Noname manuscript No.
(will be inserted by the editor)

The frequency-dependent Wright-Fisher model: diffusive and non-diffusive approximations.

Fabio A. C. C. Chalub · **Max O. Souza**

February 7, 2013

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. By considering an approximate weak formulation of the discrete problem, we are able to derive a corresponding continuous weak formulation for the probability density. Therefore, 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, small time-steps and weak selection limit. If the fitness functions are sufficiently regular, we can recast the weak formulation in a more standard formulation, without any boundary conditions, but supplemented by a number of conservation laws. The equations in this family can be purely diffusive, purely hyperbolic or of convection-diffusion type, with frequency dependent convection. 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 accurate is the convective approximation, with respect to the convective-diffusion approximation. In particular, we show that the mode, but not the expected value, of the probability distribution is modelled by the replicator dynamics. Some numerical simulations that illustrate the results are also presented.

Keywords Wright-Fisher process · diffusion approximations · continuous limits · replicator equation

Mathematics Subject Classification (2000) 92D15 · 92D25 · 35K57 · 35K67 · 35L65

Departamento de Matemática and Centro de Matemática e Aplicações, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Caparica, Portugal. E-mail: chalub@fct.unl.pt · Departamento de Matemática Aplicada, Universidade Federal Fluminense, R. Mário Santos Braga, s/n, 22240-920, Niterói, RJ, Brasil. E-mail: msouza@mat.uff.br

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 2009a). 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); see also the review in Sato (1978). Subsequently, Ethier and Kurtz (1986) systematically studied the approximation of finite Markov chain models by diffusions. In particular, they showed the validity of a diffusion approximation to a multidimensional Wright-Fisher model, in the regime of weak selection, and linear fitness. This led to a notable progress in diffusion theory, as reported for instance in (Ethier and Kurtz 1986; Stroock and Varandhan 1997). This considerable progress, in turn, led to a large use of diffusion theory in population genetics, which can be verified in contemporary introductions to the subject (see Ewens 2004; Etheridge 2011).

There is also a more heuristic 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 can be viewed as a Kramers-Moyal expansion truncated at second order. Although it is commonly claimed that the full expansion is needed in order to obtain a continuous approximation of discrete processes, it is known that under various conditions discrete Markov chains can be approximated by diffusions; cf. Ethier and Kurtz (1986) and Stroock and Varandhan (1997) for instance. In this work, we shall show that under a number of conditions similar results hold for the discrete processes considered. See (van Kampen 1981) for a discussion about this and other techniques for continuous approximations of discrete processes.

As observed above, results along similar lines had been obtained earlier by a number of authors (Feller 1951; Ethier and Kurtz 1986; Ewens 2004). These works approach the problem mostly within a probabilistic framework, while here we take a pure analytical setting, and this brings two immediate consequences: firstly, we are able to directly derive a weak formulation for the forward Kolmogorov equation, assuming only continuity of the fitness functions in contrast to the weak formulation for the backward problem for Feller processes; see Rogers and Williams (2000a,b) . The second, and possibly most important consequence, is that we are able to deal with a variety of scalings for the evolution problem. This yields a full family of evolution problems: genetic-drift dominated evolution, which is described by a diffusion equation; selection dominated evolution, which is governed by a hyperbolic equation; and an evolutionary dynamics, where the two forces are balanced, which is governed by a convection-diffusion equation that we term replicator-diffusion.

If we assume some more regularity of the fitness functions, we can then recast the weak formulation in a strong formulation. In this case, we cannot impose any boundary conditions, but we must supplement it by a number of conservation laws, namely that the probability of fixation of each type for a given probability density of population, in any time, must be the same as in the initial time. The conservation laws are used to circumvent the impossibility of imposing boundary conditions when the boundaries are absorbing.

Furthermore, by a duality argument we obtain the backward equation formulation. For the particular case of linear fitness and balanced scalings, we then recover the classical result by Ethier and Kurtz (1986). Additionally, by an appropriate combination of the weak and strong formulations, we are able to give a complete description of the forward solution.

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 dynamics[1] 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). The replicator equation was also modified to introduce stochasticity at population level (Fudenberg and Harris 1992; Foster and Young 1990). Relations between the matching scheme in a population and the deterministic approximation of its stochastic evolution are studied in (Boylan 1992, 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; Champagnat et al. 2006, 2008; Fournier and Méléard 2004; Molzon 2009; Benaim and Weibull 2003; Corradi and Sarin 2000; Traulsen et al. 2012). We will show, as in many of these works, that both descriptions — the one based on the diffusion approximation and the one based on the replicator dynamics — are correct as models for the evolutionary dynamics of a given trait, but in different scalings. As a by product, we will provide a generalisation of the Kimura equation valid for an arbitrary number of types and general fitness functions. The long time asymptotics of both descriptions will suggest that the replicator equation is a model with limited time validity, given a certain maximum admissible error. Such a limitation is to be naturally anticipated on the grounds that the diffusion process will be eventually absorbed, while the replication dynamics might converge to an equilibrium in the interior. We also confirm that the solution of the replicator equation indicates the most probable state (mode) of the population, conditional on not have been absorbed. Hence, it does not necessarily indicate the expected value of the trait.

[1] In this work, we will use the expressions "replicator dynamics", "replicator equation" and "replicator system" indistinctly.

The work presented here is a development of earlier work in (Chalub and Souza 2009a,b): the former studying the derivation and convergence of the Moran model with two types to the 1-d version of the replicator-diffusion equation discussed here, and the latter with a comprehensive analytical study of the 1-d replicator-diffusion equation. The derivation of the continuous model in Chalub and Souza (2009a) hinged on the idea that a formal expansion of master equation, but with control of the local error, and results on well-posedness of the continuous classical problem can be brought together via numerical analysis approximation results. This combination then yielded uniform convergence, in any proper closed sub interval of [0,1], of the rescaled probabilities of the discrete model to the continuous probability density. This convergence result, combined with the analytical results in Chalub and Souza (2009b) on a weak formulation that satisfies the conservation laws provided a continuous measure solution. The discrete process then converges weakly towards such a solution, on a neighbourhood of each endpoint, but uniformly as described above. To study the Wright-Fisher continuous limits, however, we took a different route. This allows us to derive an approximate discrete weak formulation of the discrete process, with global error control. Further, by embedding the discrete probabilities in an appropriate measure space, we could use compactness arguments to obtain the continuous limit. Thus, in this setting both the weak formulation and the weak convergence of the discrete model to the continuous one follows with considerable less effort, but we do not get the improved convergence on the interior.

1.1 Scalings, limits and approximations

In order to be able to study more general models, we follow the approach used by the authors in Chalub and Souza (2009a). 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 *We shall say that a simplified model $\mathscr{M}_0$ is an approximation of the family of detailed models $\mathscr{M}_\gamma$, $\gamma > 0$, in a sense χ, where χ is an appropriate metric as, for instance, any norm in a suitable space of functions (e.g., L^1, L^2, L^∞, etc) if the following holds true:*

1. *Consider a certain family of initial conditions h^{I}_γ such that $\lim_{\gamma\to 0} h^{\mathrm{I}}_\gamma = h^{\mathrm{I}}_0$, in the sense χ;*
2. *Evolve through the model $\mathscr{M}_\gamma$ the initial condition h^{I}_γ and through the model $\mathscr{M}_0$ the initial condition h^{I}_0 until the time $t < \infty$ obtaining $h_\gamma(t)$ and $h_0(t)$ respectively;*

If for all $t < \infty$ we have that $\lim_{\gamma\to 0} h_\gamma(t) = h_0(t)$, in the sense χ, then we say that the model $\mathscr{M}_\gamma$ converge in the limit $\gamma \to 0$, in the sense χ, to the model $\mathscr{M}_0$. If, furthermore, this convergence is uniformly in $t \in [0,\infty)$, then we say that the model $\mathscr{M}_\gamma$ converge in the limit $\gamma \to 0$ to the model $\mathscr{M}_0$ uniformly in time.

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

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 simplified models. The last two lines state that both the replicator equation and the replicator diffusion equation approximates the Moran process. 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 2009b,a).

In general, some extra assumptions are frequently required to allow the passage to the limit. 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 an approximation in the sense of definition 1 is the so-called *weak selection principle*, to be precisely stated in equation (7). Generally speaking, we assume that the *fitness* of a given individual converges 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 $N \to \infty$, 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.

In this work, we will consider as the detailed model, the Wright-Fisher process, to be studied in detail for finite populations in Section 3: 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 inter generation time).

In short, given the weak-selection principle, we are able to find a precise scaling that yields, as the thermodynimical limit. a parabolic equation with degenerate diffusion at the boundaries: namely the replicator-diffusion equation. Under a range of different scalings, however, we shall also obtain a simpler first order differential system: the replicator equation. This simpler model turns out to be a good approximation for the detailed model over a short time scale, if genetic drift is weak or selection is strong[2]. For the former approximation, nevertheless, we shall be able to show that $\lim_{\gamma\to 0}\lim_{t\to\infty} h_\gamma(t) = \lim_{t\to\infty}\lim_{\gamma\to 0} h_\gamma(t)$, and hence we conjecture that such an approximation should be uniform in time.

1.2 Outline

Section 2 introduces the basic notation and provides an extended abstract of our main results. In Section 3, we review some classical results about the discrete process (the finite population Wright-Fisher process); we also show the existence of a number of associated conservation laws, and explicitly obtain the first moments of the Wright-Fisher process. In Section 4, with the assumption of weak-selection, we obtain a family of continuous limits of the Wright-Fisher process depending on the scalings that are derived within a weak formulation, with solutions in appropriate measure spaces. In particular, we derive the replicator-diffusion equation, and show that it satisfies continuous counterparts of the conservation laws for the discrete process. We then continue the study of the replicator-diffusion equation in Section 5, where we derive the main properties of its solutions, including a description of the solution structure as a regular part and a sum of singular measures over the sumbsimplices, and the large time convergence to a sum of Dirac measures over the vertices of the simplex. 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 generalisation of the Kimura equation for the n types and arbitrary fitness functions. In Section 6, we study the replicator equation and show that, in the regime of strong selection the solutions to the replicator-diffusion will be well approximated by the solutions to the replicator equation within a finite time interval. Numerical examples are given in Section 7, where we also point out that, for intermediate times and large but finite populations, the replicator equation will approximate the mode of the discrete evolution, but not the expected value of a given trait. 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*

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

We also define the set of vertices of the simplex $\Delta S^{n-1} := \{\mathbf{x} \in S^{n-1} | \exists i, x_i = 1\}$, its interior $\operatorname{int} S^{n-1} := \{\mathbf{x} \in S^{n-1} | \forall i, x_i > 0\}$ and its boundary $\partial S^{n-1} = S^{n-1} \backslash \operatorname{int} S^{n-1}$. The state of the population is a vector $\mathbf{x} \in S^{n-1}$. The elements of ΔS^{n-1} are denoted $\mathbf{e}_i$,

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

$i = 1,\dots,n$ and called "homogeneous states". A vector $\mathbf{x} \in S^{n-1}\backslash\Delta S^{n-1}$ 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 *The fitness of type i, $i = 1,\dots,n$ is a continuous function $\psi^{(i)} : S^{n-1} \to \mathbb{R}$, 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 derive 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:

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

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:

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 pointwise for any time. In addition, the large time asymptotics is consistent with the discrete model.
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 suggest that the replicator dynamics approximates the discrete process for any t, however with an error increasing in t along a fixed discretisation
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 generalisation 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 4.5, 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 11), we have $p_\kappa \stackrel{\kappa\to 0}{\longrightarrow} p_0$, pointwise, but not uniformly in time.

Theorem 11 cannot be made uniform in time, for general fitness functions and initial conditions, as the Wright-Fisher process always converge in $t \to \infty$ to a linear combination of homogeneous states, while it is possible that the solution of the replicator equation converges to a stable mixed state.

The former statement 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 vertices 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 vertices of the simplex. These measures appears immediately, i.e., for any $t > 0$. This represents the fact that in a single step there is a non zero — however, small — probability that the population reaches fixation through Wright-Fisher evolution. 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 decrease 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.

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

Finally, we observe that the natural formulation for the continuous limit is the weak one. For such a formulation, we only require the fitness functions to be continuous. If, in addition, these functions are also Lipschitz, we can then recast the problem in a strong sense, provided that we supplement it with the conservation laws. Finally, requiring the fitness functions to be smooth allows for a number of results about the solutions to be easily derived. In particular, one can prove a structure theorem that show that the problem is equivalent to a hierarchy of classical degenerate problems, provided that some members are interpreted as densities for singular measures.

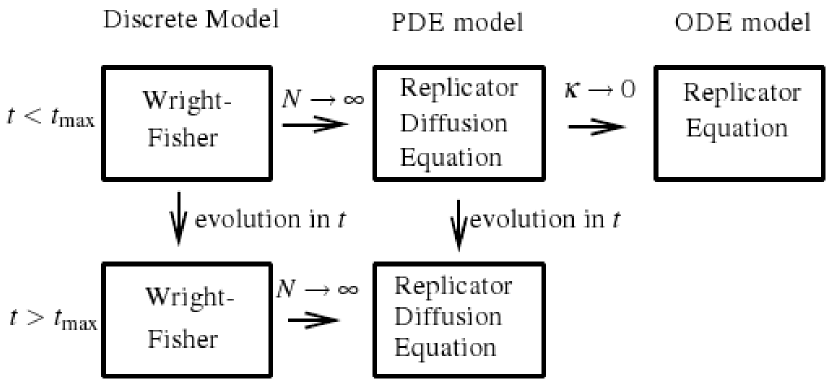

Fig. 1 The boxes in the figure represents the solutions of three different models: the Wright-Fisher process (finite population N), the replicator-diffusion equation (positive diffusion κ) and the replicator equation. The vertical axis indicates the arrow of time (top-down), and the horizontal axis indicates, first the large population limit, secondly the no-diffusion limit. Consider that there is a maximum acceptable error ε (in the L^∞ norm) between the Wright-Fisher model (suitably *Radonmised* — see Subsection 4.3) and the continuous approximation. Therefore, there is a population size N_0 such that for $N > N_0$ the difference between the replicator-diffusion equation and the discrete model is less than ε. For the replicator approximation, for any N, it may exist maximal time $t_{\max}(N) < \infty$ such that for $t > t_{\max}$ the error is too large.

