\nu > 1$: This is the *diffusive approximation*

$$(12) \quad \begin{aligned} & - \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \partial_t g(\mathbf{x}, t) \, d\mathbf{x} \, dt - \int_{S^{n-1}} p(\mathbf{x}, t_0) g(\mathbf{x}, t_0) \, d\mathbf{x} \\ & = \frac{\kappa}{2} \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t) \right) \, d\mathbf{x} \, dt. \end{aligned}$$

$\mu = 1$, $\nu = 1$: In this case there is a maximal balance of selection and genetic drift, and we find the *Replicator-diffusion equation*

$$(13) \quad \begin{aligned} & - \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \partial_t g(\mathbf{x}, t) \, d\mathbf{x} \, dt - \int_{S^{n-1}} p(\mathbf{x}, t_0) g(\mathbf{x}, t_0) \, d\mathbf{x} \\ & = \frac{\kappa}{2} \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t) \right) \, d\mathbf{x} \, dt \\ & \quad + \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left[\sum_{j=1}^{n-1} x_j \left(\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \partial_j g(\mathbf{x}, t) \right] \, d\mathbf{x}. \end{aligned}$$

Equations (11), (12) and (13) are written in the weak form. In population dynamics, and in others contexts as well, they are used casted into the strong formulation (or standard PDE formulation) as follows (see, however, remark 5):

- If $\mu > 1$ and $\nu = 1$, the *convective of drift approximation*:

$$(14) \quad \partial_t p = - \sum_{i=1}^{n-1} \partial_i \left[x_i \left(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) p \right].$$

This equation is equivalent to the replicator dynamics, showing that the Wright-Fisher process will be equivalent to the the replicator dynamics, in the limit of large population and small time-steps, if the population increases faster than the time-step decreases.

- If $\mu = 1$ and $\nu > 1$, the *diffusive approximation*

$$(15) \quad \partial_t p = \frac{\kappa}{2} \sum_{i,j=1}^{n-1} \partial_{ij} \left((x_i \delta_{ij} - x_i x_j) p \right),$$

which is relevant when the time-step decreases faster than the increase in population size.

- When there is a perfect balance between population size and time step, i.e., $\mu = \nu = 1$, we find the *replicator-diffusion approximation*, given by equation (1), which we repeat here for convenience:

$$(1') \quad \partial_t p = \frac{\kappa}{2} \sum_{i,j=1}^{n-1} \partial_{ij} \left((x_i \delta_{ij} - x_i x_j) p \right) - \sum_{i=1}^{n-1} \partial_i \left[x_i \left(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) p \right].$$

We shall focus on the last equation and on its weak formulation (13).

Remark 4. *The formal derivation carried out in this section used only general and abstract properties of the kernel $\Theta_{N,\Delta t}$. Therefore the thermodynamical limit of much more general discrete evolutionary processes but sharing the same asymptotic properties in the kernel will be the same.*

Remark 5. *We shall see below that the weak and the PDE formulations are not exactly equivalent, and that the more appropriate formulation is actually the weak one. Also, we have been quite cavalier about the nature of the test function. For now, it is sufficient to assume that g is sufficient regular so that we have all the differential expressions in the weak formulation continuous over the whole simplex S^{n-1} .*

4.2. Conservation laws from the discrete process. The conservation laws given by equation (7) now become

$$(16) \quad \frac{d}{dt} \int_{S^{n-1}} p(t, x) \varphi(x) dx = 0,$$

where φ satisfies

$$(17) \quad \frac{\kappa}{2} \sum_{i,j=1}^{n-1} D_{ij} \partial_{ij}^2 \varphi + \sum_{i=1}^{n-1} \Omega_i \partial_i \varphi = 0.,$$

Heuristically, this can be seen as follows: we define, in the weak sense,

$$\rho_i := \lim_{N \rightarrow \infty} \frac{F^{(i)}}{1/N}.$$

These functions are left eigenfunctions of the evolution operator, that is, they must be a solution of equation (17). The values of ρ_i on the vertexes are $\rho_i(\mathbf{e}_j) = \delta_{ij}$, and as we will shortly see, this is enough to properly define functions ρ_i . Then we impose that solutions of (1') should respect the following n conservation laws:

$$(18) \quad \frac{d}{dt} \int_{S^{n-1}} \rho_i(\mathbf{x}) p(\mathbf{x}, t) d\mathbf{x} = 0, \quad i = 1, \dots, n.$$

These functions can be obtained directly from the continuous problem. In particular, consider the stationary solutions of the adjoint equation (17). The limit of each of the vectors $NF^{(i)}$, $i = 1, \dots, n-1$, is a solution of the equation (17) such that $\rho_i(\mathbf{e}_j) = \delta_{ij}$. The function $\rho_i(\mathbf{x}_0)$ describes the final fixation probability of type i of a population initially in the state \mathbf{x}_0 . This can be seen by the following argument. We will prove in Theorem 2 that the final state is given by

$$p^\infty[p^1] = \lim_{t \rightarrow \infty} p(\cdot, t) = \sum_{i=1}^n \pi_i[p^1] \delta_{\mathbf{e}_i},$$

where $\delta_{\mathbf{e}_i}$ is a Dirac measure supported on the vertex $\mathbf{e}_i \in S_N^{n-1}$:

$$\int \delta_{\mathbf{e}_i}(\mathbf{x}) \phi(\mathbf{x}) d\mathbf{x} = \phi(\mathbf{e}_i).$$

Clearly, $\pi_i[p^1]$ is the fixation probability of type i in a population initially described by a probability distribution p^1 . Therefore,

$$\pi_i[\delta_{\mathbf{x}_0}] = \int \rho_i(\mathbf{x}) p^\infty(\mathbf{x}) d\mathbf{x} = \int \rho_i(\mathbf{x}) p^1(\mathbf{x}) d\mathbf{x} = \int \rho_i(\mathbf{x}) \delta_{\mathbf{x}_0}(\mathbf{x}) d\mathbf{x} = \rho_i(\mathbf{x}_0).$$

Finally, the limit of \mathbf{G} is the constant function; this happens because \mathbf{G} is the constant vector (the vector with the same value in all coordinates) and describes the conservation of probability at the discrete level. The constant vector $\rho_0(\mathbf{x}) = 1$ has the same role in the continuous case. It is clear the ρ_0 satisfies equation (17).

Remark 6. *In the neutral case, i.e., $\psi^{(i)}(\mathbf{x}) = \psi^{(j)}(\mathbf{x})$ for all $i, j = 1, \dots, n$ and $\mathbf{x} \in S^{n-1}$, we define the neutral fixation probability $\pi_i^N[\delta_{\mathbf{x}}] = x_i$, which follows from the fact that in the neutral case, $\rho_i(\mathbf{x}) = x_i$.*

Remark 7. *There is a subtle point in the derivation above, in the sense that (17) should have an infinite-dimensional kernel, thus implying an infinite number of linear conservation laws, while the discrete model yields only n linearly independent such laws. We shall see in a forthcoming work that with an appropriate definition of sufficient regular solutions, (17) has an n dimensional kernel. See (Chalub and Souza, 2011c) for further details.*

5. THE REPLICATOR-DIFFUSION APPROXIMATION

We now discuss the nature of solutions p to (1') together with the conservation laws (18). The main result of this section is theorem 2. This must be understood as the continuous counterpart of the corollary 1. We do not refer to the continuous model to prove this result. Our approach is based solely in the properties of the partial differential equation (1'), the restriction of the domain to the domain of interest, and the associated conservation laws (18). The complete proof is more technical and will be postponed to a different result. We will, however, present the main features of the result.

An outline of the proof of theorem 2 is as follows:

- (1) p can be divided in a smooth part (classical solution) defined in the simplex, plus a singular part supported on the boundary; (the classical solution was previously studied, for the case $n = 2$ and constant fitnesses functions in Feller (1951);)
- (2) the classical solution satisfies the equation (1') but not the conservation laws, therefore it cannot be the correct limit of the discrete process;
- (3) the classical solution vanishes with time, and therefore after a long time the (non-classical) solution will be supported on the boundary of the simplex; the solution will indicate, therefore, one extinction;
- (4) if we consider the evolution with one type less (i.e., the evolution on the faces of the simplex S^{n-1} , which are (homotopical to) simplexes S^{n-2} all the ideas above apply again and after a long time the solution will be supported on the faces of the faces of the simplex;

After $n - 2$ applications of the above technique, we conclude our main theorem which states that the solution of the replicator-diffusion equation converges when $t \rightarrow \infty$ to a linear combination of Dirac deltas supported on the vertexes of the simplex. For the reader uninterested in the details, this is the essence of the first section.

