

## Scaling limits of a model for selection at two scales

This content has been downloaded from IOPscience. Please scroll down to see the full text.

2017 Nonlinearity 30 1682

(<http://iopscience.iop.org/0951-7715/30/4/1682>)

View the [table of contents for this issue](#), or go to the [journal homepage](#) for more

Download details:

IP Address: 152.3.102.254

This content was downloaded on 01/06/2017 at 10:54

Please note that [terms and conditions apply](#).

You may also be interested in:

[Stochastic switching in biology: from genotype to phenotype](#)

Paul C Bressloff

[Ergodicity in randomly forced Rayleigh–Bénard convection](#)

J Földes, N E Glatt-Holtz, G Richards et al.

[Regularity of invariant densities for 1D systems with random switching](#)

Yuri Bakhtin, Tobias Hurth and Jonathan C Mattingly

[Cusping, transport and variance of solutions to generalized Fokker–Planck equations](#)

Sean Carnaffan and Reiichiro Kawai

[Anomalous diffusion for a class of systems with two conserved quantities](#)

Cédric Bernardin and Gabriel Stoltz

[Feynman–Kac equation for anomalous processes with space- and time-dependent forces](#)

Andrea Cairoli and Adrian Baule

[Mixing rates of particle systems with energy exchange](#)

A Grigo, K Khanin and D Szász

[Ergodicity of a collective random walk on a circle](#)

Michael Blank

[Stochastic partial differential equations with quadratic nonlinearities](#)

D Blömker, M Hairer and G A Pavliotis

# Scaling limits of a model for selection at two scales

Shishi Luo<sup>1,3</sup> and Jonathan C Mattingly<sup>2</sup>

<sup>1</sup> Computer Science Division and Department of Statistics, University of California - Berkeley, Berkeley, CA 94720, United States of America

<sup>2</sup> Department of Mathematics and Department of Statistical Science, Duke University, Durham, NC 27708, United States of America

E-mail: [shishi.luo@berkeley.edu](mailto:shishi.luo@berkeley.edu)

Received 3 July 2015, revised 24 August 2016

Accepted for publication 19 December 2016

Published 15 March 2017



Recommended by Professor Leonid Bunimovich

## Abstract

The dynamics of a population undergoing selection is a central topic in evolutionary biology. This question is particularly intriguing in the case where selective forces act in opposing directions at two population scales. For example, a fast-replicating virus strain outcompetes slower-replicating strains at the within-host scale. However, if the fast-replicating strain causes host morbidity and is less frequently transmitted, it can be outcompeted by slower-replicating strains at the between-host scale. Here we consider a stochastic ball-and-urn process which models this type of phenomenon. We prove the weak convergence of this process under two natural scalings. The first scaling leads to a deterministic nonlinear integro-partial differential equation on the interval  $[0,1]$  with dependence on a single parameter,  $\lambda$ . We show that the fixed points of this differential equation are Beta distributions and that their stability depends on  $\lambda$  and the behavior of the initial data around 1. The second scaling leads to a measure-valued Fleming–Viot process, an infinite dimensional stochastic process that is frequently associated with a population genetics.

Keywords: Markov chains, limiting behavior, evolutionary dynamics, Fleming–Viot process, scaling limits

Mathematics Subject Classification numbers: 35F55, 35Q92, 37L40, 60J28, 60J68, 60J70

(Some figures may appear in colour only in the online journal)

<sup>3</sup> Author to whom any correspondence should be addressed.

## 1. Introduction

We study the model, introduced in [15], of a trait that is advantageous at a local or individual level but disadvantageous at a larger scale or group level. For example, an infectious virus strain that replicates rapidly within its host will outcompete other virus strains in the host. However, if infection with a heavy viral load is incapacitating and prevents the host from transmitting the virus, the rapidly replicating strain may not be as prevalent in the overall host population as a slow replicating strain.

A simple mathematical formulation of this phenomenon is as follows. Consider a population of  $m \in \mathbb{N}$  groups. Each group contains  $n \in \mathbb{N}$  individuals. There are two types of individuals: type I individuals are selectively advantageous at the individual (I) level and type G individuals are selectively advantageous at the group (G) level. Replication and selection occur concurrently at the individual and group level according to the Moran process [8] and are illustrated in figure 1. Type I individuals replicate at rate  $1 + s$ ,  $s \geq 0$  and type G individuals at rate 1. When an individual gives birth, another individual in the same group is selected uniformly at random to die. To reflect the antagonism at the higher level of selection, groups replicate at a rate which increases with the number of type G individuals they contain. As a simple case, we take this rate to be  $w(1 + r\frac{k}{n})$ , where  $\frac{k}{n}$  is the fraction of individuals in the group that are type G,  $r \geq 0$  is the selection coefficient at the group level, and  $w > 0$  is the ratio of the rate of group-level events to the rate of individual-level events. As with the individual level, the population of groups is maintained at  $m$  by selecting a group uniformly at random to die whenever a group replicates. The offspring of groups are assumed to be identical to their parent.

As illustrated in figure 1, this two-level process is equivalent to a ball-and-urn or particle process, where each particle represents a group and its position corresponds to the number of type G individuals that are in it. We note that similar, though more general, particle models of evolutionary and ecological dynamics at multiple scales have been studied and we mention several particularly relevant works here. Dawson and Hochberg [6] also consider a population at multiple levels, albeit with one type per level, not two. In [4], Dawson and Greven consider a more general model, allowing for infinitely many hierarchical levels and for migration, selection, and mutation. Méléard and Roelly [16], in a study also inspired by host-pathogen interactions, investigate a model that allows for non-constant host and pathogen populations as well as mutation. In these more general contexts, determining long-term behavior is less straightforward than in our specific setting.