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 definitions; in Subsection 3.2 we briefly review some important results in the literature. We also prove that the discrete process has as many linear conservation laws as types. Additionally, we also show 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 discrete 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. See also Ethier and Kurtz (1986) and Etheridge (2011) and references therein for a discussion in the rôle of time scales.

3.1 Preliminaries

We consider a fixed size population of N individuals at time t consisting of a fraction $x_i \in \left\{0, \frac{1}{N}, \frac{2}{N}, \ldots, 1\right\}$ of individuals of type $i = 1, 2, \cdots, n$. 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*

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

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

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

The elements of ΔS_N^{n-1} are called the homogeneous states. *To each type we attribute a function, called fitness, $\Psi_{\Delta t}^{(i)} : S_N^{n-1} \to (0, \infty)$. It is convenient to assume that $\Psi_{\Delta t}^{(i)}$ is a discretisation of a continuous 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

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

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

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

$$\mathscr{P} \in \left\{ P : S_N^{n-1} \times \mathbb{R}_+ \to \mathbb{R}_+ | \sum_{\mathbf{x} \in S_N^{n-1}} P(\mathbf{x}, \cdot) = 1 \right\} ,$$

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 the so called "master equation":

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

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 inner product definition:

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

it follows immediately that $\langle \hat{P}_{\mathbf{x}},\hat{P}_{\mathbf{y}}\rangle=\delta_{\mathbf{x},\mathbf{y}}=1$ if $\mathbf{x}=\mathbf{y}$ and 0 otherwise.

Now, we state classical results for the Wright-Fisher process that will be useful in the sequel. Firstly, from the Perron-Frobenius theorem, the operator $\mathscr{T}^\infty : S_N^{n-1}\to S_N^{n-1}$, $\mathscr{T}^\infty P := \lim_{m\to\infty}\mathscr{T}^m P$ is well defined. For details in the lemma below, the interested reader should consult Karlin and Taylor (1975).

Lemma 1 *A function f defined in S_N^{n-1} is a fixed state of the operator $\mathscr{T}$ if, and only if, f is a linear combination of homogeneous states. In particular, $\mathscr{T}$ has exactly n linearly independent eigenfunctions associated to the eigenvalue $\lambda=1$. For all non-negative initial condition P^{I}, the final result is a linear combination of homogeneous states,*

$$\mathscr{T}^\infty P := \lim_{t\to\infty} P(\cdot,t)=\sum_{i=1}^{n} F^{(i)}_{P_{\mathbf{e}_i}}\hat{P}_{\mathbf{e}_i}\,,$$

where $F_P^{(i)} := \lim_{m\to\infty}\langle \mathbf{e}_i,\mathscr{T}^m P\rangle$ is the fixation probability of the type i in a population initially with a probability distribution $P\in S_N^{n-1}$.

Definition 5 *We define a linear conservation law as one given by a linear functional L over the functions of S_N^{n-1} such that $\mathsf{L}(\mathscr{P}(t+\Delta t))=\mathsf{L}(\mathscr{P}(t))$. A set of linear conservation laws is linearly independent, if the only linear combination providing a trivial conservation law $\mathsf{L}(\mathscr{P}(t))=0$ is the trivial one.*

Proposition 1 *Define $\mathsf{F}^{(i)} := \sum_{\mathbf{x}\in S_N^{n-1}} F^{(i)}_{\hat{P}_{\mathbf{x}}}\hat{P}_{\mathbf{x}}$, $i=1,\dots,n$, a functional over S_N^{n-1}. Therefore $\mathsf{F}^{(i)}(\mathbf{x})$ is the fixation probability of the type i associated to the initial condition $\mathbf{x}$. Finally, the set $\{\mathsf{F}^{(1)},\dots,\mathsf{F}^{(n)}\}$ is a basis for the set of linear conservation laws associated to the operator $\mathscr{T}$.*

Proof From the fact that

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

we find

$$F_P^{(i)}=\left(\mathscr{T}^\infty P\right)(\mathbf{e}_i)=\langle \mathscr{T}^\infty P,\hat{P}_{\mathbf{e}_i}\rangle=\langle P,\left(\mathscr{T}^\dagger\right)^\infty \hat{P}_{\mathbf{e}_i}\rangle\,.$$

In particular

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

Finally,

$$\left(\mathscr{T}^{\dagger}\right)^{\infty}\hat{P}_{\mathbf{e}_i}=\sum_{\mathbf{x}\in S_N^{n-1}}F^{(i)}_{\hat{P}_{\mathbf{x}}}\hat{P}_{\mathbf{x}}=\mathsf{F}^{(i)}\ .$$

Therefore, $\mathsf{F}^{(i)}$ is an eigenvector of $\mathscr{T}^{\dagger}$. In particular,

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

It is immediate to prove that they are linearly independent; let $\alpha_1,\ldots,\alpha_n$ such that $\sum_{i=1}^n\alpha_i\mathsf{F}^{(i)}=0$, i.e., for every $\mathbf{x}\in S_N^{n-1}$, $\sum_{i=1}^n\alpha_i\mathsf{F}^{(i)}(\mathbf{x})=0$. Using $\mathbf{x}=\mathbf{e}_i$, we conclude that $\alpha_i=0$, and then $\{\mathsf{F}^{(1)},\ldots,\mathsf{F}^{(n)}\}$ is a basis for the eigenspace of $\mathscr{T}^{\dagger}$ associated to $\lambda=1$.

Now, consider a linear conservation law L. From standard representation theorems, there is a vector $w\in S_N^{n-1}$ such that

$$\langle\mathscr{P}(t),w\rangle=\mathsf{L}(\mathscr{P}(t))=\mathsf{L}(\mathscr{P}(t+\Delta t))=\langle\mathscr{T}\mathscr{P}(t),w\rangle=\langle\mathscr{P}(t),\mathscr{T}^{\dagger}w\rangle\ . \qquad (4)$$

Therefore, w is an eigenvector of $\mathscr{T}^{\dagger}$ associated to $\lambda=1$ and then it is a linear combination of $\mathsf{F}^{(i)}$, $i=1,\ldots,n$.

Remark 1 *The conservation of probability (the most natural conservation law), follows directly from the equation*

$$\sum_{i=1}^n\mathsf{F}^{(i)}(\mathbf{x})=\sum_{i=1}^nF^{(i)}_{\hat{P}_{\mathbf{x}}}=1\ ,\ \forall\mathbf{x}\in S_N^{n-1}\ .$$

3.3 Properties of the transition kernel

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

$$\sum_{\mathbf{x}\in S_N^{n-1}}\Theta_{N,\Delta t}(\mathbf{y}\to\mathbf{x})=1,\qquad\forall\mathbf{y}\in S_N^{n-1}\ . \qquad (5)$$

It also follows from the definition that

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

which can be readily interpreted as the absence of mutations in the model.

It will be also convenient to write

$$S^{n-1}_{N,\mathbf{x}^{\pm}}=\left\{\mathbf{y}\in\mathbb{R}^{n-1}|\mathbf{x}\pm\mathbf{y}\in S_N^{n-1}\right\}.$$

and to introduce

$$z\tau_i=y_i,\quad z=\frac{1}{\sqrt{N}}.\quad\text{and}\quad\mathscr{S}_{\mathbf{x},z}=\{\tau\in\mathbb{R}^n|\sum_{i=1}^n\tau_i=0\ \text{and}|\tau_i|<x_i/z\}.$$

Lemma 2 *Define*

$$\widetilde{x}_i = \frac{x_i \Psi^{(i)}_{\Delta t}(\mathbf{x})}{\bar{\Psi}(\mathbf{x})}$$

and

$$\mathbb{G}_\Theta[h] = \sum_{z\tau \in S^{n-1}_{N,\mathbf{x}^+}} \Theta(\mathbf{x} \to \mathbf{x} + z\tau) h(\tau) \ ,$$

where $h : \mathscr{S}_{\mathbf{x},z} \to \mathbb{R}$.

For any N*, we have*

$$\begin{aligned}
\mathbb{G}_\Theta[1] &= 1 \\
z\mathbb{G}_\Theta[\tau_i] &= \widetilde{x}_i - x_i \\
z^2\mathbb{G}_\Theta[\tau_i\tau_j] &= (\widetilde{x}_i - x_i)(\widetilde{x}_j - x_j) + z^2\left(\delta_{ij}\widetilde{x}_i - \widetilde{x}_i\widetilde{x}_j\right) \\
z^3\mathbb{G}_\Theta[\tau_i\tau_j\tau_k] &= (\widetilde{x}_i - x_i)(\widetilde{x}_j - x_j)(\widetilde{x}_k - x_k) \\
&\quad + z^2\left[(\delta_{ij}\widetilde{x}_i - \widetilde{x}_i\widetilde{x}_j)(\widetilde{x}_k - x_k) + (\delta_{ik}\widetilde{x}_i - \widetilde{x}_i\widetilde{x}_k)(\widetilde{x}_j - x_j) + (\delta_{jk}\widetilde{x}_j - \widetilde{x}_j\widetilde{x}_k)(\widetilde{x}_i - x_i)\right] \\
&\quad + z^4\left[2\widetilde{x}_i\widetilde{x}_j\widetilde{x}_k - (\delta_{ij}\widetilde{x}_j\widetilde{x}_k + \delta_{ik}\widetilde{x}_i\widetilde{x}_j + \delta_{kj}\widetilde{x}_i\widetilde{x}_j) + \delta_{ij}\delta_{ik}\delta_{jk}\widetilde{x}_i\right]
\end{aligned}$$

Proof Let $\mathbf{q} \in S^{n-1}$, $\alpha \in NS^{n-1}_N$ and consider the multinomial distribution given by

$$f(\mathbf{q},\alpha,N) = \frac{N!}{\alpha_1!\cdots\alpha_n!}\prod_{k=1}^n q_k^{\alpha_k}, \quad \alpha = (\alpha_1,\ldots,\alpha_n), \quad \sum_i \alpha_i = N \ .$$

Then, α is a vector of random variables with first moments given by

$$\begin{aligned}
&\mathbb{E}[1] = 1 \ , \\
&\mathbb{E}[\alpha_i] = Nq_i \ , \\
&\mathbb{E}[(\alpha_i - Nq_i)(\alpha_j - Nq_j)] = \mathrm{Cov}(\alpha_i,\alpha_j) = N(\delta_{ij}q_i - q_iq_j) \ , \\
&\mathbb{E}[(\alpha_i - Nq_i)(\alpha_j - Nq_j)(\alpha_k - Nq_k)] = N\left[q_i\delta_{ij}\delta_{kj} - \left(q_iq_k\delta_{ij} + q_iq_j\delta_{kj} + q_kq_j\delta_{ik}\right) + 2q_iq_jq_k\right] ,
\end{aligned}$$

where $\mathbb{E}[\cdot]$ is the expected value under the multinomial distribution. See Taylor and Karlin (1998) for the mean and covariance; for the sake of completeness, we provide a derivation of the third moment in Appendix A.

Now, note that $\Theta_{N,\Delta t}(\mathbf{x},\mathbf{x}+z\tau) = f(\widetilde{\mathbf{x}},N(\mathbf{x}+z\tau),N)$. Therefore, $\alpha = N(\mathbf{x}+z\tau)$ is a random vector that is multinomially distributed, and upon substituting α in the multinomial moments — with $\mathbf{q} = \tilde{\mathbf{x}}$ — all the identities follow after some manipulation. □

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 ($N \to \infty$) and there is no time-separation between successive generations ($\Delta t \to 0$). The relevant variables, $\mathbf{x} \in S^{n-1}$ and $t > 0$ will be forced to be continuous.

The first four subsections will be devoted to the development of three models, based on partial differential equations obtained from the Wright-Fisher process, when $N \to \infty$ and $\Delta t \to 0$ (see equations (15), (16) and (1')), respectively). There is no "right choice" of the simplified model. As we could expect, simpler models will have a restricted application. For example, the model given by equation (15) 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 (16), is a parabolic partial differential equation that is much simpler to solve than the full model; in fact, explicit solutions are know using Gegenbauer polynomials (Ewens 2004). Our focus will be on the replicator-diffusion approximation, equation (1'), which we expect to be valid uniformly in time.

Results known for the Wright-Fisher process, and stated in Section 3 will guide the derivation, i.e., the choice of the right thermodynamical limit. We start in Subsection 4.1 by the asymptotic expansion of the transition kernel in the negligible parameters (suitable combinations of N and Δt); we plug this expansion into the master equation (3) in Subsection 4.2. In Subsection 4.3, we construct the continuous version of the discrete probability densities; in particular, we interpolate discrete probabilities in order to represent them by continuous probability measures; these measures will be central when we finally pass to the limits in Subsection 4.4, obtaining the various continuous approximations of the discrete model. Finally, in Subsection 4.5, we show that, for every conservation law of the discrete process, there exists a corresponding conservation law in the continuous model. As a by product, the final state of the continuous model shall be a linear superposition of homogeneous states (see Lemma 1 and compare it with Theorem 7).