We finally provide two applications. In subsection 5.2, we study the dual equation. This will be the continuous limit of the evolution by the dual equation (backward equation) of the discrete process and therefore its solution $f(\mathbf{k}, t)$ gives the fixation probability of a given type (to be prescribed by the boundary conditions in the dual process) for a population initially at state \mathbf{k} at time t . This gives a generalization for an arbitrary number of types and for arbitrary fitnesses of the celebrated Kimura equation with reversed time (Kimura, 1962). In the sequel, subsection 5.3, we will show that if one type dominates all other types then, for any initial condition, the fixation probability of this type will be larger than the neutral fixation probability. This shows, in particular, that for large populations, the most probable type to fixate will be the one playing the Nash-equilibrium strategy of the game (assuming the identity between fitness and pay-offs, which is standard in this framework). This is not true in general for small populations (Nowak, 2006).

5.1. Solution of the replicator-diffusion equation. We now study in more detail the features of the solution to (13) and show two important results: first that in the interior of the simplex, the solution must satisfy (1') in the classical sense; second, no classical solution to (1') can satisfy the conservation laws.

Lemma 5. *Let p be a solution to (13). Let $K \subset S^{n-1}$ be a proper compact subset. Then, in K , p satisfies (1') in the classical sense. In particular, p is $C^\infty(K)$.*

Proof. Let $g \in C_c^\infty(K)$, we have then the standard weak formulation of (1') in K . On the other hand, (1') is uniformly parabolic in any proper subset. Hence the weak and strong formulations coincide—c.f. (Evans, 2010; Taylor, 1996). \square \square

We now obtain some more information about this solution on S^{n-2} .

Lemma 6. *Let p be a classical solution to (1') over the interior of S^{n-1} . Then*

$$\lim_{t \rightarrow \infty} p(\mathbf{x}, t) = 0, \quad \mathbf{x} \in K.$$

Proof. We write the drift part as $\Omega = \frac{1}{2}\nabla\phi + \mathbf{b}$, define $\mu(\mathbf{x}) = x_1x_2 \cdots x_n$ (such that $\mu(\mathbf{x}) \geq 0$ in S^{n-1} with $\mu = 0$ if and only if $x \in \partial S^{n-1}$). In the new variable $u = \mu e^{-\phi} p$ and after some manipulations, we find

$$\begin{aligned} \partial_t u &= \omega^{-1} \sum_{i=1}^{n-1} \partial_i \left[\omega \left(\frac{1}{2} \sum_{j=1}^{n-1} D_{ij} \partial_j u - x_i \left(\beta^{(i)} - \sum_{j=1}^{n-1} x_j \beta^{(j)} \right) \right) u \right] \\ (19) \quad &= \frac{1}{\omega} \nabla \cdot \left[\omega \left(\frac{1}{2} D \nabla u - \mathbf{B} u \right) \right], \end{aligned}$$

with $\omega = e^\phi / \mu$ and $B_i = x_i \left(b^i - \sum_{j=1}^{n-1} x_j b^j \right)$.

We shall now study the eigenvalue problem

$$(20) \quad \nabla \cdot \left[\omega \left(\frac{1}{2} D \nabla \varphi + s \Phi \varphi \right) \right] = \lambda \omega \varphi, \quad \varphi = 0 \text{ in } \partial S^n.$$

We first prove that $\lambda = 0$ cannot be an eigenvalue of the operator in the left hand side of equation (20). Consider any proper subset $V \subset S^{n-1}$ and apply the maximum principle to equation (20); we conclude that the maximum of φ occurs at the boundary. But as ∂V approaches ∂S^{n-1} , this maximum can be made arbitrarily small; we conclude by continuity that $\varphi = 0$.

When $s = 0$, the left hand side of equation (20) is clearly negative-definite. Since the eigenvalues must be continuous functions of s , we have the (20) is negative for any finite s . Moreover, this also shows that there exists $\alpha > 0$, such that

$$\frac{1}{2} \partial_t \int_{S^{n-1}} u^2 \omega dV = \int_{S^{n-1}} \nabla \cdot \left[\omega \left(\frac{1}{2} D \nabla u - \mathbf{B} u \right) \right] u dV < -\alpha \int_{S^{n-1}} u^2 \omega dV.$$

Therefore

$$\int p^2 e^{-\phi} \mu dx = \int u^2 \omega dx \xrightarrow{t \rightarrow \infty} 0,$$

□

□

Lemma 7. *Equation (19) has a unique solution $u \in C(S^{n-1}) \cap C^\infty(\text{int } S^{n-1})$*

Proof. Let

$$D_{i,j}^\varepsilon = D_{i,j} + \varepsilon, \quad \text{and} \quad \nu_\varepsilon = \nu + \varepsilon.$$

For $\varepsilon > 0$, (19) is uniformly parabolic, and hence it has a unique solution with the required regularity.

We write (19) in weak form as

$$\begin{aligned} &\int_0^\infty \int_{S^{n-1}} u(t, x) \frac{\partial_t \phi(t, x)}{\omega} dx dt + \\ &\int_0^\infty \int_{S^{n-1}} \omega \left(\frac{1}{2} D \nabla u - \mathbf{B} u \right) \cdot \nabla \phi(t, x) dx dt + \int_{S^{n-1}} u(0, x) \phi(0, x) dx dt. \end{aligned}$$

Notice that $u_\varepsilon \in W_0^{2,1}(S^{n-1})$, hence by Rellich Theorem (Evans, 2010), we can select $\varepsilon_k \rightarrow 0$ such that $u_{\varepsilon_k} \rightarrow u^*$ in $L^2(S^{n-1})$. This allows us to pass the limit in the last integral. For the first integral, S^{n-1} is bounded, we have that $L^2(S^{n-1}) \subset L^1(S^{n-1})$, hence u_{ε_k} converges in $L^1(S^{n-1})$. Thus the first integral converges, by the monotone convergence theorem.

The remaining integral can be seen to converge by compactness of Sobolev inclusions.

Finally, the maximum principle can be used to show uniqueness, and regularity follows from analogous arguments. □ □

This last theorem has an important consequence

Corollary 2. *No solution to (1') in the classical sense can satisfy the required conservation laws.*

Proof. Since (1') is uniformly parabolic for any proper set of S^{n-1} , it is possible to show that $p \in C(S^{n-1})$. Given $\varepsilon > 0$, choose $K \subsetneq S^{n-1}$ such that the $\mu(S^{n-1} - K) < \varepsilon$, μ being the Lebesgue measure in S^{n-1} . Also, let $p^0 \in \text{BM}^+(S^{n-1})$ and ψ be a conservation law. Then

$$\int_{S^{n-1}} p^0(\mathbf{x})\psi(\mathbf{x}) \, d\mathbf{x} = \alpha > 0,$$

and

$$\begin{aligned} \int_{S^{n-1}} p(\mathbf{x}, t)\psi(\mathbf{x}) \, d\mathbf{x} &= \int_{S^{n-1}-K} p(\mathbf{x}, t)\psi(\mathbf{x}) \, d\mathbf{x} + \int_K p(\mathbf{x}, t)\psi(\mathbf{x}) \, d\mathbf{x} \\ &\leq C\varepsilon + \int_K p(\mathbf{x}, t)\psi(\mathbf{x}) \, d\mathbf{x} \leq C\varepsilon, \end{aligned}$$

for sufficient large t . Hence there is no conservation. □ □

Theorem 2 (Final State). *Assume that*

$$p^\infty(\mathbf{x}) = \lim_{t \rightarrow \infty} p(\mathbf{x}, t)$$

exists. Then p^∞ is a linear combination of point masses at the vertexes of S^{n-1} , i.e.,

$$(21) \quad p^\infty = \sum_{i=1}^n \pi_i [p^I] \delta_{\mathbf{e}_i} .$$

Proof. By Lemma 6, p^∞ has support only in the boundary of S_N^{n-1} . If we look at the reduced problem at each face and solve the corresponding problem with p^∞ , appropriately restricted, as the initial condition, we then see that p^∞ cannot have support at the faces either. Proceeding inductively, we see that p^∞ can possibly have support only at the vertexes. Because of probability conservation, p^∞ actually is composed only of point masses. □ □