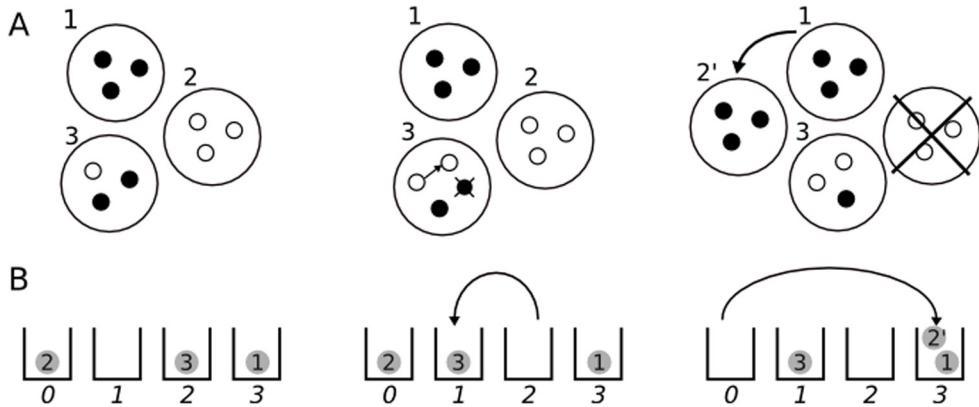
We now define the stochastic process that is the focus of this work. Let  $X_t^i$  be the number of type G individuals in group  $i$  at time  $t$ . Then

$$\mu_t^{m,n} := \frac{1}{m} \sum_{i=1}^m \delta_{X_t^i/n}$$

is the empirical measure at time  $t$  for a given number of groups  $m$  and individuals per group  $n$ .  $\delta_x(y) = 1$  if  $x = y$  and zero otherwise. The  $X_t^i$  are divided by  $n$  so that  $\mu_t^{m,n}$  is a probability measure on  $E_n := \{0, \frac{1}{n}, \dots, 1\}$ .

For fixed  $T > 0$ ,  $\mu_t^{m,n} \in D([0, T], \mathcal{P}(E_n))$ , the set of càdlàg processes on  $[0, T]$  taking values in  $\mathcal{P}(E_n)$ , where  $\mathcal{P}(S)$  is the set of probability measures on a set  $S$ . With the particle process described above,  $\mu_t^{m,n}$  has generator

$$(L^{m,n}\psi)(v) = \sum_{i,j} (R_1 + wR_2)(v, v_{ij})[\psi(v_{ij}) - \psi(v)] \quad (1)$$



**Figure 1.** Schematic of the particle process. (a) *Left*: a population of  $m = 3$  groups, each with  $n = 3$  individuals of either type G (filled small circles) or type I (open small circles). *Middle*: a type I individual replicates in group 3 and a type G individual is chosen uniformly at random from group 3 to die. *Right*: group 1 replicates and produces group 2'. Group 2 is chosen uniformly at random to die. (b) The states in (a) mapped to a particle process. *Left*: group 2 has no type G individuals, represented by ball 2 in urn 0. Similarly, group 3 is represented by ball 3 in urn 2 and group 1 by ball 1 in urn 3. *Middle*: the number of type G individuals in group 3 decreases from two to one, therefore ball 3 moves to urn 1. *Right*: a group with zero type G individuals dies, while a group with three type G individuals is born. Therefore ball 2 leaves urn 0 and appears in urn 3 as ball 2'.

where  $v_{ij} := v + \frac{1}{m}(\delta_j - \delta_i)$ ,  $\psi \in C_b(\mathcal{P}([0, 1]))$  are bounded continuous functions, and  $v \in \mathcal{P}(E_n) \subset \mathcal{P}([0, 1])$ . The transition rates ( $R_1 + wR_2$ ) are given by

$$R_1(v, v_{ij}) = \begin{cases} mv\left(\frac{i}{n}\right)i\left(1 - \frac{i}{n}\right)(1 + s) & \text{if } j = i - 1, i < n \\ mv\left(\frac{i}{n}\right)i\left(1 - \frac{i}{n}\right) & \text{if } j = i + 1, i > 0 \\ 0 & \text{otherwise} \end{cases}$$

and

$$R_2(v, v_{ij}) = mv\left(\frac{i}{n}\right)v\left(\frac{j}{n}\right)(1 + r\frac{j}{n}).$$

$R_1$  represents individual-level events while  $R_2$  represents group-level events.

## 2. Main results

We prove the weak convergence of this measure-valued process as  $m, n \rightarrow \infty$  under two natural scalings. The first scaling leads to a deterministic partial differential equation. We derive a closed-form expression for the solution of this equation and study its steady-state behavior. The second scaling leads to an infinite dimensional stochastic process, namely a Fleming–Viot process.

Let us briefly introduce some notation. By  $m, n \rightarrow \infty$  we mean a sequence  $\{(m_k, n_k)\}_k$  such that for any  $N$ , there is an  $n_0$  such that if  $k \geq n_0$ ,  $m_k, n_k \geq N$ . We define  $\langle f, v \rangle = \int_0^1 f(x)v(dx)$

where  $f$  is a test function and  $v$  a measure. Lastly,  $\delta_x$  will denote the delta measure for both continuous and discrete state spaces.