4.1 Preliminaries

From a biological point of view, the most important assumption in this derivation is the so called weak selection principle

$$\Psi^{(i)}_{\Delta t}(\mathbf{y}) = 1 + (\Delta t)^{\nu}\, \psi^{(i)}(\mathbf{y}), \tag{7}$$

where $\psi^{(i)} : S^{n-1} \to \mathbb{R}$ is a continuous function, and $\nu > 0$ is a parameter yet to be specified. In this case, we also have

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

where $\bar{\psi}(\mathbf{x}) := \sum_{i=1}^{n} x_i \psi^{(i)}(\mathbf{x})$.

As an immediate corollarium of Lemma 2 we have

Corollary 1 *Assume the weak-selection principle given by equation (7). Then, equations for $\mathbb{G}_\Theta$ in Lemma 2 are given by*

$$
\begin{aligned}
z\mathbb{G}_\Theta[\tau_i] &= x_i\,(\Delta t)^\nu\left(\psi^{(i)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)+O\left((\Delta t)^{2\nu}\right)\ ,\\
z^2\mathbb{G}_\Theta[\tau_i\tau_j] &= \frac{1}{N}\left(x_i\delta_{ij}-x_ix_j\right)+O\left(N^{-1}(\Delta t)^\nu,(\Delta t)^{2\nu}\right)\\
z^3\mathbb{G}_\Theta[\tau_i\tau_j\tau_k] &= O\left((\Delta t)^{3\nu},N^{-1}(\Delta t)^\nu,N^{-2}\right)\ .
\end{aligned}
$$

Proof All equations follow from the fact that

$$
\widetilde{x}_i = x_i\frac{1+(\Delta t)^\nu\,\psi^{(i)}(\mathbf{x})}{1+(\Delta t)^\nu\,\bar{\psi}(\mathbf{x})} = x_i\left[1+(\Delta t)^\nu\left(\psi^{(i)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)+O\left((\Delta t)^{2\nu}\right)\right]
$$

and from Lemma 2.

4.2 An asymptotic weak-discrete formulation

We rewrite the master equation (3) using displacements:

$$
p(\mathbf{x},t+\Delta t,N)=\sum_{\mathbf{y}\in S^{n-1}_{N,\mathbf{x}^-}}\Theta_{N,\Delta t}(\mathbf{x}-\mathbf{y}\to\mathbf{x})p(\mathbf{x}-\mathbf{y},t,N). \tag{8}
$$

We use our information on moments of the process as follows:

Proposition 2 *Let $g\in C^{3,1}(\Upsilon)$, where Υ is any open set such that $S^{n-1}\subset\Upsilon$, and consider its restriction to the simplex S^{n-1}. Then, we have*

$$
\begin{aligned}
&\sum_{\mathbf{x}\in S^{n-1}_N}\left(p(\mathbf{x},t+\Delta t,N)-p(\mathbf{x},t,N)\right)g(\mathbf{x},t) \qquad (9)\\
&\quad=\sum_{\mathbf{x}\in S^{n-1}_N}p(\mathbf{x},t,N)\left[\frac{1}{2N}\sum_{i,j=1}^{n-1}x_i(\delta_{ij}-x_j)\partial^2_{ij}g(\mathbf{x},t)+(\Delta t)^\nu\sum_{j=1}^{n-1}x_j\partial_{x_j}g(\mathbf{x},t)(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x}))\right]\\
&\qquad+O\left(N^{-2},(\Delta t)^{2\nu},N^{-1}(\Delta t)^\nu\right).
\end{aligned}
$$

Proof On multiplying equation (8) by $g(\mathbf{x},t)$ and summing over S^{n-1}_N, we find that:

$$
\begin{aligned}
&\sum_{\mathbf{x}\in S^{n-1}_N}p(\mathbf{x},t+\Delta t,N)g(\mathbf{x},t)=\sum_{\mathbf{x}\in S^{n-1}_N}\sum_{\mathbf{y}\in S^{n-1}_{N,\mathbf{x}^-}}\Theta_{N,\Delta t}(\mathbf{x}-\mathbf{y}\to\mathbf{x})p(\mathbf{x}-\mathbf{y},t,N)g(\mathbf{x},t)\\
&\quad=\sum_{\mathbf{x}\in S^{n-1}_N}\sum_{\mathbf{y}\in S^{n-1}_{N,\mathbf{x}^+}}\Theta_{N,\Delta t}(\mathbf{x}\to\mathbf{x}+\mathbf{y})p(\mathbf{x},t,N)g(\mathbf{x}+\mathbf{y},t)\\
&\quad=\sum_{\mathbf{x}\in S^{n-1}_N}p(\mathbf{x},t,N)\sum_{z\tau\in S^{n-1}_{N,\mathbf{x}^+}}\Theta(\mathbf{x}\to\mathbf{x}+z\tau)g(\mathbf{x}+z\tau,t)\\
&\quad=\sum_{\mathbf{x}\in S^{n-1}_N}p(\mathbf{x},t,N)\sum_{z\tau\in S^{n-1}_{N,\mathbf{x}^+}}\Theta(\mathbf{x}\to\mathbf{x}+z\tau)\left[g(\mathbf{x},t)+z\sum_{j=1}^{n-1}\tau_j\partial_{x_j}g(\mathbf{x},t)+\frac{z^2}{2}\sum_{k,l=1}^{n-1}\tau_k\tau_l\partial^2_{x_kx_l}g(\mathbf{x},t)+z^3R(\mathbf{x},\tau,t,z)\right].
\end{aligned}
$$

where there is a constant C, depending only on g, such that

$$|R(\mathbf{x},\tau,t,z)| \le C\|\tau\|^3.$$

Using Corollary 1, we obtain the result. □

Equation (9) can be seen as discrete weak formulation for $p(\mathbf{x},t,N)$ in space only, and thus any limiting argument would require some regularity assumption on $p(\mathbf{x},t,N)$ in t. In order to circumvent such assumptions, we need a full discrete weak formulation:

Proposition 3 *Let $T = M\Delta t$, where M is some fixed positive integer, and let g be an admissible test function, with support in $S^{n-1}\times[0,T]$. Let*

$$\mathsf{T} = \{k\Delta t\}, \quad k = 0,\dots,M-1.$$

Then we have that

$$\begin{aligned}
&-\sum_{t\in\mathsf{T}}\sum_{\mathbf{x}\in S_N^{n-1}} p(\mathbf{x},t,N)\left(g(\mathbf{x},t+\Delta t)-g(\mathbf{x},t)\right) - \sum_{\mathbf{x}\in S_N^{n-1}} p(\mathbf{x},t,N)g(\mathbf{x},0) \qquad (10)\\
&= \sum_{t\in\mathsf{T}}\sum_{\mathbf{x}\in S_N^{n-1}} p(\mathbf{x},t,N)\left[\frac{1}{2N}\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial^2_{ij}g(\mathbf{x},t) + (\Delta t)^{\nu}\sum_{j=1}^{n-1} x_j\partial_{x_j}g(\mathbf{x},t)(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x}))\right]\\
&\quad + O\left(N^{-2}(\Delta t)^{-1},(\Delta t)^{2\nu-1},N^{-1}(\Delta t)^{\nu-1}\right).
\end{aligned}$$

Proof Sum (9) over T, and estimate the error term by its total sum, taking into account that there are $O((\Delta t)^{-1})$ terms in this sum. This shows the right hand side of (10). To obtain the left hand side, we perform a summation by parts and use that $g(\mathbf{x},T) = 0$. See appendix B for details. □

4.3 Continuous representation

The aim is now to obtain a continuous version of (10), but *without taking any limits yet*. We first need some preliminary definitions:

Definition 6 (Piecewise time interpolation) *Let T be a set of sampling times as above, and let T_0 be a set of times such that for each $\bar{t}\in\mathsf{T}$, there exists a unique $\xi\in\mathsf{T}_0$ such that $\xi\in(\bar{t},\bar{t}+\Delta t)$. Let g be an admissible test function with support in $S^{n-1}\times[0,T]$. Observe that under the assumptions on the sets T and T_0, for each $t\in[0,T]$ there exists a unique $\bar{t}\in\mathsf{T}$ such that $t\in[\bar{t},\bar{t}+\Delta t)$, and a unique $\xi\in(\bar{t},\bar{t}+\Delta t)$. With this in mind, we define:*

$$\hat{g}(\mathbf{x},t) = g(\mathbf{x},\bar{t}), \quad t\in[\bar{t},\bar{t}+\Delta t), \quad \bar{t}\in\mathsf{T},$$

and

$$\hat{\hat{g}}(\mathbf{x},t) = g(\mathbf{x},\xi), \quad t\in[\bar{t},\bar{t}+\Delta t), \quad \xi\in(\bar{t},\bar{t}+\Delta t), \quad \bar{t}\in\mathsf{T} \text{ and } \xi\in\mathsf{T}_0.$$

Remark 2 *For fixed* $\mathbf{x}$*, we have on one hand that* $\hat{g}(\mathbf{x},t)$ *is just freezing the value of* g *on* $[\bar{t},\bar{t}+\Delta t)$ *to be the value of* $g(\mathbf{x},\bar{t})$*. On the other hand,* $\hat{\hat{g}}(\mathbf{x},t)$ *is freezing the value of* g *on the same interval to be the value of* $g(\mathbf{x},\xi)$*, with* $\xi\in(\bar{t},\bar{t}+\Delta t)$*. The natural choice for* ξ *will arise, in the present context, from applications of the mean value theorem to* g *over the interval* $[\bar{t},\bar{t}+\Delta t]$*.*

Definition 7 (Radonmisation (sic) of discrete densities) *Let* $p(\mathbf{x},t,N)$ *be a probability density defined no* $S_N^{n-1}\times\mathsf{T}$*. Let* $\delta_{\mathbf{x}}$ *denote the atomic measure at* $\mathbf{x}$*. We define*

$$p_N(\mathbf{x},t)=\sum_{\mathbf{y}\in S_N^{n-1}}p(\mathbf{y},\bar{t},N)\delta_{\mathbf{y}}(\mathbf{x}),\quad t\in[\bar{t},\bar{t}+\Delta t).$$

With these definitions we have the following result

Proposition 4 *Let* g *be an admissible test function, let* $N^{-1}=\kappa(\Delta t)^{\mu}$*, where* $\mu>0$ *is a second parameter yet to be specified, and let* $p_{\Delta t}(\mathbf{x},t)=p_{\kappa^{-1}(\Delta t)^{-\mu}}(\mathbf{x},t)$*. Then there exists a set* T_0 *as in Definition 6, such that*

$$\begin{aligned}
&-\int_0^\infty\int_{S^{n-1}}p_{\Delta t}(\mathbf{x},t)\partial_t\hat{\hat{g}}(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t-\int_{S^{n-1}}p_{\Delta t}(\mathbf{x},0)\hat{g}(\mathbf{x},0)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t\\
&\quad=\frac{\kappa(\Delta t)^{\mu-1}}{2}\int_0^\infty\int_{S^{n-1}}p_{\Delta t}(\mathbf{x},t)\left(\sum_{i,j=1}^{n-1}x_i(\delta_{ij}-x_j)\partial_{ij}^2\hat{g}(\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t\\
&\quad\quad+(\Delta t)^{\nu-1}\int_0^\infty\int_{S^{n-1}}p_{\Delta t}(\mathbf{x},t)\left[\sum_{j=1}^{n-1}x_j\left(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)\partial_j\hat{g}(\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t\\
&\quad\quad+O\left((\Delta t)^{2\mu-1},(\Delta t)^{\nu+\mu-1},(\Delta t)^{2\nu-1}\right).
\end{aligned}\tag{11}$$

Proof For the right hand side, we observe that $\hat{g}(\mathbf{x},t)=g(\mathbf{x},t)$ for $\mathbf{x}\in S_N^{n-1}$ and $k\Delta t\le t<(k+1)\Delta t$, $k=0,1,\dots$, and that this also holds for all partial derivatives of g not involving t. On using the definition of $p_{\Delta t}$, we readily obtain the equivalence between the sums over S_N^{n-1} and the integrals in $\mathbf{x}$. For the time integrals, we point out that both $p_{\Delta t}(\mathbf{x},t)$, $\hat{g}(\mathbf{x},t)$ and similarly for the derivatives of g are piecewise constant in t. Hence the summation over time can be exactly converted into a time integral with a factor of $(\Delta t)^{-1}$. As for the left hand side, apply the mean value theorem to $g(\mathbf{x},\cdot)$ to get the result and the set T_0. □

Remark 3 *The reader is cautioned that, although* (11) *has a remarkable resemblance with a weak formulation, it is not quite so, since the prospective test functions* $\hat{g}$ *and* $\hat{\hat{g}}$ *are not test functions in the usual sense.*

4.4 Passage to the limit

We now deal with the limit $\Delta t\to0$ in (11).