5.2. Duality and the Kimura equation. The formal adjoint of equation (1) (changing the flow of time from forward to backward) provides a generalization of the celebrated Kimura equation (Kimura, 1962), both including more types and allowing frequency dependent fitness:

$$(22) \quad \partial_t f = \mathcal{L}_{n-1, k}^\dagger f := \frac{\kappa}{2} \sum_{i, j=1}^{n-1} D_{ij} \partial_{ij}^2 f + \sum_{i=1}^{n-1} \Omega_i \partial_i f .$$

In diffusion theory this equation is associated with a martingale problem for the diffusive continuous process. In genetics, the meaning of equation (22) is seldom made clear and depends on the boundary conditions imposed. One possible and common interpretation is as follows: given an homogeneous state $\mathbf{e}_i \in \Delta S^{n-1}$, let $f_i(\mathbf{k}, t)$ be the probability that given a population initially in a well-defined state $\mathbf{k} \in S^{n-1}$ (i.e., $p^I(\mathbf{x}) := p(\mathbf{x}, 0) = \delta_{\mathbf{k}}(\mathbf{x})$) we find the population fixed at the homogeneous state \mathbf{e}_i at time t (or before), i.e., $f_i(\mathbf{k}, t)$ is the probability of having $p(\mathbf{x}, t) = \delta_{\mathbf{e}_i}(\mathbf{x})$ from the initial condition $\delta_{\mathbf{k}}(\mathbf{x})$, given by $\langle p(\cdot, t), \delta_{\mathbf{e}_i} \rangle$. In this case, we need to find consistent boundary conditions. See Maruyama (1977).

This follows from the interpretation of the discrete adjoint evolution:

$$Q(\mathbf{k}, t + \Delta t) = \sum_{\mathbf{k}' \in S_N^{n-1}} \Theta_{N, \Delta t}(\mathbf{k} \rightarrow \mathbf{k}') Q(\mathbf{k}', t) ,$$

which reads as follows: “the fixation probability after a time interval $t + \Delta t$ of a given type, for a population initially at state \mathbf{k} is equal to the sum over all possible states \mathbf{k}' of the fixation probability after a time interval t of a population initially at state \mathbf{k}' times the transition probability from \mathbf{k} to \mathbf{k}' ”.

Let us study the fixation of type 1, represented by the state \mathbf{e}_1 . Let us now call V_i the face of the simplex with $x_i = 0$ (type i is absent). Then, $f_i|_{V_1} = 0$. For $i \neq 1$, $f_i|_{V_i}$ is the solution of $\partial_t f = \mathcal{L}_{n-2, k}^\dagger f$, where the type i was omitted from the equation. As the faces of the simplex are invariant under the adjoint evolution (one more fact to be attributed to lack of mutations in the model), this represent the same problem in one dimension less. We continue this procedure until

we find the evolution in the edge from vertex 1 to vertex $i \neq 1$, L_{1i} . In this case, we have that $f|_{L_{1i}} : [0, 1] \rightarrow \mathbb{R}$, the restriction of f_i to this edge, with k the fraction of type 1 individuals, is the solution of

$$(23) \quad \partial_t f = \frac{\kappa}{2} k(1-k) \partial_k^2 f + k(1-k) \left(\psi^{(1)}|_{L_{1i}}(k) - \psi^{(i)}|_{L_{1i}}(k) \right) \partial_k f$$

with boundary conditions given by $f(0) = 0$ and $f(1) = 1$ and $\psi^{(j)}|_{L_{1i}}$ is the restriction of $\psi^{(j)}$ to the edge L_{1i} . The forward and backward versions of Equation (23) are fully studied in the references (Chalub and Souza, 2009a,b). For $\psi^{(1)}|_{L_{1i}} - \psi^{(i)}|_{L_{1i}}$ constant this is the Kimura equation.

5.3. Strategy dominance. Let us assume that $\psi^{(1)}(\mathbf{x}) \geq \psi^{(i)}(\mathbf{x})$ for all $\mathbf{x} \in S^{n-1}$. This happens, for example, if we identify fitness functions with pay-offs in game theory, types with strategists, and if strategist 1 plays the Nash-equilibrium strategy.

Therefore, we prove

Theorem 3. *If, for all states $\mathbf{x} \in S^{n-1}$, and all types $i = 1, \dots, n$, $\psi^{(1)}(\mathbf{x}) \geq \psi^{(i)}(\mathbf{x})$, then the fixation probability of the first type is not less than the neutral fixation probability for any initial condition p^1 ; i.e.,*

$$\pi_1[p^1] \geq \pi_1^N[p^1].$$

Proof. First note that it is enough to prove that $\pi_1[\delta_{\mathbf{x}}] \geq \pi_1^N[\delta_{\mathbf{x}}] = x_1$ for all $\mathbf{x} \in S^{n-1}$. The difference $\rho_1(\mathbf{x}) - x_1$ satisfy

$$\frac{\kappa}{2} \sum_{i,j=1}^{n-1} D_{ij} \partial_{ij}^2 (\rho_1(\mathbf{x}) - x_1) + \sum_{i=1}^{n-1} \Omega_i \partial_i (\rho_1(\mathbf{x}) - x_1) = -\Omega_1 = -x_1 \left(\psi^{(1)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \leq 0,$$

with vertex conditions $\rho_1(\mathbf{e}_i) - x_1(\mathbf{e}_i) = 0$ for $i = 1, \dots, n$. Now, we proceed by induction in n . For the case $n = 2$, the proof is in (Chalub and Souza, 2009b, section 4.3); we reproduce it here only for completeness.

We write explicitly the equation for ρ_1 :

$$\frac{\kappa}{2} x(1-x) \partial_x^2 \rho_1 + x \left(\psi^{(i)}(x) - \bar{\psi}(x) \right) \partial_x \rho_1 = 0$$

with $\rho_1(0) = 0$ and $\rho_1(1) = 1$. We simplify the equation using the fact that $\psi^{(1)}(x) - \bar{\psi}(x) = (1-x) (\psi^{(1)}(x) - \psi^{(2)}(x))$ and the solution is given by

$$\rho_1(x) = \frac{\int_0^x \exp \left[-\frac{2}{\varepsilon} \int_0^{\bar{x}} (\psi^{(1)}(\bar{x}) - \psi^{(2)}(\bar{x})) d\bar{x} \right] d\bar{x}}{\int_0^1 \exp \left[-\frac{2}{\varepsilon} \int_0^{\bar{x}} (\psi^{(1)}(\bar{x}) - \psi^{(2)}(\bar{x})) d\bar{x} \right] d\bar{x}}.$$

As $\psi^{(1)}(x) \geq \psi^{(2)}(x)$, we conclude that

$$\begin{aligned} & \frac{1}{x} \int_0^x \exp \left[-\frac{2}{\varepsilon} \int_0^{\bar{x}} (\psi^{(1)}(\bar{x}) - \psi^{(2)}(\bar{x})) d\bar{x} \right] d\bar{x} \\ & \geq \int_0^1 \exp \left[-\frac{2}{\varepsilon} \int_0^{\bar{x}} (\psi^{(1)}(\bar{x}) - \psi^{(2)}(\bar{x})) d\bar{x} \right] d\bar{x}. \end{aligned}$$

In particular, $\rho_1(x) \geq x$.

Now, assume that $\rho_1(\mathbf{x}) - x_1 \geq 0$ for all $\mathbf{x} \in \partial S^{n-1}$. (Note that ∂S^{n-1} is an union of a finite number of $n-2$ dimensional simplexes, where by the principle of induction we assume the result valid.) Finally, we use the maximum principle for subharmonic functions to conclude that the minimum cannot be in the interior of the simplex (Courant and Hilbert, 1989). Therefore $\rho_1(\mathbf{x}) \geq x_1$ for all $\mathbf{x} \in S^{n-1}$. \square

6. THE REPLICATOR DYNAMICS

We showed in previous sections that the evolution of a large, but finite, population can be well approximated in all time scales by a partial differential equation of drift-diffusion type. Now, we show that locally in time, the dynamics can be approximated (in the DCA sense) by the much simpler equation (14). We start in subsection 6.1 showing that the equation (14) is formally equivalent to the replicator system. Afterwards, in subsection 6.2 we answer a central question: what exactly is the replicator equation modelling? In particular, we will show, using a simple argument, that the replicator equation does not model the evolution of the expected value (of a given trait) in the population, but the evolution of the most common trait (the mode of the probability distribution). Finally, we show, in subsection 6.3 that the replicator ordinary differential equation is a good approximation for the initial dynamics of the Wright-Fisher process, when the population is large.