To provide intuition for the two scalings and the corresponding limits, take  $\psi$  to be of the form  $\psi(v) = F(\langle g, v \rangle)$ , where  $g$  is some suitable function on  $[0,1]$ , and apply the generator in (1) to it:

$$\begin{aligned} L^{m,n}\psi(\mu) &= F' \cdot \left\{ \sum_i \left[ \frac{1}{n} g''\left(\frac{i}{n}\right) - sg'\left(\frac{i}{n}\right) \right] \frac{i}{n} \left(1 - \frac{i}{n}\right) \mu\left(\frac{i}{n}\right) \right. \\ &\quad \left. + wr \left[ \sum_i \frac{i}{n} g\left(\frac{i}{n}\right) \mu\left(\frac{i}{n}\right) - \sum_i g\left(\frac{i}{n}\right) \mu\left(\frac{i}{n}\right) \sum_j \frac{j}{n} \mu\left(\frac{j}{n}\right) \right] \right\} \\ &\quad + \frac{1}{m} wF'' \cdot \left\{ \sum_i g\left(\frac{i}{n}\right)^2 \mu\left(\frac{i}{n}\right) - \left( \sum_i g\left(\frac{i}{n}\right) \mu\left(\frac{i}{n}\right) \right)^2 \right. \\ &\quad \left. + \frac{1}{2} r \sum_{i,j} \left( g\left(\frac{i}{n}\right) - g\left(\frac{j}{n}\right) \right)^2 \frac{j}{n} \mu\left(\frac{i}{n}\right) \mu\left(\frac{j}{n}\right) \right\} + o\left(\frac{1}{m}\right) + o\left(\frac{1}{n}\right) \end{aligned} \quad (2)$$

This suggests two natural scalings. The first is to take  $m, n \rightarrow \infty$  without rescaling any parameters. The  $g''$  and  $F''$  terms vanish and we have a deterministic process. The second is to let  $s = \frac{\sigma}{n}$ ,  $r = \frac{\rho}{m}$ , and  $\frac{n}{m} \rightarrow \theta$ . The terms  $F''$  and  $g''$  no longer vanish and the process converges to a limit that is stochastic. The precise statement of the weak convergence of the finite state space system to the deterministic limit is in terms of a weak measure-valued solution to a partial differential equation:

**Theorem 1.** *Suppose the particles in the system described by  $\mu_t^{m,n}$  are initially independently and identically distributed according to the measure  $\mu_0^{m,n}$ , where  $\mu_0^{m,n} \rightarrow \mu_0 \in \mathcal{P}([0, 1])$  as  $m, n \rightarrow \infty$ . Then, as  $m, n \rightarrow \infty$ ,  $\mu_t^{m,n} \rightarrow \mu_t \in D([0, T], \mathcal{P}([0, 1]))$  weakly, where  $\mu_t$  solves the differential equation*

$$\frac{d}{dt} \langle f, \mu_t \rangle = -\langle x(1-x)f', \mu_t \rangle + \lambda [\langle xf, \mu_t \rangle - \langle f, \mu_t \rangle \langle x, \mu_t \rangle] \quad (3)$$

for any positive-valued test function  $f \in C^1([0, 1])$  and with initial condition  $\langle f, \mu_0 \rangle$ . Here,  $\lambda := \frac{wr}{s}$  and time has been sped up by a factor of  $s$ .

Throughout we will denote the measure-valued solutions to (3) by  $\mu_t(dx)$ . We note that strong, density-valued solutions, denoted by  $\eta_t(x)$ , solve:

$$\frac{\partial}{\partial t} \eta_t = \frac{\partial}{\partial x} [x(1-x)\eta_t] + \lambda \eta_t \cdot \left( x - \int_0^1 y \eta_t(y) dy \right) \quad (4)$$

with initial density  $\eta_0(x)$ . In this more transparent form one can see that the first term on the right is a flux term that transports density towards  $x = 0$  whereas the second term is a forcing term that increases the density at values of  $x$  above the mean of the density. The flux corresponds to the individual-level moves: nearest neighbor moves in the particle system. The forcing term corresponds to group-level moves: moves to occupied sites in the particle system.

We will see that if we start with an initial measure  $\mu_0$  which is the sum of delta measures, then the solution  $\mu_t$  retains the same form. More explicitly, if

$$\mu_0(dx) = \sum_i a_i(0) \delta_{x_i(0)}(dx)$$

where  $x_i(0) \in [0, 1]$ ,  $a_i(0) > 0$ , and  $\sum a_i(0) = 1$ , then we will see (from lemma 5) that the solution  $\mu_t$  to (3) has the form

$$\mu_t(dx) = \sum_i a_i(t) \delta_{x_i(t)}(dx).$$

Moreover, the parameters  $(a_i(t), x_i(t))$  satisfy the following set of coupled equations

$$\begin{cases} \frac{dx_i}{dt} = -x_i(1 - x_i) \\ \frac{da_i}{dt} = \lambda a_i(x_i - \langle y, \mu_t \rangle) = \lambda a_i \left( x_i - \sum_j a_j x_j \right). \end{cases} \quad (5)$$

Notice that the positions of the delta masses change according to a negative logistic function, independently of the other masses and the density. The weight  $a_i$  increases at time  $t$  if the position of the particle  $x_i$  is above the mean,  $\sum a_j x_j$ , and decreases if it is below the mean. To build intuition, it is instructive to consider some simple examples of this form.

**Example 1.** According to (5), if  $\mu_0 = \delta_1$ , then  $\mu_t = \mu_0$ . This can also be seen directly from (3). In the case of an initial condition containing some delta mass at 1, all of the rest of the mass will migrate towards zero. Eventually all of the mass will be below the mean as the mass at one will not move and will ever be increasing its mass as it is always above the mean. Once this happens it is clear that all of the mass will drain from all of the points not at one and hence  $\mu_t \rightarrow \delta_1$  as  $t \rightarrow \infty$ . This reasoning holds in a more general setting and is included in theorem 3.