Theorem 1 *Under the same assumptions of Proposition 4, we have that, for any choice of parameters* μ *and* ν*, there exists* $p\in L^\infty([0,T],\mathsf{BM}^+\left(S^{n-1}\right)))$*, where* $\mathsf{BM}^+\left(S^{n-1}\right)$

is the set of positive measures of bounded variation on S^{n-1}, such that $p_{\Delta t}(\mathbf{x},t) \to p(\mathbf{x},t)$ weakly as $\Delta t \to 0$. Moreover, the following limits also hold:

$$\int_0^\infty \int_{S^{n-1}} p_{\Delta t}(\mathbf{x},t)\partial_t \hat{\hat{g}}(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \to \int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\partial_t g(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

$$\int_0^\infty \int_{S^{n-1}} p_{\Delta t}(\mathbf{x},t)\left(\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial^2_{ij}\hat{g}(\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t$$
$$\to \int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\left(\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial^2_{ij}g(\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

$$\int_0^\infty \int_{S^{n-1}} p_{\Delta t}(\mathbf{x},t)\left[\sum_{j=1}^{n-1} x_j\left(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)\partial_j\hat{g}(\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t$$
$$\to \int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\left[\sum_{j=1}^{n-1} x_j\left(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)\partial_j g(\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t$$

Proof From the tightness of Radon measures, cf. Billingsley (1999), we have that there exists a sequence $\Delta t_n > 0$, with $\Delta t_n \downarrow 0$ as $n \to \infty$, and $p \in L^\infty([0,T],\mathsf{BM}^+\left(S^{n-1}\right))$, such that

$$\lim_{n\to\infty} p_{\Delta t_n}(\mathbf{x},t) = p(\mathbf{x},t).$$

The convergence of the integrals follows from the weak convergence of $p_{\Delta t_n} \to p$, and from the fact that for a continuous function h, we have

$$\lim_{\Delta t\to 0}\|h-\hat{h}\|_\infty = \lim_{\Delta t\to 0}\|h-\hat{\hat{h}}\|_\infty = 0.$$

□

If either $\mu < 1$ or $\nu < 1$, we can multiply (11) by $(\Delta t)^{-\min(\nu-1,\mu-1)}$. It is then easily verified that the error term vanishes in the limit, as well as the term with a time derivative. Thus, in this case, we obtain stationary limits governed by the steady version of the equations derived below. Now let us assume that $\mu,\nu \geq 1$. It is easily verified that the error term will be small. If both $\mu,\nu > 1$, we have stationary solutions given by the initial condition.

The other cases are as follows:

Theorem 2 *There exists $p \in L^\infty([0,T];\mathsf{BM}^+\left(S^{n-1}\right))$ such that*

If $\mu > 1$, $\nu = 1$, the convective or drift approximation:

$$-\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\partial_t g(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \int_{S^{n-1}} p(\mathbf{x},t_0)g(\mathbf{x},t_0)\,\mathrm{d}\mathbf{x}$$
$$= \int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\left[\sum_{j=1}^{n-1} x_j\left(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)\partial_j g(\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t. \quad (12)$$

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

$$-\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\partial_t g(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \int_{S^{n-1}} p(\mathbf{x},t_0)g(\mathbf{x},t_0)\,\mathrm{d}\mathbf{x}$$
$$= \frac{\kappa}{2}\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\left(\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial_{ij}^2 g(\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t. \tag{13}$$

If $\mu = 1$, $\nu = 1$, *the case where there is a maximal balance of selection and genetic drift; we find the* replicator-diffusion equation

$$-\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\partial_t g(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \int_{S^{n-1}} p(\mathbf{x},t_0)g(\mathbf{x},t_0)\,\mathrm{d}\mathbf{x}$$
$$= \frac{\kappa}{2}\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\left(\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial_{ij}^2 g(\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t \tag{14}$$
$$+\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\left[\sum_{j=1}^{n-1} x_j\left(\psi^{(j)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)\partial_j g(\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t.$$

Proof The result follows from Theorem 1, and from straightforward bookkeeping of the Δt orders of the terms in (11). □

Remark 4 *Alternatively, Theorems 1 and 2 can be seen together as an existence theorem for equations* (12), (13) *and* (14). *Under additional regularity hypothesis on the fitness functions we have uniqueness — see Section 5 — and then the limit is unique.*

Equations (12), (13) and (14) 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*:

$$\partial_t p = -\sum_{i=1}^{n-1} \partial_i\left[x_i\left(\psi^{(i)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)p\right]. \tag{15}$$

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*

$$\partial_t p = \frac{\kappa}{2}\sum_{i,j=1}^{n-1}\partial_{ij}\left((x_i\delta_{ij}-x_ix_j)p\right), \tag{16}$$

which is relevant when the fitness converges to 1 as $\Delta t \to 0$ faster than $N \to \infty$.

– 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:

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

We shall focus on the last equation and on its weak formulation (14).

Remark 5 *We shall see in Section 5 that the weak and the PDE formulations are not equivalent, and that the correct formulation is actually the weak one.*

4.5 Conservation laws from the discrete process

Let us write $\mathfrak{S}$ for the set of all functions $g : S^{n-1} \times [0,+\infty)$ such that there exist an open set $\bar{\Upsilon} \supset S^{n-1}$ and a function $G : \bar{\Upsilon} \times [0,+\infty) \to \mathbb{R}$ such that g is the restriction of G to S^{n-1} and $G \in C^{2,1}(\bar{\Upsilon})$.

Notice that in the right hand side of (14), p is multiplied by:

$$\frac{\kappa}{2}\sum_{i,j=1}^{n-1} D_{ij}\partial^2_{ij}g + \sum_{i=1}^{n-1} \Omega_i\partial_i g = 0. \qquad (17)$$

Equation (17) is readily seen to be a steady backward equation. We now show that the weak solutions have also conservation laws.

Theorem 3 *Let p be a solution to* (14) *(we shall take* (13) *as a special case). Let $\varphi \in \mathfrak{S}$ be in the kernel of* (17)*. Then*

$$\int_{S^{n-1}} p(\mathbf{x},t)\varphi(\mathbf{x})\,\mathrm{d}\mathbf{x} = \int_{S^{n-1}} p(\mathbf{x},0)\varphi(\mathbf{x})\,\mathrm{d}\mathbf{x},$$

for almost every $t \in [0,\infty)$.

Proof Let $\eta(t) \in C_c([0,\infty))$, with $\eta(0) = 1$. Then

$$g(\mathbf{x},t) = \eta(t)\varphi(\mathbf{x})$$

is an admissible test function. On substituting in (14), we find that

$$\int_0^\infty \int_{S^{n-1}} p(\mathbf{x},t)\varphi(\mathbf{x})\eta'(t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t + \int_{S^{n-1}} p(\mathbf{x},0)\varphi(\mathbf{x})\,\mathrm{d}\mathbf{x} = 0.$$

Since η is an arbitrary function with compact support in $[0,\infty)$, the result follows. □

A similar argument shows also the following

Theorem 4 *Let p be a solution to* (12)*. Then*

$$\int_{S^{n-1}} p(\mathbf{x},t)\,\mathrm{d}\mathbf{x} = \int_{S^{n-1}} p(\mathbf{x},0)\,\mathrm{d}\mathbf{x},$$

for almost every $t \in [0,\infty)$.

Therefore, the conservation laws given by equation (4) now become

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{S^{n-1}} p(t,x)\varphi(x)\mathrm{d}x = 0, \tag{18}$$

where φ satisfies (17). In principle the condition set out by (18) seems to imply an infinite (likely to be uncountable) number of conservation laws. The following result shows that it is actually much more conspicuous:

Theorem 5 *Let $\mathbf{e}_i$ denote the vertices of S^{n-1}. Then there exist unique ρ_i, $i=1,\dots,n-1$, with $\rho_i(\mathbf{e}_j)=\delta_{ij}$ that are solutions to* (17). *In addition, let $\rho_0\equiv 1$. Then, any solution to* (17) *in $\mathfrak{S}$ can be written as a linear combination of ρ_i, $i=0,\dots,n-1$. In particular its kernel, for solutions in $\mathfrak{S}$, has dimension n.*

Proof Given a vertex $\mathbf{e}_i$, let $\mathbf{e}_j$ be an adjacent vertex. Now we solve (17) in the segment $\overline{\mathbf{e}_j\mathbf{e}_i}$ with boundary values δ_{ij}. In the segments not adjacent to $\mathbf{e}_i$ define the solution to be zero. This defines the solution in all one-dimensional simplices. For each two-dimensional subsimplex, we now solve the Dirichlet problem with the data from the previous step. Now, assume that we have the solution uniquely defined in all subsimplices of dimension m. Repeating the construction above yields the solution in all subsimplices of dimension $m+1$. Proceeding inductively, this yields a solution in S^{n-1} that is unique and admissible. Uniqueness follows from the maximum principle applied at each subsimplex level. Let $\varphi\in\mathfrak{S}$, and let

$$\Phi(\mathbf{x}) = \sum_{i=1}^{n}\varphi(\mathbf{e}_i)\rho_i(\mathbf{x}).$$

Then $\Phi\in\mathfrak{S}$. By the preceding argument, Φ and φ must agree at all edges of S^{n-1}. Proceeding inductively once again yields that $\varphi=\Phi$ in S^{n-1}. □

Thus, the solutions to (14) must satisfy:

$$\frac{\mathrm{d}}{\mathrm{d}t}\int_{S^{n-1}}\rho_i(\mathbf{x})p(\mathbf{x},t)\mathrm{d}\mathbf{x} = 0\ , \quad i=1,\cdots,n\ . \tag{19}$$

From a probabilistic viewpoint, the ρ_i, $i=1,\dots,n$, are naturally identified with the fixation probability of type i. We now give a pure analytical argument for this fact. In Section 5, we shall prove Theorem 7 which shows that the final state is given by

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

where $\delta_{\mathbf{e}_i}$ is a Dirac measure supported on the vertex $\mathbf{e}_i\in S_N^{n-1}$. Clearly, $\pi_i[p^{\mathrm{I}}]$ is the fixation probability of type i in a population initially described by a probability distribution p^{I}.

Therefore,

$$\pi_i[\delta_{\mathbf{x_0}}] = \int\rho_i(\mathbf{x})p^{\infty}(\mathbf{x})\mathrm{d}\mathbf{x} = \int\rho_i(\mathbf{x})p^{\mathrm{I}}(\mathbf{x})\mathrm{d}\mathbf{x} = \int\rho_i(\mathbf{x})\delta_{\mathbf{x}_0}(\mathbf{x})\mathrm{d}\mathbf{x} = \rho_i(\mathbf{x}_0).$$

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 *$\pi_i^{\mathrm{N}}[\delta_{\mathbf{x}}]=x_i$, which follows from the fact that in the neutral case, $\rho_i(\mathbf{x})=x_i$.*

5 The Replicator-Diffusion approximation

We now discuss the nature of solutions p to (1') together with the conservation laws (19). The main result of this section is Theorem 7. This must be understood as the continuous counterpart of the Lemma 1. We do not refer to the discrete 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 (19).

An outline of the proof of Theorem 7 is as follows: First, we show that a solution to (14) can be written as regular part plus a singular measure over the boundary. Moreover, the regular part vanishes for large time. Repeating these arguments over the lower dimensional subsimplices, and using the projection result in Proposition 5, we arrive at a representation of p as a sum of its classical solution and a sum of singular measures that are uniformly supported on the descending chain of subsimplices of S^{n-1} down to the zeroth dimension. Since the solutions over the subsimplices also have a regular part that vanishes, we can show that all measures that are not atomically supported at the vertices should vanish for large time. Thus, conservation of probability implies that the steady state of (14) is a sum of deltas.

Finally, we 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 at time t of a given type (to be prescribed by the boundary conditions in the dual process) for a population initially at state $\mathbf{k}$. 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 (14) 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. Throughout this section, we shall have the further assumption that the fitnesses are smooth.

We start by showing that, in the interior, a weak solution is regular enough to be a classical solution.

Lemma 3 *Let p be a solution to (14). Let $K\subset S^{n-1}$ be a proper compact subset. Then, in K, p satisfies (1') in the classical sense. In particular, $p\in C([0,T];C^\infty(\mathrm{int}(S)))$.*

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). The last statement follows from $\mathrm{int}(S) = \cup_{K\subset S} K$, with K compact and $K \cap \partial S^{n-1} = \emptyset$. □

The next two Lemmas show existence of a unique classical solution, and that such a solution decays to zero for large time.

Lemma 4 *Let p be a classical solution to (1'). Then*

$$\lim_{t\to\infty} p(\mathbf{x},t) = 0, \quad \mathbf{x} \in \mathrm{int}\, S^{n-1}.$$

Proof We define $\mu_{\mathrm{S}}(\mathbf{x}) = x_1 x_2 \cdots x_n$ (such that $\mu_{\mathrm{S}}(\mathbf{x}) \geq 0$ in S^{n-1} with $\mu_{\mathrm{S}} = 0$ if and only if $\mathbf{x} \in \partial S^{n-1}$). Note that

$$\sum_{j=1}^{n-1} \partial_j \left(\frac{D_{ij}}{\mu_{\mathrm{S}}} \right) = \mu_{\mathrm{S}}^{-1} \left[\sum_{j=1}^{n-1} (\delta_{ij} - \delta_{ij} x_j - x_i) - \sum_{j=1}^{n-1} (x_j \delta_{ij} - x_i x_j) \left(\frac{1}{x_j} - \frac{1}{x_n} \right) \right] = 0\,. \tag{20}$$

We introduce the new variable $u = \mu_{\mathrm{S}} p$ and after some manipulations, we find

$$\partial_t u = \mu_{\mathrm{S}} \nabla \cdot \left[\mu_{\mathrm{S}}^{-1} \left(\frac{\kappa}{2} D \nabla u - \Omega u \right) \right]\,, \tag{21}$$

with $D = (D_{ij})_{i,j=1,\dots,n-1}$ and $\Omega = \sum_{i=1}^{n-1} \Omega_i \mathbf{e}_i$.