6.1. The replicator ODE and PDE. We shall now study in more detail the equation (14), which has a close connection with the replicator dynamics as shown below:

Theorem 4. *Let $\Phi_t(\mathbf{x})$ the flow map of*

$$(24) \quad \frac{d\mathbf{x}}{dt} = \Omega(\mathbf{x}(t)).$$

and let

$$Q(\mathbf{x}, t) = - \int_0^t (\nabla \cdot \Omega)(\Phi_{s-t}(\mathbf{x})) ds.$$

then the solution to (14) with a C^1 initial condition p_0 is given by

$$(25) \quad p(\mathbf{x}, t) = e^{Q(\mathbf{x}, t)} p_0(\Phi_{-t}(\mathbf{x})).$$

Proof. Clearly $Q(\mathbf{x}, 0) = 0$, and $\Phi_0(\mathbf{x}) = \mathbf{x}$. Hence the initial condition is satisfied.

Let

$$R(\mathbf{z}, t) := e^{-Q(\Phi_t(\mathbf{z}), t)} p(t, \Phi_t(\mathbf{z}))$$

On one hand, (25) shows that

$$R(\mathbf{z}, t) = p_0(\mathbf{z})$$

Therefore

$$\frac{dR}{dt}(\mathbf{z}, t) = 0.$$

On the other hand, one can compute

$$\frac{dR}{dt}(\mathbf{z}, t) = e^{-Q(\Phi_t(\mathbf{z}), t)} (\partial_t p + \nabla p \cdot \dot{\mathbf{x}} + (\nabla \cdot \Omega) p) = e^{-Q(\Phi_t(\mathbf{z}), t)} (\partial_t p + \nabla \cdot (p\Omega)).$$

We then conclude that $p(\mathbf{x}, t)$ is the solution of equation (14). □ □

6.2. Peak and average dynamics. We start by showing that the dynamics of the average in the Wright-Fisher process, even in the thermodynamical limit, is not governed by the replicator equation. Consider for example, a population of n types, evolving according to the replicator-diffusion equation with fitness functions given by $\psi^{(i)} : S^{n-1} \rightarrow \mathbb{R}_+$.

From the fact that the final state of the replicator-diffusion equation is given by equation (21), the coefficients $\pi_i[p^I]$, $i = 1, \dots, n$ can be calculated in two ways:

$$\int \rho_i(\mathbf{x}) p^I(\mathbf{x}) d\mathbf{x} = \int \rho_i(\mathbf{x}) p^\infty(\mathbf{x}) d\mathbf{x} = \pi_i[p^I] = \int x_i p^\infty(\mathbf{x}) d\mathbf{x} =: \langle p^\infty \rangle_i.$$

Therefore the average of the probability distribution will converge to a certain point of the simplex depending on the initial condition. This is completely different from the replicator dynamics, as its solution converges to a single attractor, periodic orbits, chaotic attractors, etc (Hofbauer and Sigmund, 1998).

Now, we show that the probability distribution concentrates in the ESS; this shows that the peak will behave in manner similar to the solutions of the replicator dynamics.

Recall that (Hofbauer and Sigmund, 1998), that an ESS that lies in interior of S^{n-1} must be a globally stable point of (24). We have then the following result

Theorem 5. *Assume $p^I(\mathbf{x})$ is smooth, and assume that (24) has an ESS \mathbf{x}^* . Then the solution of equation (14) is such that*

$$\lim_{t \rightarrow \infty} p(t, \mathbf{x}) = \delta_{\mathbf{x}^*}.$$

Proof. Since \mathbf{x}^* is a globally stable equilibrium, for any given $\delta > 0$, we can find $T > 0$, such that, for $t > T$, we have that for any proper compact subset $K \subset S^{n-1}$.

$$\Phi_t(K) \subset B_\delta(\mathbf{x}^*).$$

Let $\psi(\mathbf{x})$ be a continuous function with support contained in K . Then, for $t > T$, we have that

$$\int_{S^{n-1}} p(t, \mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x} = \int_{B_\delta(\mathbf{x}_1)} p(t, \mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x}.$$

But, let $\epsilon > 0$ be given. Since ψ is continuous, possibly with a smaller $\delta > 0$, we must have

$$(26) \quad \psi(\mathbf{x}^*) - \epsilon \leq \int_{B_\delta(\mathbf{x}_1)} p(t, \mathbf{x}) \psi(\mathbf{x}) \, d\mathbf{x} \leq \psi(\mathbf{x}^*) + \epsilon,$$

Now take $(\delta_k, \epsilon_k) \downarrow 0$ such that (26) is satisfied. This yields a sequence of times T_k such that $T_k \rightarrow \infty$ and

$$\lim_{k \rightarrow \infty} \int_{S^{n-1}} p(T_k, \mathbf{x}) \, d\mathbf{x} = \psi(\mathbf{x}^*).$$

Since $\Phi_s(K) \subset \Phi_t(K)$, for $s > t$, the claim follows. \square \square

The previous result shows that, for sufficient large time, the support of the solution of the replicator PDE, equation (14), will be concentrated in sufficient small neighborhoods of \mathbf{x}^* . In particular, this will be true for the maximum. For the replicator-diffusion equation (1) this cannot be valid for any value of $\kappa > 0$ (as it was proved in theorem 2); however, the initial dynamics given by the replicator diffusion equation is similar to the one given by the replicator ODE; in the language of the DCA, we say that the replicator ODE approximate the initial dynamics of the replicator-diffusion equation (and, by extension, the discrete Wright-Fisher process). This is justified by theorem 6. We start with:

Lemma 8. *Let $\mathbf{x}_* \in S^{n-1}$ be the only ω -limit of the replicator dynamics (24). Then, there exists a (sufficiently small) set $V \in S^{n-1}$ such that $\mathbf{x}_* \in V$ and*

$$\int_{\partial V} \Omega \cdot \hat{n} \, dS < 0,$$

where ∂V is the boundary of V , \hat{n} its exterior normal and dS the surface element.

Proof. We divide the proof in two cases: when $\mathbf{x}_* \in \text{int } S^{n-1}$ and when $\mathbf{x}_* \in \partial S^{n-1}$. For the first case, the field vector Ω points to \mathbf{x}_* for points close enough; for the second case, it is clear that $\Omega \cdot \hat{n}|_{\partial S^{n-1}} = 0$ and the result follows considering the decomposition $V = (V \cap \text{int } S^{n-1}) \cup (V \cap \partial S^{n-1})$. \square \square

Theorem 6. *Let $\mathbf{x}_* \in S^{n-1}$ be the only ω -limit of the replicator dynamics (24), and consider a constant initial condition p^I . Let p_κ be the solution of equation (1'). Then there is a time $t_* > 0$ and $V \ni \mathbf{x}^*$ such that*

$$\partial_t \int_V p_\kappa \, d\mathbf{x} > 0, \quad \forall t < t_*.$$

Proof. Note that

$$\partial_t p_0 = -\nabla \cdot \Omega p_0 - \Omega \cdot \nabla p_0.$$

Initially, $\nabla p^I = 0$ and thus with $V \ni \mathbf{x}_*$ sufficiently small, we have from lemma 8

$$(27) \quad \lim_{t \rightarrow 0} \partial_t \int_V p_0 \, d\mathbf{x} = - \int_V \nabla \cdot \Omega p^I \, d\mathbf{x} = -p^I \int_{\partial V} \Omega \cdot \hat{n} \, dS > 0.$$

On the other hand, p_κ converges locally in time to p_0 weakly, and therefore

$$\lim_{t \rightarrow 0} \partial_t \int_V p_\kappa d\mathbf{x} > 0$$

for $\kappa < \kappa_0 > 0$. From the conservation of probability, it is clear that

$$\lim_{t \rightarrow 0} \partial_t \int_{S^{n-1} \setminus V} p_\kappa d\mathbf{x} < 0 .$$

We conclude that for an initial constant distribution, p_κ will concentrate initially near the ω -limits of the replicator dynamics. □ □

Remark 8. *If, instead of a uniform distribution, we consider a distribution with a peak and symmetric around the peak, we will have, for small enough V*

$$\partial_t \int_V p_0 d\mathbf{x} \approx - \int_V \nabla \cdot \Omega p_0 d\mathbf{x} = -p_0|_{\partial V} \int_{\partial V} \Omega \cdot \hat{n} dS > 0 ,$$

and, following the same arguments as above, the peak will move from its initial location to \mathbf{x}_* . This movement will be along the lines that maximize $\int_{\partial V} \Omega \cdot \hat{n} dS$ for small V . Obviously, a small asymmetry around the peak will not be able to invert the previous inequality.