**Example 2.** According to (5), if  $\mu_0 = \delta_0$ , then  $\mu_t = \mu_0$ . This too can be seen directly from (3). In the case of an initial condition containing no mass at one and only finite number of masses total, the mass will eventually all move towards zero and hence hence  $\mu_t \rightarrow \delta_0$  as  $t \rightarrow \infty$ . If an infinite number of masses are allowed the situation is not as simple. Theorem 3 hints at the possible complications by giving an example of a density which is invariant.

These simple examples correspond to similarly straightforward biological scenarios: once the population is entirely composed of type I individuals or of type G individuals, it stays in that state. Though  $\delta_0$  is a fixed point of the system attracting many initial configurations, it is not Lyapunov stable. This means that even small perturbations of  $\delta_0$  can lead to an arbitrary large excursion away from  $\delta_0$  even though the system eventually returns to  $\delta_0$ . Rather than making a precise statement which would require quantifying the size of a perturbation, consider the example of  $\mu_0 = (1 - \varepsilon)\delta_0 + \varepsilon\delta_{1-\alpha}$ . As  $\varepsilon \rightarrow 0$ , the distance between  $\mu_0$  and  $\delta_0$  goes to zero in any reasonable metric. If we write  $\mu_t = (1 - a_t)\delta_0 + a_t\delta_{x_t}$  then as  $\alpha \rightarrow 0$  one can ensure that the system spends arbitrarily long time with  $x_t > \frac{1}{2}$  and hence  $a_t$  will grow to as close to one as one wants in this time. Thus the system could be described as making an an arbitrarily big excursion away from  $\delta_0$  even though  $\mu_t \rightarrow \delta_0$  as  $t \rightarrow \infty$ .

It natural to ask if there are other fixed points beyond  $\delta_0$  and  $\delta_1$ .

**Lemma 2 (Fixed points).** *The measures delta  $\delta_0$ ,  $\delta_1$ , and densities in the Beta( $\lambda - \alpha, \alpha$ ) family of distributions:*

$$\frac{1}{B(\lambda - \alpha, \alpha)} x^{\lambda - \alpha - 1} (1 - x)^{\alpha - 1}$$

with  $\alpha \in (0, \lambda)$ , are fixed points of (3).  $B(\lambda - \alpha, \alpha)$  is the normalizing constant that makes the density integrate to 1 over the interval  $[0, 1]$ .

For measure-valued initial data, we show that the basins of attraction for the fixed points are determined by whether they charge the point  $x = 1$  and their Hölder exponent around  $x = 1$ .

**Theorem 3 (Steady state behavior).** Consider measure valued solution  $\mu_t(dx)$  to (3) with initial probability measure  $\mu_0(dx)$ . If  $\mu_0(\{1\}) > 0$  then

$$\mu_t \rightarrow \delta_1 \quad \text{as} \quad t \rightarrow \infty$$

and if  $\mu_0([1 - \varepsilon, 1]) = 0$  for some  $\varepsilon > 0$  then

$$\mu_t \rightarrow \delta_0 \quad \text{as} \quad t \rightarrow \infty.$$

Alternatively, suppose that for some  $\alpha > 0$  and  $C > 0$

$$x^{-\alpha} \mu_0([1 - x, 1]) \rightarrow C \quad \text{as} \quad x \rightarrow 0.$$

If  $\alpha < \lambda$ , then

$$\mu_t(dx) \rightarrow \text{Beta}(\lambda - \alpha, \alpha) \quad \text{as} \quad t \rightarrow \infty.$$

Otherwise, if  $\alpha \geq \lambda$ ,

$$\mu_t(dx) \rightarrow \delta_0(dx) \quad \text{as} \quad t \rightarrow \infty$$

The  $\alpha < \lambda$  case in the above theorem is particularly relevant to theoretical evolutionary biology because it implies that coexistence of the two types is possible in this infinite population limit.

The results of theorem 3 should be contrasted with the original Markov chain before taking the limit  $m, n \rightarrow \infty$ . In the Markov chain, all individuals eventually become either entirely type G or type I. These two homogeneous states are absorbing states for the individual level dynamics. The population level state made of individuals that are all either homogeneous of type G or I is absorbing for the group level dynamics. Hence, the state of the system eventually becomes composed entirely of homogeneous groups of solely G or I and stays in that state for all future times. These two absorbing states of the Markov chain, with finite  $m$  and  $n$ , correspond to the states  $\delta_0$  and  $\delta_1$  in the scaling limit. Hence the natural discretization for the Beta distribution to the lattice  $\{\frac{k}{n} : 0 < k < n\}$ , given by

$$\frac{1}{Z(m, n, \lambda, \alpha)} \left( \frac{k}{n} \right)^{\lambda - \alpha - 1} \left( 1 - \frac{k}{n} \right)^{\alpha - 1},$$

cannot be invariant. (Here  $Z$  is the normalization constant which ensures the probabilities sum to one.) However for large  $m$  and  $n$ , it is reasonable to expect it to be nearly invariant in the sense that if the initial states  $\{X_i(0) : 1 \leq i \leq m\}$  are independent and distributed as the discrete Beta distribution then the Markov chain dynamics will keep the distribution close to the product of discretized Beta distributions for a long time. The expectation of this time will grow to infinity as  $m, n \rightarrow \infty$ . In the context of evolutionary biology, this suggests that although a large finite population ultimately becomes fixed in one of two homogeneous states, it may be trapped for a long time in a state where both types coexist. Furthermore, this nearly invariant state should be similar to the discretization of the Beta distribution above.