We now show that the last equation is well defined in S^{n-1}. For the second order term, this follows from a new application of equation (20). For the first order term, note that

$$\mu_{\mathrm{S}} \nabla \cdot \left(\frac{\Omega u}{\mu_{\mathrm{S}}} \right) = \nabla \cdot \Omega u - \frac{\Omega \cdot \nabla \mu_{\mathrm{S}}}{\mu_{\mathrm{S}}} u + \Omega \cdot \nabla u\,.$$

Furthermore

$$\frac{\Omega \cdot \nabla \mu_{\mathrm{S}}}{\mu_{\mathrm{S}}} = \sum_{i=1}^{n-1} \Omega_i \left(\frac{1}{x_i} - \frac{1}{x_n} \right) = \sum_{i=1}^{n} \left(\psi^{(i)}(\mathbf{x}) - \bar{\psi}(\mathbf{x}) \right)\,.$$

We shall now study the eigenvalue problem associated to (21) by considering the dual problem, with respect to the measure $(\mu_S)^{-1} \mathrm{d}\mathbf{x}$, and with regularised coefficients:

$$\begin{cases} \mu_{\mathrm{S}}^{(\varepsilon)} \nabla \cdot \left[\frac{\kappa}{2\mu_{\mathrm{S}}^{(\varepsilon)}} D^{(\varepsilon)} \nabla \varphi^{(\varepsilon)} \right] + s\Omega \cdot \nabla \varphi^{(\varepsilon)} = \lambda^{(\varepsilon)} \varphi^{(\varepsilon)}, \\ \varphi^{(\varepsilon)} = 0 \text{ in } \partial S^{n-1}\,, \end{cases} \tag{22}$$

where $D^{(\varepsilon)}(\mathbf{x})$ is a positive defined matrix in S^{n-1}, with $D^{(\varepsilon)} \xrightarrow{\varepsilon\to 0^+} D$ uniformly in $\mathbf{x}$, and $\mu_{\mathrm{S}}^{(\varepsilon)} > 0$ in S^{n-1}, $\mu_{\mathrm{S}}^{(\varepsilon)} \xrightarrow{\varepsilon\to 0^+} \mu_{\mathrm{S}}$ uniformly in $\mathbf{x}$, and s is a real parameter.

First we observe that, for $\varepsilon \geq 0$, equation (22) satisfies a maximum principle for solutions in $C^2(\mathrm{int}(S^{n-1}))$; therefore if $\lambda^{(\varepsilon)} = 0$ we have $\varphi^{(\varepsilon)} = 0$ in S^{n-1}; see Crandall et al. (1992). We conclude that $\lambda^{(\varepsilon)} \neq 0$, for $s \in \mathbb{R}$ and $\varepsilon \geq 0$. Additionally, since the coefficients are smooth, the solution to equation (22) is smooth in the interior by standard elliptic regularity.

For $\varepsilon > 0$, the dominant eigenvalue $\lambda_0^{(\varepsilon)}$ is real and from the maximum principle it follows that $\lambda_0^{(\varepsilon)} \neq 0$. For $s = 0$, $\lambda_0^{(\varepsilon)}$ is negative and therefore from its continuity in s, we conclude that $\lambda_0^{(\varepsilon)} < 0$ for any s. Therefore, for any other eigenvalue $\mathrm{Re}\left(\lambda^{(\varepsilon)}\right) \leq \lambda_0^{(\varepsilon)} < 0$ (see (Evans 2010) for further details).

Moreover, let $\varepsilon_k \to 0$ be a decreasing sequence of positive numbers, and $\varphi^{(\varepsilon_k)} \geq 0$ be the normalised eigenfunctions for the corresponding leading eigenvalues. Since the coefficients are assumed smooth, the eigenfunctions are also smooth. Hence, by Rellich theorem, there is a subsequence ε_{k_j} such that $\varphi^{(\varepsilon_{k_j})}$ converges in $L^2(S^{n-1})$. By considering the weak formulation for equation (22), we immediately see that, for this subsequence, we must also have $\lambda_0^{(\varepsilon_{k_j})} \to \lambda_0$.

We thus have obtained a real negative eigenvalue λ_0, with a real eigenfunction that is single signed. Using the same argument in Evans (2010), we conclude that λ_0 is the principal eigenvalue for the nonregularised problem. Hence, any other eigenvalue will satisfy $\mathrm{Re}(\lambda) \leq \lambda_0$.

This also shows that there exists $\alpha > 0$, such that

$$\frac{1}{2}\partial_t \int_{S^{n-1}} u^2 \mu_S^{-1} d\mathbf{x} = \int_{S^{n-1}} \mu_\mathrm{S} \nabla \cdot \left[\mu_\mathrm{S}^{-1}\left(\frac{1}{2}D\nabla u - \Omega u\right)\right] u\, \mu_\mathrm{S}^{-1} \mathrm{d}\mathbf{x} < -\alpha \int_{S^{n-1}} u^2\, \mu_\mathrm{S}^{-1} \mathrm{d}\mathbf{x}.$$

Therefore

$$\int p^2 \mu_\mathrm{S} \mathrm{d}\mathbf{x} = \int u^2 \mu_\mathrm{S}^{-1} \mathrm{d}\mathbf{x} \overset{t\to\infty}{\to} 0\ .$$

□

Lemma 5 *Equation (21) has a unique solution* $u \in C\left([0,\infty); C^\infty(\mathrm{int}\left(S^{n-1}\right)\right)$.

Proof Consider equation (21) with $D = D^{(\varepsilon)}$ and $\mu_\mathrm{S} = \mu_\mathrm{S}^{(\varepsilon)}$, with $D^{(\varepsilon)}$ and $\mu_\mathrm{S}^{(\varepsilon)}$ as in Lemma 4. For $\varepsilon > 0$, it is uniformly parabolic, and hence it has a unique solution with the required regularity.

We write (21) in weak form as

$$\begin{aligned}
&\int_0^\infty \int_{S^{n-1}} u^{(\varepsilon)}(t,x)\partial_t \phi(t,x) \left(\mu_\mathrm{S}^{(\varepsilon)}\right)^{-1} \mathrm{d}\mathbf{x}\mathrm{d}t \\
&\quad + \int_0^\infty \int_{S^{n-1}} \left(\mu_\mathrm{S}^{(\varepsilon)}\right)^{-1} \left(\frac{1}{2}D\nabla u^{(\varepsilon)} - \Omega u^{(\varepsilon)}\right) \cdot \nabla\phi(t,x) \mathrm{d}\mathbf{x}\mathrm{d}t \\
&\quad + \int_{S^{n-1}} u^{(\varepsilon)}(0,x)\phi(0,x) \left(\mu_\mathrm{S}^{(\varepsilon)}\right)^{-1} \mathrm{d}\mathbf{x} = 0.
\end{aligned}$$

We now observe that any such solution is bounded in $\aleph = L^2((0,T); W_0^{2,1})$. Hence, one can select a sequence $\varepsilon_k \downarrow 0$ such that $u^{(\varepsilon_k)} \to u^* \in \aleph$. Since (21) is weakly parabolic, such a solution must be unique — Lieberman (1996). Finally, regularity follows from Lemma 3. □

The preceding two lemmas have an important consequence:

Corollary 2 *No solution to (1') in the classical sense can satisfy the required conservation laws. In particular, this shows that the weak-formulation presented in Theorem 2 is not only a device for obtaining the continuous limit, but it turns out to be the correct formulation.*

In view of Corollary 2, we turn back to the weak formulation given by (14) or, equivalently, to (1') together with the conservation laws (19). In what follows, we obtain some more information about such solutions.

Remark 7 *As an extension of Lemma 4.1 in Chalub and Souza (2009b), we observe that if p is a Radon measure in S^{n-1}, we can write $p = q + r$, with* $\operatorname{sing\,supp}(q) \in \partial S^{n-1}$ *and* $\operatorname{sing\,supp}(r) \in \operatorname{int} S^{n-1}$.

Proposition 5 (Face Projections) *Let $1 \leq k < n$ and let p be a solution to* (14)*. Let S^{k-1} be a face of S^k. Assume that* $\operatorname{sing\,supp}(p) \cap S^{k-1} \neq \emptyset$*. Then, over S^{k-1}, p satisfies* (14) *in one less dimension with forcing given by the regular part of p evaluated at $x_i = 0$, for a certain value of i.*

Proof Assume, without loss of generality, that $i = 1$. In view of remark 7, we can write $p = q + r$, where $\operatorname{sing\,supp}(q) \subset S^{n-2}$ with the singular support of r lying in the complement with respect to the full simplex. Moreover, we can also assume, without loss of generality, that S^{n-2} is given by the intersection of the hyperplane $x_1 = 0$ with S^{n-1}. Let us write $\mathbf{x} = (x_2, \ldots, x_n)$. Let h be an appropriate test function in S^{n-2}, satisfying $h(\mathbf{x}, 0) = 0$ and let $\eta(x_1) \in C_c([0,1])$, with $\eta(0) = 1$. Then $g = \eta h$ is an appropriate test function for S^{n-1} and a direct computation with (14) then yields

$$
\begin{aligned}
&-\int_0^\infty \int_{S^{n-1}} p(x_1,\mathbf{x},t)\partial_t g(x_1,\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&\quad= \frac{\kappa}{2}\int_0^\infty \int_{S^{n-1}} p(x_1,\mathbf{x},t)\left(\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial_{ij}^2 g(x_1,\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&\qquad+\int_0^\infty \int_{S^{n-1}} p(x_1,\mathbf{x},t)\left[\sum_{j=1}^{n-1} x_j\left(\psi^{(j)}(x_1,\mathbf{x}) - \bar{\psi}(x_1,\mathbf{x})\right)\partial_j g(x_1,\mathbf{x},t)\right]\mathrm{d}\mathbf{x}.
\end{aligned}
$$

Over $x_1 = 0$, on using the definition of g, we find that

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

For r, we have

$$
\begin{aligned}
&-\int_0^\infty \int_{S^{n-1}} r(x_1,\mathbf{x},t)\partial_t g(x_1,\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&= \frac{\kappa}{2}\int_0^\infty \int_{S^{n-1}} r(x_1,\mathbf{x},t)\left(\sum_{i,j=1}^{n-1} x_i(\delta_{ij}-x_j)\partial^2_{ij}g(x_1,\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&\quad+\int_0^\infty \int_{S^{n-1}} r(x_1,\mathbf{x},t)\left[\sum_{j=1}^{n-1} x_j\left(\psi^{(j)}(x_1,\mathbf{x})-\bar{\psi}(x_1,\mathbf{x})\right)\partial_j g(x_1,\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t.
\end{aligned}
$$

By Lemma 3, r is smooth. Therefore, the above equation can be integrated by parts to yield, an integral on S^{n-1} that will cancel out identically, since r is a classical solution to (1'), and a number of integrals over the various faces of S^{n-1}. In particular, at $x_1 = 0$, we find that

$$
0 = -\frac{\kappa}{2}\int_0^\infty \int_{S^{n-2}} r(0,\mathbf{x},t)h(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t.
$$

By collecting together the two calculations on $x_1 = 0$, we obtain the result. □

In what follows, we shall need some preliminaries. Recall — see Stanley (1996) — that to the simplex S^{n-1} is associated a corresponding f-vector, such that the entry $i+1$ (f_{i+1}) is the number of i-dimensional subsimplices of S^{n-1}. We shall assume that, for each dimension i, there is a definite order of the subsimplices $S^{i,j}$, with $i = 0,\ldots,n-1$ and $j = 1,\ldots,f_{i+1}$. Moreover, we define the adjacent operator by $\mathrm{ad}(j,k)$ which denotes the kth adjacent subsimplex of dimension $i+1$ to S^{ij}. Notice that there are $n-i$ such simplexes.

Theorem 6 (Solution Structure) *Equation* (14)*, with a given initial condition* $p^{\mathrm{I}} \in \mathsf{BM}^+\left(S^{n-1}\right)$*, has a unique solution* $p \in L^\infty\left([0,T];\mathsf{BM}^+\left(S^{n-1}\right)\right)$*. Moreover, let* δ^{ij} *be the Radon measure with unit mass uniformly supported on* S^{ij}*. Then the solution* p *can be written as*

$$
p(t,x) = p_{n1} + \sum_{(i,j)\in\mathscr{I}} p_{ij}\delta^{ij}, \quad \mathscr{I} = \left(\cup_{i=0}^{n-1}\{i\}\right)\times\{1,\ldots,f_{i+1}\}, \tag{23}
$$

where p_{ij} *satisfies*

$$
\begin{aligned}
&-\int_0^\infty \int_{S^{ij}} p_{ij}(\mathbf{x},t)\partial_t g(\mathbf{x},t)\,\mathrm{d}\mathbf{x}\,\mathrm{d}t - \int_{S^{ij}} p_{ij}(\mathbf{x},t_0)g(\mathbf{x},t_0)\,\mathrm{d}\mathbf{x} \\
&= \frac{\kappa}{2}\int_0^\infty \int_{S^{ij}} p_{ij}(\mathbf{x},t)\left(\sum_{r,s=1}^{n-1} x_r(\delta_{rs}-x_s)\partial^2_{rs}g(\mathbf{x},t)\right)\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&\quad+\int_0^\infty \int_{S^{ij}} p_{ij}(\mathbf{x},t)\left[\sum_{s=1}^{n-1} x_s\left(\psi^{(s)}(\mathbf{x})-\bar{\psi}(\mathbf{x})\right)\partial_s g(\mathbf{x},t)\right]\mathrm{d}\mathbf{x}\,\mathrm{d}t \\
&\quad+\int_0^\infty \int_{S^{ij}} \sum_{k=1}^{n-i} p_{(i+1)\mathrm{ad}(j,k)}\big|_{S^{ij}}\,\mathrm{d}\mathbf{x}\,\mathrm{d}t.
\end{aligned} \tag{24}
$$

The initial condition for (24) *will be denoted* p^{I}_{ij}*, and it is obtained from* p^{I} *by applying the decomposition described in Remark 7 recursively.*

Proof By direct substitution into (14) and after a integrating by parts starting from S^{n-1}, proceeding downwards until the vertices and using Lemma 5 one can verify that (23) is indeed a solution. To verify uniqueness, let $\tilde{p}$ a solution to (14). Consider $\tilde{p}$ restricted to S^{n-1} and let p_{n1} be the classical solution guaranteed by Lemma 5. By considering (14) with test functions with compact support on $\operatorname{int} S^{n-1}$, we see that $\tilde{p}-p_{n1}$ vanishes. Therefore sing $\operatorname{supp}(\tilde{p}-p_{n1}) \subset \partial S^{n-1}$. By Remark 7, we can write $\tilde{p}=p_{n1}+q$, with sing $\operatorname{supp}(q) \subset \partial S^{n-1}$. Now ∂S^{n-1} is the union of f_{n-1} copies of S^{n-2}. By Proposition 5, q must satisfies (24) in one less dimension in each of the subsimplices. Proceeding inductively, we can now choose a subsimplex S^{n-3} of S^{n-2}. Now, we repeat the argument above for each simplex S^{n-2} which has S^{n-3} as a subsimplex. Iterate until arrive at the simplices of zero dimension to get the result.