6.3. Local in time convergence. Let

$$0 < \varepsilon \ll 1.$$

For sake of simplicity, we shall consider uniform strong selection, i.e

$$(28) \quad \tilde{\psi}^{(i)} = \varepsilon^{-1} \psi^{(i)},$$

and a fast time scale

$$(29) \quad \tilde{t} = \varepsilon \tau.$$

If we perform a regular asymptotic expansion, i.e., if we write $p_\varepsilon \approx p^{(0)} + \varepsilon p^{(1)} + \dots$, then we find, for times $\tau \ll \varepsilon^{-1}$, that the leading order dynamics is given by

$$(30) \quad \partial_\tau p + \nabla \cdot (p\Omega) = 0.$$

This asymptotic Ansatz can be made rigorous by a recast of a similar result for the Moran dynamics in Chalub and Souza (2011a)

Theorem 7. *Assume that the fitness are $C^2(S^{n-1})$ functions, and that the initial condition p^I is also $C^2(S^{n-1})$. Let p_ε be the solution of (1), with $\varepsilon \geq 0$. Then p_0 is $C^2(S^{n-1})$, and satisfies the conservation law (18). Moreover, given ε and K positive, there exists a C such that, for $\tau \leq C$, we have*

$$\|p_\varepsilon(\tau, \cdot) - p_0(\tau, \cdot)\|_\infty \leq C\varepsilon$$

and

$$\|\partial_x^2 p_0(\tau, \cdot)\|_\infty < K$$

Thus p_0 is the leading order asymptotic approximation to p_ε , for $t < \varepsilon C$.

Proof. The statements about p_0 follows straightforward by obtaining the solution by the method of characteristics.

Let $w_\varepsilon = p_\varepsilon - p_0$. Then w_ε satisfies

$$\partial_\tau w_\varepsilon = \frac{\varepsilon}{2} \sum_{i=1}^{n-1} \partial_i^2 (x_i w_\varepsilon) - \frac{\varepsilon}{2} \sum_{i,j=1}^{n-1} \partial_{ij}^2 (x_i x_j w_\varepsilon) - \sum_{i=1}^{n-1} \partial_i \left(x_i \left(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) w_\varepsilon \right) + \frac{\varepsilon}{2} g_0(\tau, x)$$

with null initial condition, where

$$g_0(\tau, x) = \sum_{i=1}^{n-1} \partial_i^2 (x_i p_0) - \sum_{i,j=1}^{n-1} \partial_{ij}^2 (x_i x_j p_0).$$

Notice that, because of the assumptions on p_0 , we have that g_0 is uniformly bounded in time.

The solution for such a problem is given by Duhammel principle. Let $S(t, t_0)$ be associated solution operator. We have that

$$w_\varepsilon(\tau, \mathbf{x}) = \frac{\kappa}{2} \int_0^\tau S(\tau, s) g_0(s, x) ds.$$

By the maximum principle applied to the semigroup $S(t_2, t_1)$, we have that $\|S(\tau, s) g_0(s, x)\| \leq M_s$, and by the uniform bound on g_0 , we have that there exists a constant M such that $M_s \leq M$. Thus, we find that

$$\|S(\tau, s) g_0(s, x)\|_\infty \leq M.$$

Hence, for $\tau < C$, we have a constant M' such that

$$\|w_\varepsilon(\tau, \cdot)\|_\infty \leq \varepsilon M'.$$

□

□

Remark 9. *It should be also possible to have asymptotic results valid for $\tau \sim \varepsilon^{-1}$, but this will be discussed elsewhere. Nevertheless, we notice that for the Moran process such results can be found in Chalub and Souza (2011a).*

7. NUMERICAL RESULTS

We show, in this section, numerical results for two variants of the Rock-Scissor-Paper game (Hofbauer and Sigmund, 1998); i.e., fitness are identified with the pay-off from game theory. In subsection 7.1, we study the evolution of the discrete evolution numerically in time, and show that the peak of distribution behaves accordingly to the replicator equation while the average value of the same distribution converges to a point which is not the ESS. In subsection 7.2 we obtain explicitly the fixation probability of a given type for the symmetric Rock-Scissor-Paper game. A full animation is available in the website indicated in the caption of figure 2.

7.1. Forward equation. We use evolutionary game theory (Smith, 1982; Hofbauer and Sigmund, 1998) to define the fitness function. More precisely, we define a pay-off matrix $\mathbf{M} = (M_{ij})_{i,j=1,\dots,n}$ such that M_{ij} is the gain (in fitness) of the i type against the j type. The fitness of the i type in a population at state \mathbf{x} is

$$(31) \quad \psi^{(i)}(\mathbf{x}) = \sum_{j=1}^n M_{ij} x_j = (\mathbf{M}\mathbf{x})_j .$$

The replicator dynamics is given by the system of differential equation $\dot{x}_i = x_i(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}))$, where $\bar{\psi}(\mathbf{x}) = \mathbf{x} \cdot \mathbf{M}\mathbf{x}$.

We consider in Figure 2 the evolution of a discrete population of $N = 150$ individuals with the pay-off matrix given by

$$(32) \quad \mathbf{M} = \begin{pmatrix} 30 & 81 & 29 \\ 6 & 30 & 104 \\ 106 & 4 & 30 \end{pmatrix} .$$

This is know as the generalized Rock-Scissor-Paper game and presents an evolutionary stable state (ESS) $(x^*, y^*, z^*) = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Furthermore, the flow of the replicator dynamics converges in spirals to the ESS. The vertexes as well as (x^*, y^*, z^*) are equilibrium points for the continuum dynamics. See Hofbauer and Sigmund (1998) for the choice of values of the matrix \mathbf{M} .

Note that the peak moves in inward spirals around the central equilibrium, following the trajectories of the replicator dynamics, while all the mass diffuses to the boundary.

The green spot indicates the average value for x and y ; at first it moves in spirals close to the trajectories of the replicator dynamics. After a time depending on the value of N it starts to move in the direction of its final point $(x^\infty, y^\infty, z^\infty) = (\pi_1[p^I], \pi_2[p^I], \pi_3[p^I])$. This point can be calculated using equation (21) and the $n = 3$ independent conservation laws. Effectively, let x_i denote a given vertex of the simplex

$$\lim_{t \rightarrow \infty} \langle x_i \rangle(t) = \lim_{t \rightarrow \infty} \int x_i p(t, x) dx = \pi_i[p^I] ,$$