We will not pursue a rigorous proof of this near or quasi invariance here. Nonetheless, we now briefly sketch the argument as we understand it, giving the central points. If the distribution of the Markov chain is close to a product of discretized Beta distributions, then the empirical mean will be highly concentrated around the mean of continuous Beta when  $m$  and  $n$  are large. Hence the generator projected on to any  $X_i$  is nearly decoupled from the other particles and close to being Markovian. More precisely, the dynamics of any fixed  $X_i$  is well approximated in this setting by the one-dimensional Markov chain obtained by replacing the mean of the empirical measure in the full generator with the mean of the Beta distribution. It is straightforward to see that for  $m$  and  $n$  large the discretized Beta distribution is an approximate left-eigenfunction of this one-dimensional generator with an eigenvalue which goes to zero as  $m, n \rightarrow \infty$ .

All of these observations can be combined to show that if the systems starts in the product discretized Beta distribution then it will stay close to the product discretized Beta distribution for a long time if  $m$  and  $n$  are large.

We now turn to the second scaling. Let  $s = \frac{\sigma}{n}$ ,  $r = \frac{\rho}{m}$ , and  $\frac{n}{m} \rightarrow \theta$ , and let  $\nu_t^{m,n}$  denote the empirical measure under this scaling. The terms  $F''$  and  $g''$  in the generator (1) no longer vanish and the process converges to a limit that is stochastic. Our weak convergence result is proved and stated in terms of a martingale problem.

**Theorem 4.** Suppose  $\frac{n}{m} \rightarrow \theta$ ,  $w = O(1)$ ,  $s = \frac{\sigma}{n}$ ,  $r = \frac{\rho}{m}$ , and we speed up time by a factor of  $n$ . Suppose the particles in the rescaled  $\nu_t^{m,n}$  process are initially independently and identically distributed according to the measure  $\nu_0^{m,n}$  where  $\nu_0^{m,n} \rightarrow \nu_0$  as  $m, n \rightarrow \infty$ . Then the rescaled process converges weakly to  $\nu_t$  as  $m, n \rightarrow \infty$ , where  $\nu_t$  satisfies the following martingale problem:

$$\begin{aligned} N_t(f) &= \langle f, \nu_t \rangle - \langle f, \nu_0 \rangle - \int_0^t \langle Af, \nu_z \rangle dz \\ &\quad - w\theta\rho \int_0^t \left\{ \int_0^1 \int_0^1 f(x) V(z, \nu_z, y) Q(\nu_z; dx, dy) \right\} dz \end{aligned} \quad (6)$$

is a martingale with conditional quadratic variation

$$\langle N(f) \rangle_t = 2w\theta \int_0^t \left[ \int_0^1 \int_0^1 f(x) f(y) Q(\nu_\tau; dx, dy) \right] d\tau \quad (7)$$

where

$$\begin{aligned} Af(x) &= x(1-x) \left[ \frac{d^2}{dx^2} f(x) - \sigma \frac{d}{dx} f(x) \right] \\ V(t, \nu, x) &= x \\ Q(\nu; dx, dy) &= \nu(dx)(\delta_x(dy) - \nu(dy)) \end{aligned}$$

and  $f \in C^2([0, 1])$ .

The drift part of the martingale (6) comprises a second order partial differential operator  $A$  and the centering term from the global jump dynamics (the expression in curly brackets). Note that  $A$  is in fact the generator of the Wright–Fisher diffusion with selection [8]. The entire process is a Fleming–Viot process [11]. Fleming–Viot processes frequently arise in models of population genetics (for example [3, 12]; see [10] for a review). In these contexts,

the variable  $x$  can represent the geographical location of an individual, or as in the original paper of Fleming and Viot [11], the genotype of an individual (where genotype is a continuous instead of a discrete variable).

As an aside, it may be helpful to mention an alternative characterization of this Fleming–Viot process as an infinite system of ordinary stochastic differential equations. Instead of a martingale problem, where both  $m, n$  have been taken to  $\infty$ , we consider the  $n \rightarrow \infty$  limit first. In this case, we have a finite collection of delta masses (each of mass  $\frac{1}{m}$ ) moving on the interval  $[0, 1]$ . The positions of these delta masses can be represented by a coupled system of stochastic differential equations (SDEs). From the generator equation (2), one can see that each SDE comprises a diffusion part (corresponding to the individual-level dynamics) and a jump process (corresponding to the group-level dynamics). Specifically, a delta mass jumps to the position of another delta mass according to a Poisson process with rates dependent on the positions of the delta masses. Donnelly and Kurtz [7] characterize a population process in terms of such a system of SDES and show that the infinite population limit corresponds to a martingale problem for the Fleming–Viot process.

We briefly discuss other scalings one might obtain from the particle system and their biological significance. The two scalings studied here correspond, respectively, to what is called ‘strong’ selection and ‘weak’ selection occurring at both levels. (In the field of theoretical evolutionary biology, strong selection is defined as the selection parameter being constant in the population size whereas weak selection has the selection parameter scaling with the inverse of the population size). One can also use the same techniques to characterize the limiting system when selection is strong at one level and weak at the other. The dynamics of these limiting systems are more straightforward. For example, if selection is weak at the individual level ( $s = O(\frac{1}{n})$ ) and strong at the group level (no rescaling of  $r$ ), one can see from the generator equation (2) that the highest order term corresponds to selection at the group level. The limit is therefore deterministic and the steady-state is a population homogeneous in the group type with the largest proportion of type G individuals present in the initial state. Note that it is possible, by further rescaling  $w$ , to obtain a limit from a mixture of weak and strong selection. A biological interpretation of these observations is that for selection to manifest itself at two biological levels, the selective forces must be comparable in some sense: either both levels undergo the same type of selection (weak or strong), or if weak selection acts on one level but not the other, this weak selection must be compensated by a faster timescale.