□

This theorem leads to the following result:

Theorem 7 (Final State) *Let*

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

where p is the solution of equation (1) subject to conservation laws (19). Then p^{∞} is a linear combination of point masses at the vertices of S^{n-1},i.e,

$$p^{\infty} = \sum_{i=1}^{n} \pi_i \left[p^{\mathrm{I}}\right] \delta_{\mathbf{e}_i} \ . \tag{25}$$

Proof First, we observe that Lemma 4 still holds if applied to inhomogeneous version of (21), provided that the forcing decays for large times. The results now follows from a straightforward application of Proposition 5 together with Lemma 4 applied in a descending chain of simplices down to dimension 1. Conservation of probability then yields that p^{∞} must be a sum of atomic measures at the vertices of S^{n-1}. On using the other conservation laws, we obtain the coefficients, and hence the result.

□

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:

$$\partial_t f = \mathscr{L}^{\dagger}_{n-1,k} f := \frac{\kappa}{2} \sum_{i,j=1}^{n-1} D_{ij} \partial^2_{ij} f + \sum_{i=1}^{n-1} \Omega_i \partial_i f \ . \tag{26}$$

In diffusion theory this equation is associated with a martingale problem for the diffusive continuous process. In genetics, the meaning of equation (26) 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.,

$p^{\mathrm{I}}(\mathbf{x}) := p(\mathbf{x},0) = \delta_{\mathbf{k}}(\mathbf{x}))$ we find the population fixed at the homogeneous state $\mathbf{e}_i$ at time t (or before), i.e., $f_i(\mathbf{k},t) = \langle p(\cdot,t), \delta_{\mathbf{e}_i}\rangle$. In this case, we need to find consistent boundary conditions. See Maruyama (1977); Etheridge (2011).

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, $f_i\big|_{V_1} = 0$. For $i \neq 1$, $f_i\big|_{V_i}$ is the solution of $\partial_t f = \mathscr{L}^{\dagger}_{n-2,k} f$, 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 $f\big|_{L_{1i}} : [0,1] \to \mathbb{R}$, the restriction of f_i to this edge, with x_1 being the fraction of type 1 individuals, is the solution of

$$\partial_t f = \frac{\kappa}{2} x_1(1-x_1)\partial_1^2 f + x_1(1-x_1)\left(\psi^{(1)}_{1i}(x_1) - \psi^{(i)}_{1i}(x_1)\right)\partial_1 f \qquad (27)$$

with boundary conditions given by $f(0) = 0$ and $f(1) = 1$ and $\psi^{(j)}_{1i}(x_1) = \psi^{(j)}(x_1\mathbf{e}_1 + (1-x_1)\mathbf{e}_i)$ is the restriction of $\psi^{(j)}$ to the edge L_{1i}. The forward and backward versions of Equation (27) are fully studied in the references (Chalub and Souza 2009b,a). For $\psi^{(1)}_{1i} - \psi^{(i)}_{1i}$ 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 8 *If, for all states $\mathbf{x} \in S^{n-1}$, and all types $i = 1, \ldots, n$, $\psi^{(1)}(\mathbf{x}) \geq \psi^{(i)}(\mathbf{x})$, then the fixation probability of the first type is not less than the neutral fixation probability for any initial condition p^{I}; i.e,*

$$\pi_1[p^{\mathrm{I}}] \geq \pi_1^{\mathrm{N}}[p^{\mathrm{I}}] \ .$$

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

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

with vertex conditions $\rho_1(\mathbf{e}_i) - x_1(\mathbf{e}_i) = 0$ for $i = 1, \ldots, n$. Now, we proceed by induction in n. For the case $n = 2$, the proof is in (Chalub and Souza 2009a, 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 $\psi^{(1)}(x)-\bar{\psi}(x)=(1-x)\left(\psi^{(1)}(x)-\psi^{(2)}(x)\right)$ and the solution is given by

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

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

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

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 sub-harmonic 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 $\mathbf{x}\in S^{n-1}$. □

6 The Replicator Dynamics

In Section 4, we proved that, when genetic drift and selection balance, then there is a special timescale such that the evolution of an infinite population can be described by a parabolic partial differential equation. Nevertheless, in applications one is usually interested in large but finite populations. In this case, an exact limit is not taken, and (14) can be taken as an approximation of this evolution. We shall discuss this further in the conclusions, but we observe that this equation might be a good approximation even when balance is not exact, i.e., when ν and μ are close but not equal to one. This could typically lead to an equation with κ being either quite large or small. In the former case, a regular expansion in κ shows that the evolutions is governed by (13). On the other hand, in the latter case, one expects that the much simpler transport equation (15) will be a good approximation for the evolution. Indeed, in this section we show that (1') can be uniformly approximated by (15) in proper compact subsets of the simplex, and over a time interval shorter than κ^{-1}.

We start in Subsection 6.1 showing that the equation (15) is formally equivalent to the replicator system. Afterwards, in Subsection 6.2, we answer what we believe to be an important 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 conditional on the absence of extinctions. 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 κ is small. As in Section 5, we shall assume that the fitness functions are smooth.

6.1 The replicator ODE and PDE

We shall now study in more detail the equation (15), which has a close connection with the replicator dynamics as shown below:

Theorem 9 *Assume that Ω is Lipschitz. Let $\Phi_t(\mathbf{x})$ the flow map of*

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t} = \Omega(\mathbf{x}(t)). \tag{28}$$

and let

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

Let $p^I \in \mathsf{BM}^+\left(S^{n-1}\right)$ and assume that $\operatorname{sing\,supp}(p^I) \subset \operatorname{int}(S^{n-1})$ *(see Remark 7). Then the solution to* (15) *with initial condition p^I is given by*

$$p(\mathbf{x},t) = \mathrm{e}^{Q(\mathbf{x},t)} p^{\mathrm{I}}\left(\Phi_{-t}(\mathbf{x})\right). \tag{29}$$

Proof We observe that, since Ω is Lipschitz, the push-forward of p^I by Φ_t is well defined—(Ambrosio et al. 2005). Hence, the proof is based on the methods of characteristics; see John (1991); Evans (2010). See DiPerna and Lions (1989) for an approach that works even if Ω fails to be Lipschitz continuous. □

Remark 8 *If p^{I} gives mass to the boundaries of S^{N-1}, we write, as in Theorem 6:*

$$p^{\mathrm{I}} = \sum_{i,j} p^{\mathrm{I}}_{i,j}, \quad \textit{with} \quad \operatorname{sing\,supp}(p^{\mathrm{I}}_{i,j}) \subset S^{i,j}.$$

Moreover, notice that Ω restricted to $S^{i,j}$ is tangent to $\partial S^{i,j}$. Hence, the restricted dynamics is always well defined, and we can write $\Phi^{i,j}_t$ for the flow map of

$$\frac{\mathrm{d}\mathbf{x}^{i,j}}{\mathrm{d}t} = \Omega^{i,j}(\mathbf{x}^{i,j}(t)) \quad \textit{and} \quad Q^{i,j}(\mathbf{x},t) = -\int_0^t \left(\nabla^{i,j}\cdot\Omega^{i,j}\right)\left(\Phi^{i,j}_{s-t}(\mathbf{x})\right)\mathrm{d}s,$$

where $\Omega^{i,j}$ is the restriction of Ω to $S^{i,j}$, and $\nabla^{i,j}\cdot\Omega^{i,j}$ is divergence in $S^{i,j}$. Then, repeated applications of Theorem 9 lead to the conclusion that the solution to (15) *is given by*

$$p(\mathbf{x},t) = \sum_{i,j} \mathrm{e}^{Q^{i,j}(\mathbf{x},t)} p^{\mathrm{I}}_{i,j}(\Phi^{i,j}_{-t}(\mathbf{x})).$$

6.2 Peak and average dynamics

We start by showing that the long term 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 $\psi^{(i)}: S^{n-1}\to\mathbb{R}$.

From the fact that the final state of the the replicator-diffusion equation is given by equation (25), the coefficients $\pi_i[p^{\mathrm{I}}]$, $i=1,\dots,n$ can be calculated in two ways:

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

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) an ESS that lies in interior of S^{n-1} must be a global attractor of the replicator equation (28). We have then the following result

Theorem 10 *Assume* $p^{\mathrm{I}} \in \mathsf{BM}^{+}\left(S^{n-1}\right)$, sing supp $p^{\mathrm{I}} \subset \operatorname{int} S^{n-1}$ *and assume that (28) has a unique point* $\mathbf{x}^*$ *such that for any initial condition* $\mathbf{x}(0) \in \operatorname{int} S^{n-1}$, $\lim_{t\to\infty} \mathbf{x}(t) = \mathbf{x}^*$. *Then the solution of equation (15) is such that*

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

Proof Assume, initially, that $\mathbf{x}^* \in \operatorname{int} S^{n-1}$. Since $\mathbf{x}^*$ is a globally stable equilibrium for interior initial points, we can find $T > 0$, such that, for $t > T$, and sufficiently small $\delta > 0$, we have, for any proper compact subset $K \subset S^{n-1}$, that:

$$\Phi_t(K) \subset B_\delta(\mathbf{x}^*) \subset \operatorname{int} S^{n-1}.$$

where $B_\delta(\mathbf{x}^*)$ is the open ball of radius δ and centered at $\mathbf{x}^*$.

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

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

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

$$\eta(\mathbf{x}^*) - \varepsilon \le \int_{B_\delta(\mathbf{x}^*)} p(\mathbf{x},t)\eta(\mathbf{x})\,\mathrm{d}\mathbf{x} \le \eta(\mathbf{x}^*) + \varepsilon, \tag{30}$$

Now take $(\delta_k, \varepsilon_k) \downarrow 0$ such that (30) is satisfied. This yields a sequence of times T_k such that $T_k \to \infty$ and

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

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

For the case $\mathbf{x}^* \in \partial S^{n-1}$, the result follows from similar arguments, replacing $B_\delta(\mathbf{x}^*)$ by $B_\delta(\mathbf{x}^*) \cap S^{n-1}$. □

6.3 Asymptotic approximation

Let

$$0 < \kappa \ll 1.$$

If we perform a regular asymptotic expansion, i.e., if we write $p_\kappa \approx p_0 + \kappa p_1 + \cdots$, then we find, for times $t \ll \kappa^{-1}$, that the leading order dynamics is given by

$$\partial_t p_0 + \nabla \cdot (p_0 \Omega) = 0. \quad (31)$$

The next theorem shows that this indeed the case, provided we see p_0 as the leading order dynamics with respect to the regular part of the probability density.

Theorem 11 *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 r_κ be the regular part of the solution of (1), with $\kappa \geq 0$. Then p_0 is $C^2(S^{n-1})$, and satisfies the conservation law (19). Moreover, if $\nabla \cdot \Omega \geq 0$, then given κ and K positive, there exits a C such that, for $t \ll C\kappa^{-1}$, we have*

$$\|r_\kappa(\cdot,t) - p_0(\cdot,t)\|_\infty \leq C\kappa$$

and

$$\|\partial_x^2 p_0(\cdot,t)\|_\infty > K$$

Thus p_0 is the leading order asymptotic approximation to r_κ, for $t \ll \kappa^{-1}C$.