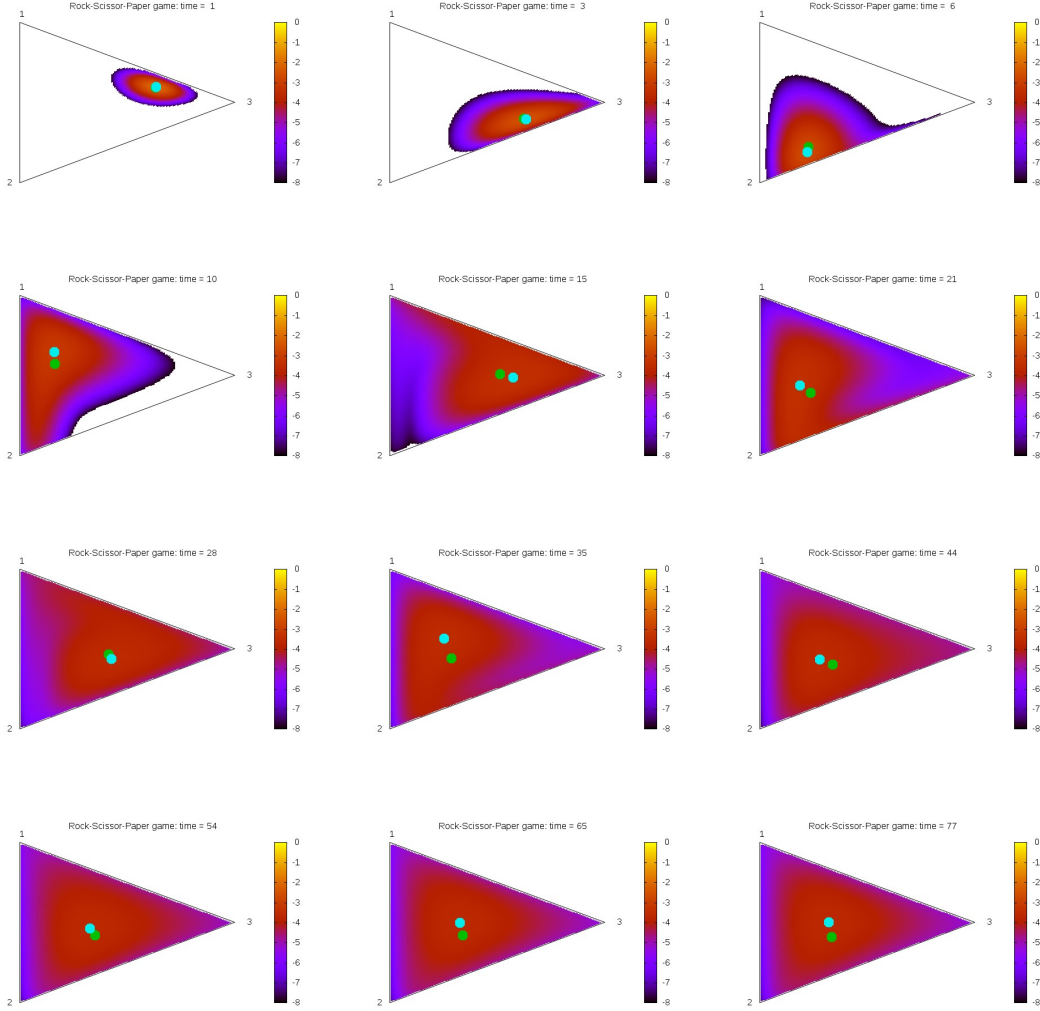


FIGURE 2. Solution for short times (1,3,6,10,15,21,28,35,44,54,65,77) of the Wright-Fisher evolution for a population of 150 individuals of two given types, with fitness given by equations (31) and (32) for a distribution initially concentrated in an interior non-stationary point. The value of the distribution $P(x, y, t)$ is in logarithmic scale. Note that the cyan spot, marking the interior peak of the probability distribution rotates and converges to the ESS $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$ (along characteristics of the PDE or, equivalently, the trajectories of the replicator dynamics). At the same time, the green spot marks the mean value of the probability distribution and also rotates initially. After a long time, it moves toward its final position, given by $\mathbf{x}^\infty := (c_1[p^I], c_2[p^I], 1 - c_1[p^I] - c_2[p^I]) \approx (0.331, 0.227, 0.442)$. For a full animation, also for different population sizes N , see <http://d1.dropbox.com/u/11325424/WFsim/RSPFinal.html>

where $\pi_i[p^I]$ is the fixation probability of type i associated to the initial condition p^I .

7.2. Backward equation and the decay of the interior L^1 -norm. The stationary state of the backward equation (22) represents the fixation of probability of a given type. This type is specified by the associated boundary conditions. Let us consider, as an example, that $n = 3$, the

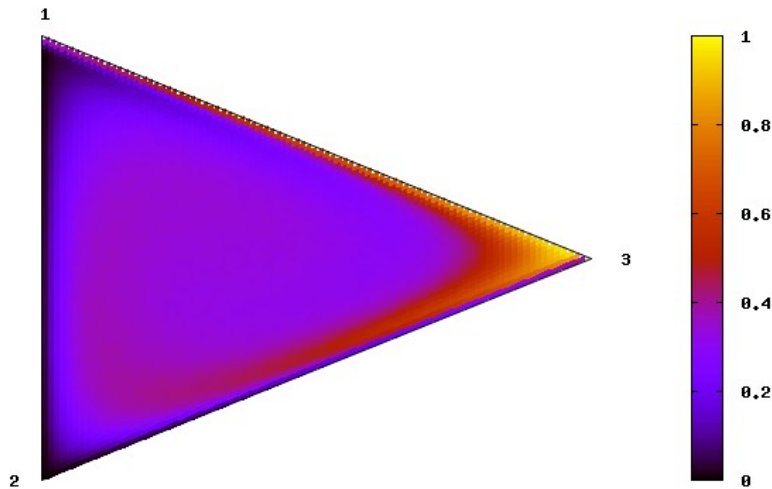


FIGURE 3. Fixation probability of the third type, in a Rock-Scissor-Paper game. This is the numerical solution of the stationary state of the equation 22, simulated by a Wright-Fisher process with $N = 150$ and pay-off matrix $([[20, 0, 40], [40, 20, 0], [0, 40, 20]])$. Note that higher values of the fixation probability “rotates” around the center of the simplex (the stationary state of the replicator dynamics).

evolution is given by the Rock-Scissor-Paper game defined by the matrix

$$(33) \quad \mathbf{M} = \begin{pmatrix} 0 & 40 & 20 \\ 20 & 0 & 40 \\ 40 & 20 & 0 \end{pmatrix},$$

and we study the fixation probability of the third type. An exact solution is difficult to obtain, as it would be necessary to solve an hierarchy of equations, representing boundary conditions of larger sets; however, a numerical solution is extremely easy to compute, as the Wright-Fisher process is a natural discretization of the (forward as well as the) backward equation. This is probably computationally inefficient, and different processes can be compatible with the same limit equations. See figure 3 for an illustration.

In figure 4, we plot the L^1 norm in the interior of the simplex and all subsimplexes, showing that that the probability mass flows from the simplex S^{n-1} to the faces (which are equivalent to the simplexes S^{n-2}); the solution behaves on the faces as the solution of the replicator-diffusion problem with one dimension less. The probability flows to the “faces of the faces”, i.e., to simplexes S^{n-3} until it reaches the absorbing state \mathbf{e}_i (simplexes S^0) for $i = 1, \dots, n$. We may think in a stochastic process reaching and sticking to the faces of the simplex until they reach their final spot, the vertexes.

8. CONCLUSIONS

The use of ordinary differential equations in population dynamics is widespread. However, as they are valid only for infinite populations, and real populations are always finite, the precise justification of its use and the precise meaning of its solution is seldom made clear. In this paper, we showed, in a limited framework, but expanding results from previous works (Chalub and Souza,

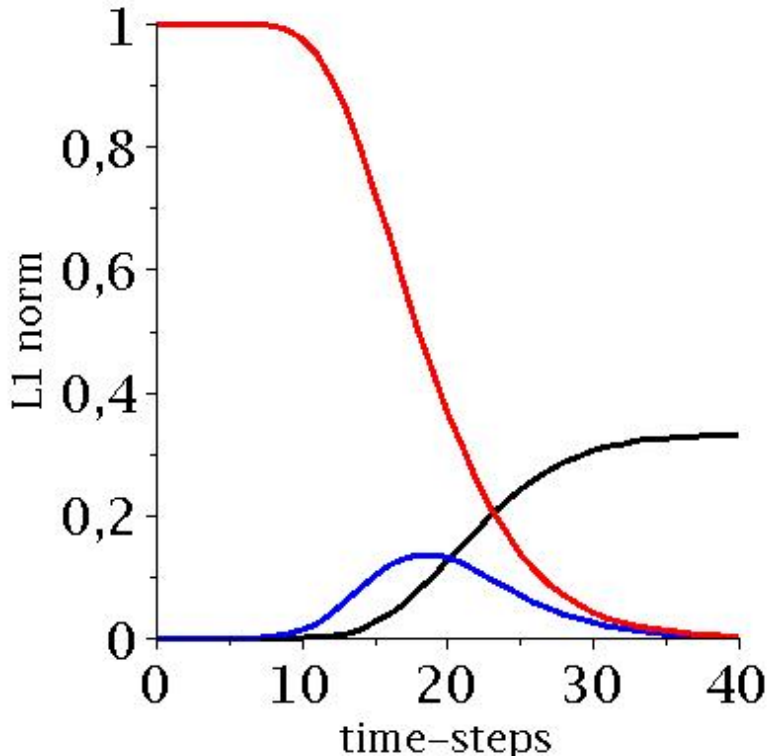


FIGURE 4. Evolution of the probability mass, for the Rock-Scissor-Paper game given by matrix (33) and with initial condition concentrated in the ESS, $p^1 = \delta_{(\frac{1}{3}, \frac{1}{3})}$. The red line indicates the mas (L^1 -norm) in the interior of the simplex; the blue line, the mass in the interior any of the faces, and the black line, the mass in any of the vertexes.