The dynamical properties of the deterministic partial differential equation (3) are the focus of the next section. The proofs of weak convergence (theorems 1 and 4) are deferred to section 4.

### 3. Properties of the deterministic limit

We begin with a closed-form expression for solutions to the deterministic partial differential equation (3).

**Lemma 5.** *The solution to the deterministic partial differential equation (3) with initial measure  $\mu_0$  is given by*

$$\mu_t(dx) = (G_t \mu_0)(dx) = (\mu_0 \phi_t^{-1})(dx) \cdot w_t(x) \quad (8)$$

where

$$\phi_t^{-1}(x) = \frac{x}{e^{-t} + x(1 - e^{-t})}$$

$$w_t(x) = \left[ (e^{-t} + x(1 - e^{-t})) e^{t - \int_0^t h(z) dz} \right]^\lambda$$

and  $h(t)$  satisfies  $h(t) = \langle x, \mu_t \rangle$

**Remark 1.**  $(\mu_0 \phi_t^{-1})(dx) := \mu_0(\phi_t^{-1}(dx))$  captures the changes in the initial data that are solely due to the flux term. This expression is also known as the push-forward measure of  $\mu_0$  under the dynamics of  $\phi$ . As we will see in the proof,  $\phi_t(x)$  is precisely the characteristic curve for the spatial variable  $x$  and includes a normalizing constant. The multiplication by  $w_t(x)$  captures the changes in the initial data that are due to the forcing term in (3) and includes a normalizing factor.

**Remark 2.** Density-valued solutions are given by

$$\begin{aligned} \eta_t(x) &= \eta_0(\phi_t^{-1}(x)) \partial_x \phi_t^{-1}(x) \cdot w_t(x) \\ &= \eta_0 \left( \frac{x}{e^{-t} + x(1 - e^{-t})} \right) [e^{-t} + x(1 - e^{-t})]^{\lambda-2} e^{(\lambda-1)t - \lambda \int_0^t h(z) dz} \end{aligned} \quad (9)$$

To see this, suppose  $\mu_0(dx) = \eta_0(x)dx$ . Then for any test function  $f$ ,

$$\int_0^1 f(x) (\mu_0 \phi_t^{-1})(dx) = \int_0^1 (f \circ \phi_t)(x) \mu_0(dx) = \int_0^1 f(y) \eta_0(\phi_t^{-1}(y)) \partial_y \phi_t^{-1}(y) dy$$

The first equality follows from the change-of-variable property of push-forward measures and the second from a standard change of variables. The limits of integration do not change because 0 and 1 are fixed points of both  $\phi_t$  and  $\phi_t^{-1}$ .

**Proof of lemma 5.** We apply the method of characteristics (see for example [17]) to obtain a formula for a density-valued solution. We then prove that the weak, measure-valued analog of this solution satisfies (3). Consider the following modification of (4):

$$\frac{\partial}{\partial t} \xi(t, x) = \frac{\partial}{\partial x} [x(1 - x)\xi(t, x)] + \lambda \xi(t, x) [x - h(t)] \quad (10)$$

where  $h(t)$  is a general function in time and  $\xi_0 \in C^1([0, 1])$ . Note that when  $h(t) = \int_0^1 y \xi(t, y) dy$ , this differential equation is equivalent to (4). To be clear about which equation we are solving, we use  $\xi(t, x)$  to denote solutions when  $h(t)$  is unspecified.

Rewriting (10):

$$\left( \frac{\partial}{\partial t} \xi, \frac{\partial}{\partial x} \xi, -1 \right) \cdot (1, -x(1 - x), [(1 - 2x) + \lambda(x - h(t))] \xi) = 0$$

The second vector is therefore tangent to the solution surface and gives the rates of change for the  $t$ ,  $x$ , and  $\xi$  coordinates. Let the initial condition be parameterized as  $(0, x, \xi_0(x)) = (0, p, \xi_0(p))$ . The  $t$ ,  $x$ , and  $\xi$  coordinates change according to the characteristic equations

$$\begin{aligned}
\frac{dt}{dq} &= 1 & t(0, p) &= 0 \\
\frac{dx}{dq} &= -x(1 - x) & x(0, p) &= p \\
\frac{d\xi}{dq} &= [(1 - 2x(q, p)) + \lambda(x(q, p) - h(t(q, p)))] \xi & \xi(0, p) &= \xi_0(p)
\end{aligned}$$

where  $q$  is the parameter as we move through the solutions in time. The first two ordinary differential equations have solutions

$$\begin{aligned}
t(q, p) &= q \\
x(q, p) &= \frac{p}{p - (p - 1)e^q} =: \phi_q(p)
\end{aligned} \tag{11}$$

From this, the third differential equation can be solved exactly:

$$\begin{aligned}
\frac{d\xi}{dq} &= \left[ 1 + \frac{p}{p - (p - 1)e^q} (\lambda - 2) - \lambda h(q) \right] \xi \\
\xi(q, p) &= \xi_0(p) \exp \left\{ q - \lambda \int_0^q h(z) dz + (\lambda - 2) \int_0^q \frac{p}{p - (p - 1)e^z} dz \right\} \\
&= \xi_0(p) e^{q - \lambda \int_0^q h(z) dz} [p(e^{-q} - 1) + 1]^{(-\lambda+2)}
\end{aligned}$$