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

Let $w_\kappa = r_\kappa - p_0$. Then w_κ satisfies

$$\partial_t w_\kappa = \frac{\kappa}{2}\sum_{i,j=1}^{n-1} \partial_{ij}^2 (D_{ij} w_\kappa) - \sum_{i=1}^{n-1} \partial_i (\Omega_i w_\kappa) + \frac{\kappa}{2} g_0(\mathbf{x},t)$$

with null initial condition, where

$$g_0(\mathbf{x},t) = \sum_{i,j=1}^{n-1} \partial_{i,j}^2 (D_{ij} p_0).$$

Notice that, because of the assumptions on p^{I}, 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

$$w_\kappa(\mathbf{x},t) = \frac{\kappa}{2}\int_0^t S(t,s) g_0(s,x)\mathrm{d}s.$$

By the maximum principle applied to the semigroup $S(t_2,t_1)$, we have that $\|S(t,s)g_0(s,x)\| \leq M_s$, 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

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

Hence

$$|w_\kappa(\mathbf{x},t)| \leq \kappa t \frac{M}{2}.$$

Therefore, taking $C = 2M^{-1}$, we find, for $t \ll C\kappa^{-1}$, that:

$$\|w_\kappa(t,\cdot)\|_\infty \ll 1.$$

□

Remark 9 *If the condition on $\nabla \cdot \Omega$ is not satisfied, a similar proof shows that if $t \ll -\log(\kappa)$ then the same conclusion holds. Notice also that this condition is satisfied if the replicator has a globally stable equilibrium in the interior of S^{n-1}.*

Proposition 6 *Under the same hypothesis of Theorem 11, let U be an open set such that $U \subset S^{n-1}$ and $\bar{U} \cap \partial S^{n-1} = \emptyset$. Then, there exists $C > 0$, such that*

$$\left| \int_U (p_\kappa(\mathbf{x},t) - p_0(\mathbf{x},t))\, \mathrm{d}\mathbf{x} \right| < C\kappa t,$$

for any t.

Proof By Theorem 11, there exists $C' > 0$ such that $\|r_\kappa(\cdot,t) - p_0(\cdot,t)\|_\infty < C'\kappa t$, which we write as:

$$-C'\kappa t \leq r_k(\mathbf{x},t) - p_0(\mathbf{x},t) < C'\kappa t.$$

Integrating in U and using that sing $\mathrm{supp}(q_\kappa) \cap U = \emptyset$, the result follows.

Theorem 10 shows that, for sufficient large time, the support of the solution of the replicator PDE, equation (15), will be concentrated in sufficiently small neighbourhoods 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 7); however, for strong selection, the initial dynamics given by the replicator-diffusion equation is similar to the one given by the replicator ODE. This is justified by the following result:

Theorem 12 *Assume that the replicator has a unique global attractor. Then, under the same hypothesis of Theorem 11, we have that given $\varepsilon > 0$ and $\delta > 0$ there exist a time t^* and a constant $C > 0$, depending only on the initial condition, such that*

$$\left| \int_{B_\varepsilon(\mathbf{x}_*)} p_\kappa(\mathbf{x},t)\, \mathrm{d}\mathbf{x} - 1 \right| < C\kappa t + \delta,$$

for $t > t^*$.

Proof We have

$$\left| \int_{B_\varepsilon(\mathbf{x}_*)} p_\kappa(\mathbf{x},t)\, \mathrm{d}\mathbf{x} - 1 \right| \leq \left| \int_{B_\varepsilon(\mathbf{x}_*)} (p_\kappa(\mathbf{x},t) - p_0(\mathbf{x},t)\ \mathrm{d}\mathbf{x} \right| + \left| \int_{B_\varepsilon(\mathbf{x}_*)} p_0(\mathbf{x},t)\, \mathrm{d}\mathbf{x} - 1 \right|.$$

From Theorem 11, we have a constant $C > 0$ such that

$$\left| \int_{B_\varepsilon(\mathbf{x}_*)} (p_\kappa(\mathbf{x},t) - p_0(\mathbf{x},t))\ \mathrm{d}\mathbf{x} \right| < C\kappa t.$$

From Theorem 10, we have that there exists a time t^* such that, for $t > t^*$, we have

$$\left| \int_{B_\varepsilon(\mathbf{x}_*)} p_0(\mathbf{x},t)\,\mathrm{d}\mathbf{x} - 1 \right| < \delta .$$

Combining these two calculations yields the result.

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 $\mathbf{M} = (M_{ij})_{i,j=1,\cdots,n}$ 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

$$\Psi^{(i)}(\mathbf{x}) = \sum_{j=1}^{n} M_{ij} x_j = (\mathbf{M}\mathbf{x})_i \ . \tag{32}$$

In a simulation where both effects, drift and diffusion, are apparent, we have $\mu = \nu = 1$ and $\kappa = O(1)$. The last identity implies $\Delta t = O\left(N^{-1}\right)$. Furthermore, from equation (7), we have $\psi^{(i)}(\mathbf{x}) = \frac{1}{\Delta t}\left(\Psi^{(i)}(x) - 1\right)$, and therefore, in order to see both effects we need strong fitness functions and long times, i.e., $\Delta t \approx N^{-1} \ll 1$.

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

$$\mathbf{M} = \begin{pmatrix} 30 & 81 & 29 \\ 6 & 30 & 104 \\ 106 & 4 & 30 \end{pmatrix} \ . \tag{33}$$

This is know as the generalized Rock-Scissor-Paper game and presents an evolutionary stable state (ESS) $(x^*, y^*, z^*) = \left(\frac{1}{3}, \frac{1}{3}, \frac{1}{3}\right)$. Furthermore, the flow of the replicator dynamics converges in spirals to the ESS. The vertices 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 $\mathbf{M}$.

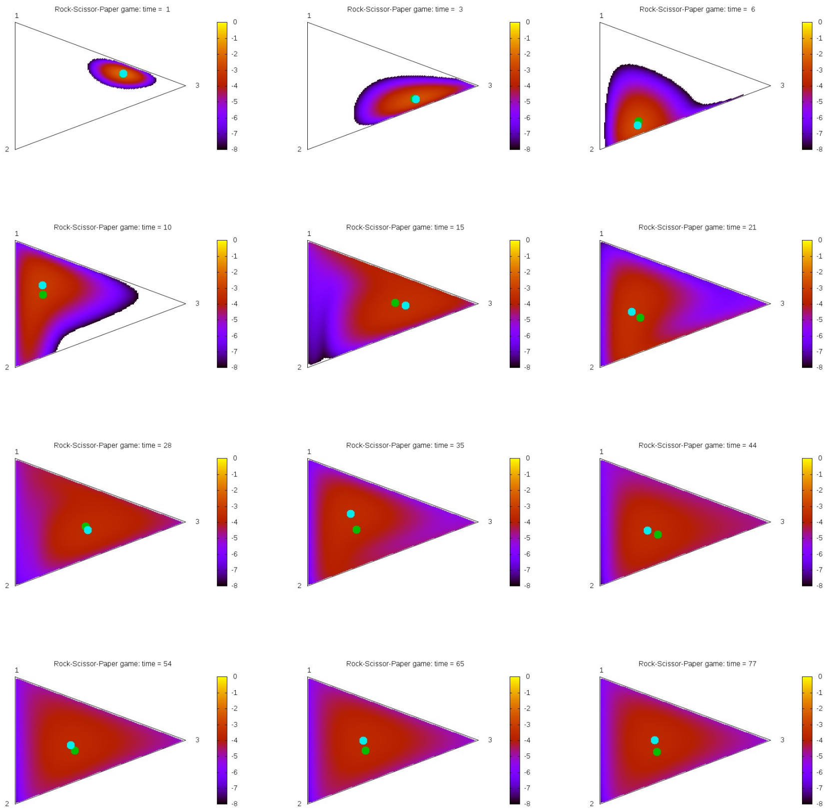

Fig. 2 Solution for short times (1,3,6,10,15,21,28,35,44,54,65,77) of the Wright-Fisher evolution for a population of 150 individuals of two given types, with fitness given by equations (32) and (33) for a distribution initially concentrated in the interior non-stationary point $\frac{1}{150}(70,70,10)$. 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 $\left(\frac{1}{3},\frac{1}{3},\frac{1}{3}\right)$ (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 $\mathbf{x}^{\infty} := \left(F^{(1)}_{p^{\mathrm{I}}}, F^{(2)}_{p^{\mathrm{I}}}, 1-F^{(1)}_{p^{\mathrm{I}}}-F^{(2)}_{p^{\mathrm{I}}}\right) \approx (0.331, 0.227, 0.442)$. For a full animation, also for different population sizes N, see `http://dl.dropbox.com/u/11325424/WFsim/RSPFinal.html`

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 $(x^{\infty}, y^{\infty}, z^{\infty}) = (\pi_1[p^{\mathrm{I}}], \pi_2[p^{\mathrm{I}}], \pi_3[p^{\mathrm{I}}])$. This point can be calculated using equation (25) and the $n = 3$ independent conservation laws.

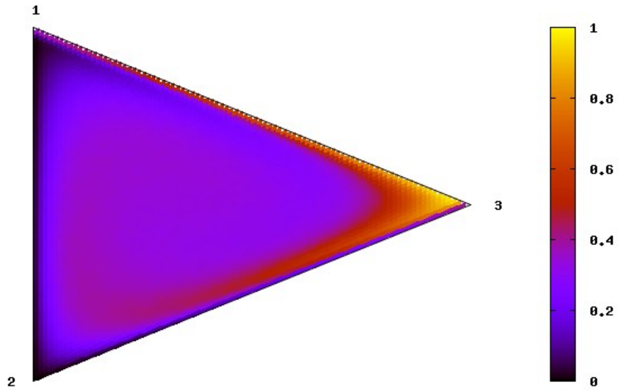

Fig. 3 Fixation probability of the third type, in a Rock-Scissor-Paper game. This is the numerical solution of the stationary state of the equation (26), simulated by a Wright-Fisher process with $N = 150$ and pay-off matrix $([[20,0,40],[40,20,0],[0,40,20]])$. Note that higher values of the fixation probability "rotates" around the center of the simplex (the stationary state of the replicator dynamics).

7.2 Backward equation and the decay of the interior L^1-norm

The stationary state of the backward equation (26) 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 evolution is given by the Rock-Scissor-Paper game defined by the matrix

$$\mathbf{M} = \begin{pmatrix} 0 & 40 & 20 \\ 20 & 0 & 40 \\ 40 & 20 & 0 \end{pmatrix} , \tag{34}$$

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, each solution representing boundary conditions of a larger set; however, a numerical solution is extremely easy to compute, as the Wright-Fisher process is a natural discretisation of the (forward as well as the) backward equation (cf. Theorem 5). 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 of a stochastic process reaching and

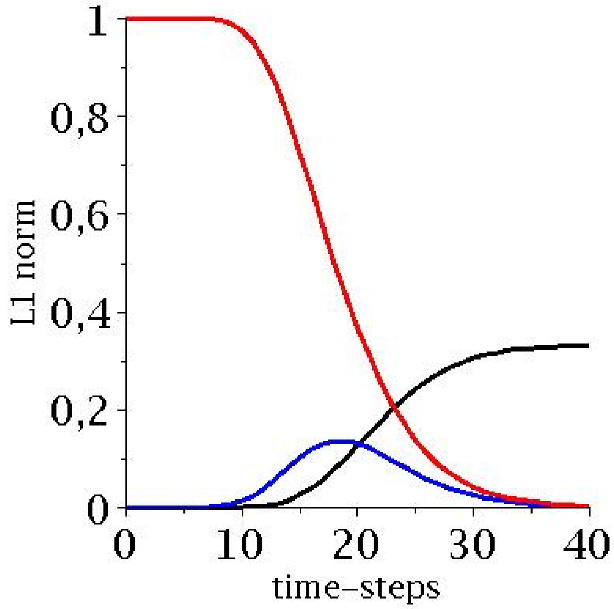

Fig. 4 Evolution of the probability mass, for the Rock-Scissor-Paper game given by matrix (34) and with initial condition concentrated in the ESS, $p^{\mathrm{I}} = \delta_{(\frac{1}{3},\frac{1}{3})}$. The red line indicates the mas (L^1-norm) in the interior of the simplex; the blue line, the mass in the interior any of the faces, and the black line, the mass in any of the vertices.

sticking to the faces of the simplex until they reach their final spot, the vertices. We further observe that, the probability mass in the interior of the simplex is the so-called quasi-stationary distribution of the process, namely, the probability distribution given that it has not been absorbed. See Méléard and Villemonais (2012) for a recent survey on the topic.

8 Conclusions

We present a derivation of continuous limits of discrete Markov chain evolutionary models, that are frequency-dependent extensions the classical Wright-Fisher model, through pure analytical techniques. The derivation presented pays close attention to the variety of possible time scalings possible as related to the selection and population size that are measured by two parameters $\mu,\nu \geq 1$. The balance of diffusion and selection ($\mu = \nu = 1$ in our terminology) can be seen as slight extension of the results in (Ethier and Kurtz 1986, Chapter 10) using analytical methods instead of probabilistic arguments, and that favours the forward Fokker-Planck equation instead of the backward. In this sense, from a mathematical point of view, the weak formulation presented for the forward equation seems to be new, in particular for the minimal assumptions on the fitness functions. The case $\mu > \nu = 1$ yields a hyperbolic equation that is the PDE version of the replicator equation. An apparently similar result can be found in (Ethier and Kurtz 1986, Chapter 11), which would correspond formally to take $\Delta t = 1$ and $N \gg 1$, without using explicit scaling between these two variables.

With some additional regularity assumptions on the fitnesses functions, we can show that (14) is equivalent to (1') together with the conservation laws (19). In particular, this allows to characterise the behaviour of p on the lower dimensional subsimplices of S^{n-1}. This can be used to obtain equations for the probability of extinction among other information.

The results here are also related to results in Champagnat et al. (2006, 2008), where the idea that the underling scaling influences the macroscopic model was already present, although in a less explicit way than here. Nevertheless, the accelerated birth-death regime in Champagnat et al. (2008) can be seen as a counterpart to our scaling of Δt and N. On the other hand, the scaling for the fitness are taken as fixed (corresponding to our $\nu = 1$ in our terminology), and this explains why they do not obtain the pure diffusive limit in the large population regime. Notice also that the large population regime taken there seems to annihilate any stochastic effects coming from births and deaths, and that the stochastic effects in this limit are due only to the mutation process.