2009a,b), that ODEs can be justifiably used to model the evolution of a population. However, the validity of the modeling is necessarily limited in time (increasing with the population size), and the solution of the differential equation models the most probable state of the system (therefore, the differential equation would give answers compatible with the most likelihood method, but not necessarily compatible with other estimators).

We also overcame the limitations of the use of ODEs, obtaining a partial differential equation that approximates the evolution of the discrete process in all time scales. The price to pay is that this equations requires sophisticated techniques in its analysis, as it is a degenerated drift-diffusion equation without boundary conditions. We guarantee however the existence and uniqueness of solution in a restricted class of functions; this unique solution is the right limit of the Wright-Fisher process — in order to obtain the right class of functions, we imposed a set of conservation laws which comes directly from the discrete process.

The technique developed here was initially dedicated to obtain continuous limits from the Moran processes with two types. The case for the Wright-Fisher process with arbitrary number of types turned out to be far more sophisticated. This is consequence of the fact that arbitrarily large jumps in the population size are allowed in a single step (although the probability decreases with $e^{-\sigma^2}$, where σ is the size of the jump) and from the fact that the topology of a higher dimensional simplex is far more complicated than the $[0, 1]$ interval, specially with respect to the boundaries.

We are currently applying a similar technique to epidemiological models; in this case it is necessary to impose boundary conditions in part of the boundary (as an homogeneous population of infected individual is not stationary, as infected individuals become, with time, removed or

even susceptible) and it is impossible to impose boundary conditions in part of the boundary (a population of susceptible remains in this state for ever). Early results were already published in Chalub and Souza (2011b). The same problem, regarding the imposition of boundary conditions is true if we include mutations in the Moran or Wright-Fisher model. This is work in progress.

ACKNOWLEDGEMENTS

FACCC was partially supported by CMA/FCT/UNL, financiamento base 2011 ISFL-1-297 and projects PTDC/FIS/101248/2008, PTDC/FIS/70973/2006 from FCT/MCTES/Portugal. FACCC also acknowledges the hospitality of CRM/Barcelona where part of this work was performed and discussions with J. J. Velazquez (Madrid). MOS was partially supported by CNPq and FAPERJ.

APPENDIX A. WEAK FORMULATION IN TIME

In order to obtain a truly weak formulation, without any requirement upon the regularity of p , we observe that the equation above is valid for any time $t_k = t_0 + k\Delta t$. Hence, if we also let $T = (m + 1)\Delta t$ in the equation above, and sum over k , we obtain that

$$\begin{aligned} & \sum_{k=0}^m \int_{S^{n-1}} (p(\mathbf{x}, t_{k+1}) - p(\mathbf{x}, t_k)) g(\mathbf{x}, t_k) \, d\mathbf{x} \\ &= \sum_{k=0}^m \frac{1}{2N} \int_{S^{n-1}} p(\mathbf{x}, t_k) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t_k) \right) \, d\mathbf{x} \\ & \quad + \sum_{k=0}^m (\Delta t)^\nu \int_{S^{n-1}} p(\mathbf{x}, t_k) \left[\sum_{j=1}^{n-1} x_j \left(\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \partial_j g(\mathbf{x}, t_k) \right] \, d\mathbf{x} + O(N^{-2}). \end{aligned}$$

On summing by parts the left hand side, we obtain

$$\begin{aligned} & - \sum_{k=0}^{m-1} \int_{S^{n-1}} p(\mathbf{x}, t_k) (g(\mathbf{x}, t_{k+1}) - g(\mathbf{x}, t_k)) \, d\mathbf{x} \\ & - \int_{S^{n-1}} p(\mathbf{x}, t_0) g(\mathbf{x}, t_0) \, d\mathbf{x} + \int_{S^{n-1}} p(\mathbf{x}, T) g(\mathbf{x}, T) \, d\mathbf{x} \\ &= \sum_{k=0}^m \frac{1}{2N} \int_{S^{n-1}} p(\mathbf{x}, t_k) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t_k) \right) \, d\mathbf{x} \\ & \quad + \sum_{k=0}^m (\Delta t)^\nu \int_{S^{n-1}} p(\mathbf{x}, t_k) \left[\sum_{j=1}^{n-1} x_j \left(\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \partial_j g(\mathbf{x}, t_k) \right] \, d\mathbf{x} + O(N^{-2}). \end{aligned}$$

Now, let us take g to be C^1 in t , with support on $[t_0, T]$. Assuming Δt is sufficient small, we may write the previous equation as

$$\begin{aligned} & - \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \partial_t g(\mathbf{x}, t) \, d\mathbf{x} \, dt - \int_{S^{n-1}} p(\mathbf{x}, t_0) g(\mathbf{x}, t_0) \, d\mathbf{x} \\ &= \frac{(\Delta t)^{-1}}{2N} \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left(\sum_{i,j=1}^{n-1} x_i (\delta_{ij} - x_j) \partial_{ij}^2 g(\mathbf{x}, t) \right) \, d\mathbf{x} \, dt \\ & \quad + (\Delta t)^{\nu-1} \int_0^\infty \int_{S^{n-1}} p(\mathbf{x}, t) \left[\sum_{j=1}^{n-1} x_j \left(\psi^{(j)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right) \partial_j g(\mathbf{x}, t) \right] \, d\mathbf{x} + O(N^{-1}). \end{aligned}$$

REFERENCES

- Claude Bardos, François Golse, and David Levermore. Fluid dynamic limits of kinetic equations. I. Formal derivations. *J. Statist. Phys.*, 63(1-2):323–344, 1991.
- Claude Bardos, François Golse, and C. David Levermore. Fluid dynamic limits of kinetic equations. II. Convergence proofs for the Boltzmann equation. *Comm. Pure Appl. Math.*, 46(5):667–753, 1993.
- James D. Bjorken and Sidney D. Drell. *Relativistic Quantum Mechanics*. McGraw-Hill Book Co., New York, 1964.
- R. W. Carrol and R. Schowalter. *Singular and Degenerate Cauchy Problems*. Academic Press, 1976.
- C. Cercignani. The Boltzmann equation and fluid dynamics. In *Handbook of mathematical fluid dynamics, Vol. I*, pages 1–69. North-Holland, Amsterdam, 2002. doi: 10.1016/S1874-5792(02)80003-9. URL [http://dx.doi.org/10.1016/S1874-5792\(02\)80003-9](http://dx.doi.org/10.1016/S1874-5792(02)80003-9).
- Fabio A. C. C. Chalub and Max O. Souza. Multiscaling in evolutionary models. In preparation, 2011a.
- Fabio A. C. C. Chalub and Max O. Souza. A non-standard evolution problem arising in population genetics. *Commun. Math. Sci.*, 7(2):489–502, 2009a. ISSN 1539-6746. URL <http://projecteuclid.org/getRecord?id=euclid.cms/1243443991>.
- Fabio A. C. C. Chalub and Max O. Souza. From discrete to continuous evolution models: A unifying approach to drift-diffusion and replicator dynamics. *Theor. Pop. Biol.*, 76(4):268–277, 2009b.
- Fabio A. C. C. Chalub and Max O. Souza. The SIR epidemic model from a PDE point of view. *Math. And Comp. Model.*, 53(7-8):1568–1574, 2011b.
- Fabio A. C. C. Chalub and Max O. Souza. Continuous genetic models and generalized Fokker-Planck equations. In preparation, 2011c.
- Fabio A. C. C. Chalub, Peter A. Markowich, Benoît Perthame, and Christian Schmeiser. Kinetic models for chemotaxis and their drift-diffusion limits. *Monatsh. Math.*, 142(1-2):123–141, 2004. ISSN 0026-9255. doi: 10.1007/s00605-004-0234-7. URL <http://dx.doi.org/10.1007/s00605-004-0234-7>.
- Ross J. Cirincione and Paul R. Chernoff. Dirac and Klein-Gordon equations: convergence of solutions in the nonrelativistic limit. *Comm. Math. Phys.*, 79(1):33–46, 1981. ISSN 0010-3616. URL <http://projecteuclid.org/getRecord?id=euclid.cmp/1103908886>.
- R. Courant and D. Hilbert. *Methods of Mathematical Physics. Vol. II*. Wiley Classics Library. John Wiley & Sons Inc., New York, 1989. ISBN 0-471-50439-4. Partial differential equations, Reprint of the 1962 original, A Wiley-Interscience Publication.
- Emanuelle DiBenedetto. *Degenerate Parabolic Equations*. Springer-Verlag, 1993.
- Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, second edition, 2010. ISBN 978-0-8218-4974-3.
- Warren J. Ewens. *Mathematical Population Genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. ISBN 0-387-20191-2. Theoretical introduction.
- William Feller. Diffusion processes in genetics. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950*, pages 227–246, Berkeley and Los Angeles, 1951. University of California Press.
- R. A. Fisher. On the dominance ratio. *Proc. Royal Soc. Edinburgh*, 42:321–341, 1922.
- R. A. Fisher. The distribution of gene ratios for rare mutations. *Proc. Royal Soc. Edinburgh*, 50: 214–219, 1930.
- Gerald B. Folland. *Introduction to Partial Differential Equations*. Princeton University Press, Princeton, NJ, second edition, 1995. ISBN 0-691-04361-2.
- Daniel L. Hartle and Andrew G. Clark, editors. *Principles of Population Genetics*. Sinauer, Massachusetts, 2007.
- Klaus Hepp. The classical limit for quantum mechanical correlation functions. *Comm. Math. Phys.*, 35:265–277, 1974.