Next, make the substitutions  $q = t$  and  $p = \phi_t^{-1}(x)$  from (11) to obtain  $\xi$  in terms of  $t$  and  $x$ :

$$\begin{aligned}
\xi(t, x) &= (\xi_0 \circ \phi_t^{-1})(x) [e^{-t} + x(1 - e^{-t})]^{(\lambda-2)} e^{(\lambda-1)t - \lambda \int_0^t h(z) dz} \\
&= (\xi_0 \circ \phi_t^{-1})(x) \partial_x \phi_t^{-1}(x) \cdot w_t(x)
\end{aligned} \tag{12}$$

If  $h(t)$  satisfies  $h(t) = \int_0^1 y \xi(t, y) dy$ , then by definition,  $\xi(t, x)$  solves the partial differential equation (4). Conversely, if  $\xi(t, x)$  solves the partial differential equation (4), it also solves the differential equation (10) with  $h(t) = \int_0^1 y \xi(t, y) dy$ . Therefore this above expression, along with the condition  $h(t) = \int_0^1 y \xi(t, y) dy$ , are equivalent to solutions of (4).

To extend this result to measures, suppose we have a strong solution  $\eta_t(x)$  with initial condition  $\eta_0$ :

$$\eta_t(x) = (\eta_0 \circ \phi_t^{-1})(x) \partial_x \phi_t^{-1}(x) \cdot w_t(x)$$

Using a similar calculation as that in remark 2, the measure  $\mu_t(dx)$  corresponding to  $\eta_t(x)$  is given by

$$\mu_t(dx) = (\mu_0 \phi_t^{-1})(dx) \cdot w_t(x)$$

It remains to check that this satisfies the weak deterministic partial differential equation (3), with  $h(t) = \langle x, \mu_t \rangle$ . The left hand side of the equation is

$$\frac{d}{dt} \langle f, \mu_t \rangle = \frac{d}{dt} \int_0^1 f(x) w_t(x) (\mu_0 \phi_t^{-1})(dx) = \frac{d}{dt} \int_0^1 f(\phi_t(x)) w_t(\phi_t(x)) \mu_0(dx)$$

Differentiating under the integral sign, expanding out the expressions for  $\partial_t \phi_t$  and  $\partial_t(w_t(\phi_t(x)))$ , and applying change of variables for push-forward measures again, we obtain

$$\frac{d}{dt} \langle f, \mu_t \rangle = - \int_0^1 x(1-x)f'(x) w_t(x) (\mu_0 \phi_t^{-1})(dx) + \lambda \int_0^1 [x - h(t)] f(x) w_t(x) (\mu_0 \phi_t^{-1})(dx)$$

This matches right hand side of the weak deterministic partial differential equation (3).  $\square$

In practice, the condition  $h(t) = \langle x, \mu_t \rangle$  is difficult to use. The following provides an equivalent and simpler condition.

**Lemma 6 (Conservation of measure condition).** *Suppose  $\xi$  is a weak measure-valued solution to the deterministic partial differential equation (10) with initial condition  $\int_0^1 \xi_0(dx) = 1$ . Then*

$$h(t) = \int_0^1 y \xi(t, dy) \quad \text{if and only if} \quad \int_0^1 \xi(t, dy) = 1 \quad \forall t > 0$$

**Proof.** ( $\Rightarrow$  direction) Suppose  $h(t) = \int_0^1 y \xi(t, dy)$ . Then  $\xi$  is a weak measure-valued solution to (3). Taking the test function  $f \equiv 1$ , we obtain

$$\frac{d}{dt} \langle 1, \xi \rangle = 0 + \lambda [\langle x, \xi \rangle - \langle 1, \xi \rangle \langle x, \xi \rangle] = 0$$

Thus, if the initial data has total measure 1,  $\langle 1, \xi \rangle$  remains constant at 1 for all  $t \geq 0$ .

( $\Leftarrow$  direction) Suppose  $\int_0^1 \xi(t, dx) = 1$  for all  $t > 0$ . Again take the test function  $f \equiv 1$  but this time with unspecified  $h(t)$ :

$$0 = \frac{d}{dt} \langle 1, \xi \rangle = 0 + \lambda [\langle x, \xi \rangle - \langle 1, \xi \rangle h(t)] = \lambda [\langle x, \xi \rangle - h(t)].$$

For this to hold, we must have  $h(t) = \int_0^1 x \xi(t, dx)$ .  $\square$

The above lemmas imply that solutions  $\mu_t(dx)$  to (3) can be obtained by using formula (8) from lemma 5 and imposing the conservation of measure condition  $\langle 1, \mu_t \rangle \equiv 1$  from lemma 6. We illustrate this with some examples of exactly solvable solutions for special choices of initial data. We will see that the long time behavior of the examples is consistent with results stated in theorem 3.

**Example 3.** Initial measure concentrated at  $x_0 \in [0, 1]$ , i.e.  $\mu_0 = \delta_{x_0}$  Using formula (8),

$$\begin{aligned} \int f(x) \mu_t(dx) &= \int f(x) w_t(x) \delta_{x_0}(\phi_t^{-1}(dx)) = f(\phi_t(x_0)) w_t(\phi_t(x_0)) \\ &= \int f(x) w_t(x) \delta_{\phi_t(x_0)}(dx) \end{aligned}$$

Thus  $\mu_t(dx) = w_t(x) \delta_{\phi_t(x_0)}(dx)$ . Imposing the conservation of measure condition gives  $\mu_t(dx) = \delta_{\phi_t(x_0)}(dx)$ . In other words, an initial delta measure at  $x_0$  moves as a delta measure

along the  $x$  axis with position given by  $\phi_t(x_0)$ , the solution to the negative logistic equation with initial position  $x_0$ .