However, as pointed out above, as we allow more flexibility in the scaling laws, we are able to highlight any of these two factors independently; more precisely, for certain choices of the scaling in the fitnesses functions (namely, the exponent $\nu > 1$), their influence in the dynamics goes to zero so fast that the limit model is purely diffusive. On the other hand, if we grow the population size fast enough (i.e., $\mu > 1$) then we highlight the deterministic evolution, providing a direct way to compare the replicator equation with the Wright-Fisher process (or, for that matter, also with the Moran process, but, naturally, in a different time scale). To the best of our knowledge, this explicit comparison is new. See also Fournier and Méléard (2004) for a similar approach.

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 2009b,a), that ODEs can be justifiably used to model the evolution of a population. However, the validity of the modelling 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 maximum likelihood method, but not necessarily compatible with other estimators).

One of the central issues of the present work is to discuss the possibility of using diffusive approximation for large, but finite, N. However, a major challenge to any one interested to use the replicator-diffusion equation to fit experimental data is the value of κ.

From the derivation of the replicator-diffusion equation, we see that κ is directly linked to the variance of the diffusion, while $\|\psi\|_\infty$ is directly linked with natural selection. Hence, their ratio is an adimensional measure of the relative relevance of the genetic drift with respect to natural selection. If we normalise the fitness functions such that $\|\psi\|_\infty = 1$, then κ becomes a measure of such relative relevance.

In this sense, κ^{-1} could be an alternative definition of effective population size (see also Etheridge (2011) for usual definitions). Only when the population is small or times are long in the evolutionary scale, we would expect order 1 values of κ.

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 (2011). 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 grants #s 309616/2009-3 and 451313/2011-9, and FAPERJ grant # 110.174/2009. We thank the careful reading and comments of three anonymous referees.

A Third moment of multinomial distributions

Let α be a multinomially distributed vector. Let also $\partial_i = \frac{\partial}{\partial q_i}$ and $|\mathbf{q}| = \sum_i q_i$ (not necessarily equal to 1). Evidently $\partial_i q_j = \delta_{ij}$. Then:

$$\begin{aligned}\mathbb{E}[\alpha_i\alpha_j\alpha_k] &= \left\{\sum_\alpha \alpha_i\alpha_j\alpha_k f(\mathbf{q},\alpha,N)\right\}_{|\mathbf{q}|=1} = \left\{q_i\partial_i\left[q_j\partial_j\left(q_k\partial_k|\mathbf{q}|^N\right)\right]\right\}_{|\mathbf{q}|=1}\\ &= N[q_i\delta_{ij}\delta_{kj}] + N(N-1)\left[q_iq_k\delta_{ij} + q_iq_j\delta_{kj} + q_kq_j\delta_{ki}\right] + N(N-1)(N-2)q_iq_jq_k\ .\end{aligned}$$

The expression for the third moment now follows from a straightforward calculation.

B 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 (9) 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

(9), and sum over k, we obtain that

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

On summing by parts the left hand side, we obtain

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

References

Ambrosio L, Gigli N, Savaré G (2005) Gradient flows in metric spaces and in the space of probability measures. Birkhäuser

Bardos C, Golse F, Levermore CD (1993) Fluid dynamic limits of kinetic equations. II. Convergence proofs for the Boltzmann equation. Comm. Pure Appl. Math. 46(5):667–753. doi: 10.1002/cpa.3160460503

Bardos C, Golse F, Levermore D (1991) Fluid dynamic limits of kinetic equations. I. Formal derivations. J. Statist. Phys. 63(1-2):323–344. doi: 10.1007/BF01026608

Benaim M, Weibull JW (2003) Deterministic approximation of stochastic evolution in games. Econometrica 71(3):873–903. doi: 10.1111/1468-0262.00429

Billingsley P (1999) Convergence of probability measures. John Wiley & Sons, New York

Bjorken JD, Drell SD (1964) Relativistic quantum mechanics. McGraw-Hill Book Co., New York

Boylan RT (1992) Laws of large numbers for dynamic-systems with randomly matched individuals. J. Econ. Theory 57(2):473–504. doi: 10.1016/0022-0531(92)90046-K

Boylan RT (1995) Continuous approximation of dynamical-systems with randomly matched individuals. J. Econ. Theory 66(2):615–625. doi: 10.1006/jeth.1995.1060

Carrol RW, Schowalter R (1976) Singular and degenerate Cauchy problems. Academic Press, New York

Cercignani C: 2002 The Boltzmann equation and fluid dynamics. Handbook of mathematical fluid dynamics, Vol. I, North-Holland, Amsterdam, pp 1–69

Chalub FACC, Markowich PA, Perthame B, Schmeiser C (2004) Kinetic models for chemotaxis and their drift-diffusion limits. Monatsh. Math. 142(1-2):123–141. doi: 10.1007/s00605-004-0234-7

Chalub FACC, Souza MO (2009a) From discrete to continuous evolution models: A unifying approach to drift-diffusion and replicator dynamics. Theor. Pop. Biol. 76(4):268–277. doi: 10.1016/j.tpb.2009.08.006

Chalub FACC, Souza MO (2009b) A non-standard evolution problem arising in population genetics. Commun. Math. Sci. 7(2):489–502

Chalub FACC, Souza MO (2011) The SIR epidemic model from a PDE point of view. Math. And Comp. Model. 53(7-8):1568–1574. doi: 10.1016/j.mcm.2010.05.036

Champagnat N, Ferrière R, Méléard S (2006) Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models. Theor. Popul. Biol. 69(3):297–321. doi: 10.1016/j.tpb.2005.10.004

Champagnat N, Ferrière R, Méléard S (2008) From individual stochastic processes to macroscopic models in adaptive evolution. Stoch. Models 24(suppl. 1):2–44. doi: 10.1080/15326340802437710
Cirincione RJ, Chernoff PR (1981) Dirac and Klein-Gordon equations: convergence of solutions in the nonrelativistic limit. Comm. Math. Phys. 79(1):33–46
Corradi V, Sarin R (2000) Continuous approximations of stochastic evolutionary game dynamics. J. Econ. Theory 94(2):163–191. doi: 10.1006/jeth.1999.2596
Courant R, Hilbert D (1989) Methods of mathematical physics. Vol. II. John Wiley & Sons Inc., New York
Crandall MG, Ishii H, Lions PL (1992) User's guide to viscosity solutions of second order partial differential equations. Bulletin of AMS 27(1):1–67. doi: 10.1090/S0273-0979-1992-00266-5
DiBenedetto E (1993) Degenerate parabolic equations. Springer-Verlag
DiPerna RJ, Lions PL (1989) Ordinary differential equations, transport theory and sobolev spaces. Inventiones Mathematicae 98:511–547. doi: 10.1007/BF01393835
Etheridge A (2011) Some mathematical models from population genetics. Springer, Heidelberg
Ethier SN, Kurtz TG (1986) Markov processes. John Wiley & Sons Inc., New York
Evans LC (2010) Partial differential equations. 2nd edn. American Mathematical Society, Providence, RI
Ewens WJ (2004) Mathematical population genetics. I. 2nd edn. Springer-Verlag, New York
Feller W (1951) Diffusion processes in genetics. Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability, 1950, University of California Press, Berkeley and Los Angeles, pp 227–246
Fisher RA (1922) On the dominance ratio. Proc. Royal Soc. Edinburgh 42:321–341. doi: 10.1007/BF02459576
Fisher RA (1930) The distribution of gene ratios for rare mutations. Proc. Royal Soc. Edinburgh 50:214–219
Foster D, Young P (1990) Stochastic evolutionary game dynamics. Theor. Popul. Biol. 38(2):219–232. doi: 10.1016/0040-5809(90)90011-J
Fournier N, Méléard S (2004) A microscopic probabilistic description of a locally regulated population and macroscopic approximations. Ann. Appl. Probab. 14(4):1880–1919. doi: 10.1214/105051604000000882
Fudenberg D, Harris C (1992) Evolutionary dynamics with aggregate shocks. J. Econ. Theory 57(2):420–441. doi: 10.1016/0022-0531(92)90044-I
Hartle DL, Clark AG (2007) Principles of population genetics. Sinauer, Massachussets
Hepp K (1974) The classical limit for quantum mechanical correlation functions. Comm. Math. Phys. 35:265–277
Hillen T, Othmer HG (2000) The diffusion limit of transport equations derived from velocity-jump processes. SIAM J. Appl. Math. 61(3):751–775 (electronic). doi: 10.1137/S0036139999358167
Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge, UK
Imhof LA, Nowak MA (2006) Evolutionary game dynamics in a Wright-Fisher process. J. Math. Biology 52(5):667–681. doi: 10.1007/s00285-005-0369-8
John F (1991) Partial differential equations. 4th edn. Springer-Verlag, New York
Karlin S, Taylor HM (1975) A first course in stochastic processes. 2nd edn. Academic Press [A subsidiary of Harcourt Brace Jovanovich, Publishers], New York-London
Keller L (ed.) (1999) Levels of selection in evolution. Princeton University Press, Princeton, NJ
Kimura M (1962) On the probability of fixation of mutant genes in a population. Genetics 47:713–719
Lessard S (2005) Long-term stability from fixation probabilities in finite populations: New perspectives for ESS theory. Theoret. Popul. Biology 68(1):19 – 27. doi: 10.1016/j.tpb.2005.04.001
Lessard S, Ladret V (2007) The probability of fixation of a single mutant in an exchangeable selection model. J. Math. Biol. 54:721–744. doi: 10.1007/s00285-007-0069-7
Lieberman G (1996) Second order parabolic differential equations. World Scientific
Maruyama T (1977) Stochastic problems in population genetics. Springer-Verlag, Berlin
McKane AJ, Waxman D (2007) Singular solutions of the diffusion equation of population genetics. J. Theoret. Biol. 247(4):849–858. doi: 10.1016/j.jtbi.2007.04.016
Méléard S, Villemonais D (2012) Quasi-stationary distributions and population processes. Probab. Surveys 9:340–410. doi: 10.1214/11-PS191
Metz JAJ (2011) Thoughts on the geometry of meso-evolution: collecting mathematical elements for a post-modern synthesis. In: Chalub FACC, Rodrigues JF (eds) The Mathematics of Darwin's Legacy, Springer, Basel, pp 193–231

Molzon R (2009) Deterministic approximation of stochastic evolutionary dynamics. 2009 international conference on game theory for networks (GAMENETS 2009), pp 323–332

Nowak MA (2006) Evolutionary dynamics — exploring the equations of life. The Belknap Press of Harvard University Press, Cambridge, MA

Othmer HG, Hillen T (2002) The diffusion limit of transport equations. II. Chemotaxis equations. SIAM J. Appl. Math. 62(4):1222–1250 (electronic). doi: 10.1137/S0036139900382772

Rogers L, Williams D (2000a) Diffusions, Markov processes, and martingales: Volume 1, Foundations. Cambridge University Press

Rogers L, Williams D (2000b) Diffusions, Markov processes and martingales: Volume 2, Itô calculus. Cambridge University Press

Sato KI (1976) Class of Markov-chains related to selection in population-genetics. J. Math. Soc. of Japan 28(4):621–637

Sato KI (1978) Convergence to a diffusion of a multi-allelic model in population genetics. Adv. in Appl. Probab. 10(3):538–562

Sato KI (1983) Limit diffusions of some stepping-stone models. J. Appl. Prob. 20(3):460–471

Smith JM (1982) Evolution and the theory of games. Cambridge University Press, Cambridge, UK

Smith JM (1988) Can a mixed strategy be stable in a finite population? J. of Theor. Biol. 130(2):247–251. doi: 10.1016/S0022-5193(88)80100-0

Stanley RP (1996) Combinatorics and commutative algebra. Birkhauser

Stevens A (2000) The derivation of chemotaxis equations as limit dynamics of moderately interacting stochastic many-particle systems. SIAM J. Appl. Math. 61(1):183–212 (electronic). doi: 10.1137/S0036139998342065

Stroock D, Varandhan S (1997) Multidimensional diffusion processes. Springer-Verlag

Taylor HM, Karlin S (1998) An introduction to stochastic modeling. 3rd edn. Academic Press Inc., San Diego, CA

Taylor ME (1996) Partial differential equations. I. Springer-Verlag, New York

Traulsen A, Claussen JC, Hauert C (2005) Coevolutionary dynamics: From finite to infinite populations. Phys. Rev. Lett. 95(23):238701. doi: 10.1103/PhysRevLett.95.238701

Traulsen A, Claussen JC, Hauert C (2012) Stochastic differential equations for evolutionary dynamics with demographic noise and mutations. Phys. Rev. E 85(4, Part 1). doi: 10.1103/PhysRevE.85.041901

van Kampen NG (1981) Stochastic processes in physics and chemistry. North-Holland Publishing Co., Amsterdam

Vickery WL (1988) Can a mixed strategy be stable in a finite population? — reply. J. of Theor. Biol. 132(3):375–378. doi: 10.1016/S0022-5193(88)80222-4

Waxman D (2011) 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. doi: 10.1016/j.jtbi.2010.10.014

Weibull JW (1995) Evolutionary game theory. The MIT Press, Cambridge, Massachussets

Wright S (1937) The distribution of gene frequencies in populations. Proc. Nat. Acad. Sci. US 23:307–320

Wright S (1938) The distribution of gene frequencies under irreversible mutations. Proc. Nat. Acad. Sci. US 24:253–259