- Thomas Hillen and Hans G. Othmer. The diffusion limit of transport equations derived from velocity-jump processes. *SIAM J. Appl. Math.*, 61(3):751–775 (electronic), 2000. ISSN 0036-1399. doi: 10.1137/S0036139999358167. URL <http://dx.doi.org/10.1137/S0036139999358167>.
- Josef Hofbauer and Karl Sigmund. *Evolutionary Games and Population Dynamics*. Cambridge University Press, Cambridge, UK, 1998.
- L. A. Imhof and M. A. Nowak. Evolutionary game dynamics in a Wright-Fisher process. *J. Math. Biology*, 52(5):667–681, MAY 2006.
- Fritz John. *Partial Differential Equations*, volume 1 of *Applied Mathematical Sciences*. Springer-Verlag, New York, fourth edition, 1991. ISBN 0-387-90609-6.
- Laurent Keller, editor. *Levels of Selection in Evolution*. Monographs in Behavior and Ecology. Princeton University Press, Princeton, NJ, 1999.
- Motoo Kimura. On the probability of fixation of mutant genes in a population. *Genetics*, 47: 713–719, 1962.
- Sabin Lessard. Long-term stability from fixation probabilities in finite populations: New perspectives for ess theory. *Theoret. Popul. Biology*, 68(1):19 – 27, 2005.
- Sabin Lessard and Véronique Ladret. The probability of fixation of a single mutant in an exchangeable selection model. *J. Math. Biol.*, 54:721–744, 2007.
- Takeo Maruyama. *Stochastic problems in population genetics*, volume 17 of *Lecture Notes in Biomathematics*. Springer-Verlag, Berlin, 1977. ISBN 3-540-08349-9.
- A. J. McKane and D. Waxman. Singular solutions of the diffusion equation of population genetics. *J. Theoret. Biol.*, 247(4):849–858, 2007.
- J. A. J. Metz. Thoughts on the geometry of meso-evolution: collecting mathematical elements for a post-modern synthesis. In F. A. C. C. Chalub and J. F. Rodrigues, editors, *The Mathematics of Darwin's Legacy*, Mathematics and Biosciences in Interaction, pages 193–231, Basel, 2011. Springer.
- J. D. Murray. *Mathematical Biology. II*, volume 18 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, third edition, 2003. ISBN 0-387-95228-4. Spatial models and biomedical applications.
- Martin A. Nowak. *Evolutionary Dynamics — Exploring the Equations of Life*. The Belknap Press of Harvard University Press, Cambridge, MA, 2006. ISBN 978-0-674-02338-3; 0-674-02338-2.
- Hans G. Othmer and Thomas Hillen. The diffusion limit of transport equations. II. Chemotaxis equations. *SIAM J. Appl. Math.*, 62(4):1222–1250 (electronic), 2002. ISSN 0036-1399. doi: 10.1137/S0036139900382772. URL <http://dx.doi.org/10.1137/S0036139900382772>.
- Robert D. Richtmyer and K. W. Morton. *Difference Methods for Initial-Value Problems*. John Wiley & Sons, 1967.
- K. I. Sato. Class of Markov-chains related to selection in population-genetics. *J. Math. Soc. of Japan*, 28(4):621–637, 1976.
- Ken-Iti Sato. Convergence to a diffusion of a multi-allelic model in population genetics. *Adv. in Appl. Probab.*, 10(3):538–562, September 1978. ISSN 00018678. URL <http://www.jstor.org/stable/1426633>.
- Ken-Iti Sato. Limit diffusions of some stepping-stone models. *J. Appl. Prob.*, 20(3):460–471, September 1983. ISSN 00219002. URL <http://www.jstor.org/stable/3213884>.
- J. M. Smith. Can a mixed strategy be stable in a finite population? *J. of Theor. Biol.*, 130(2): 247–251, JAN 21 1988. ISSN 0022-5193.
- John Maynard Smith. *Evolution and the theory of games*. Cambridge University Press, Cambridge, UK, 1982.
- Angela Stevens. The derivation of chemotaxis equations as limit dynamics of moderately interacting stochastic many-particle systems. *SIAM J. Appl. Math.*, 61(1):183–212 (electronic), 2000. ISSN 0036-1399. doi: 10.1137/S0036139998342065. URL <http://dx.doi.org/10.1137/S0036139998342065>.
- D.S. Strock and S.R.S. Varadhan. *Multidimensional Diffusion Processes*. Springer-Verlag, 1997.
- Michael E. Taylor. *Partial Differential Equations. I*, volume 115 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 1996. ISBN 0-387-94653-5. Basic theory.

- Arne Traulsen, Jens Christian Claussen, and Christoph Hauert. Coevolutionary dynamics: From finite to infinite populations. *Phys. Rev. Lett.*, 95(23):238701, Dec 2005. doi: 10.1103/PhysRevLett.95.238701.
- N. G. Van Kampen. *Stochastic Processes in Physics and Chemistry*. Amsterdam: North-Holland, 2001.
- W. L. Vickery. Can a mixed strategy be stable in a finite population? — reply. *J. of Theor. Biol.*, 132(3):375–378, JUN 7 1988. ISSN 0022-5193.
- D. Waxman. Comparison and content of the Wright-Fisher model of random genetic drift, the diffusion approximation, and an intermediate model. *J. of Theoret. Biol.*, 269(1):79 – 87, 2011.
- Jörgen W. Weibull. *Evolutionary Game Theory*. The MIT Press, Cambridge, Massachussets, 1995.
- S Wright. The distribution of gene frequencies under irreversible mutations. *Proc. Nat. Acad. Sci. US*, 24:253–259, 1938.
- S Wright. The distribution of gene frequencies in populations. *Proc. Nat. Acad. Sci. US*, 23: 307–320, 1937.

DEPARTAMENTO DE MATEMÁTICA AND CENTRO DE MATEMÁTICA E APLICAÇÕES, UNIVERSIDADE NOVA DE LISBOA, QUINTA DA TORRE, 2829-516, CAPARICA, PORTUGAL.
E-mail address: `chalub@fct.unl.pt`

DEPARTAMENTO DE MATEMÁTICA APLICADA, UNIVERSIDADE FEDERAL FLUMINENSE, R. MÁRIO SANTOS BRAGA, S/N, 22240-920, NITERÓI, RJ, BRASIL.
E-mail address: `msouza@mat.uff.br`