**Example 4.** Initial uniform density:  $\eta_0(x) = 1$ , i.e  $\mu_0(dx) = dx$  Using formula (9),

$$\eta_t(x) = e^{(\lambda-1)t - \lambda \int_0^t h(z)dz} [e^{-t} + x(1 - e^{-t})]^{(\lambda-2)}$$

Imposing conservation of measure:

$$\begin{aligned} e^{(\lambda-1)t - \lambda \int_0^t h(z)dz} &= \left[ \int_0^1 [e^{-t} + x(1 - e^{-t})]^{(\lambda-2)} dx \right]^{-1} \\ &= \begin{cases} \frac{(\lambda-1)(1-e^{-t})}{1-e^{-(\lambda-1)t}} & \text{if } \lambda \neq 1 \\ \frac{1-e^{-t}}{t} & \text{if } \lambda = 1 \end{cases} \end{aligned}$$

Thus,

$$\eta_t(x) = \begin{cases} \frac{(\lambda-1)(1-e^{-t})}{1-e^{-(\lambda-1)t}} [e^{-t} + x(1 - e^{-t})]^{(\lambda-2)} & \text{if } \lambda \neq 1 \\ \frac{1-e^{-t}}{t} [e^{-t} + x(1 - e^{-t})]^{(\lambda-2)} & \text{if } \lambda = 1 \end{cases}$$

Note that  $\eta_0 \equiv 1$  corresponds to an initial condition satisfying the hypothesis of theorem 3 with  $\alpha = 1$ . As predicted when  $\lambda > 1$ , we obtain  $\eta(t, x) \rightarrow (\lambda-1)x^{\lambda-2} = \text{Beta}(\lambda-1, 1)$  as  $t \rightarrow \infty$ .

The following is an example with  $\alpha > 1$ .

**Example 5.** If  $\eta_0(x) = 2(1-x)$ , i.e.  $\mu_0([1-x, 1]) = x^2$ , then the corresponding  $\alpha$  from theorem 3 is  $\alpha = 2$ .

Using formula (9)

$$\eta_t(x) = 2e^{(\lambda-2)t - \lambda \int_0^t h(z)dz} (1-x) [e^{-t} + x(1 - e^{-t})]^{(\lambda-3)}$$

Imposing the condition in lemma 6 to solve for the  $h(z)$  term

$$\begin{aligned} e^{(\lambda-2)t - \lambda \int_0^t h(z)dz} &= \left[ 2 \int_0^1 (1-x) [e^{-t} + x(1 - e^{-t})]^{(\lambda-3)} dx \right]^{-1} \\ &= \begin{cases} \frac{(\lambda-2)(1-e^{-t})}{2} \left[ \frac{1}{(\lambda-1)(1-e^{-t})} - \frac{e^{-(\lambda-1)t}}{(\lambda-1)(1-e^{-t})} - e^{-(\lambda-2)t} \right]^{-1} & \text{if } \lambda \neq 2 \\ \frac{(1-e^{-t})^2}{2te^{-t}} & \text{if } \lambda = 2 \end{cases} \end{aligned}$$

As predicted by theorem 3 for  $\lambda > 2 = \alpha$ ,

$$\eta_t(x) \rightarrow \frac{1}{2}(\lambda-2)(\lambda-1)(1-x)x^{\lambda-3} = \text{Beta}(\lambda-2, 2)$$

as  $t \rightarrow \infty$ .

**Example 6.**  $\eta_0(x) = \frac{1}{c} \cdot 1_{[0,c]}(x)$  with  $c < 1$ .

Using formula (9)

$$\eta_t(x) = \frac{1}{c} \mathbf{1}\{x \leq \phi_t(c)\} w_t(x) \partial_x \phi_t^{-1}(x)$$

Since  $\phi_t(c) = \frac{ce^{-t}}{1-c+ce^{-t}} \rightarrow 0$  as  $t \rightarrow \infty$ ,  $\eta_t(x) \rightarrow 0$  for any  $x > 0$ . Since  $\eta$  must have total mass 1, it follows that regardless of the value of  $\lambda$ ,  $\eta_t(x)dx \rightarrow \delta_0(dx)$  for any  $c < 1$ . This can also be seen by applying theorem 3 and noting that  $\mu_0([1-c, 1]) = \int_{1-c}^1 \eta_0(x)dx = 0$ .

We end these examples with solutions for  $\mu_0$  that are mixtures of delta measures and densities. First, note that it is straightforward to extend example 3 to the case where  $\mu_0(dx) = \sum a_i \delta_{x_i}(dx)$  is a linear combination of delta measures,  $a_i > 0$  for all  $i$ . Applying (8), we obtain

$$\mu(t, dx) = \sum_i a_i w_t(x) \delta_{\phi_t(x_i)}(dx) = \sum_i a_i(t) \delta_{x_i(t)}(dx)$$

where  $x_i(t) = \phi_t(x_i)$  and  $a_i(t) = a_i w_t(x)|_{x=x_i(t)}$ . Our earlier system of equation (5) is obtained from this and the definitions of  $\phi_t(x)$  and  $w_t(x)$ .

Second, we consider a combination of a delta measure and a density

$$\mu_0(dx) = a \delta_{x_0}(dx) + (1-a) v_0(x) dx$$

Notice that the formula for the solution (8) at first seems linear in the initial condition:

$$\begin{aligned} \int f(x) \mu_t(dx) &= \int f(x) (G_t \mu_0)(dx) \\ &= \int f(x) w_t(x) [a \delta_{\phi_t(x_0)}(dx) + (1-a) v_0(\phi_t^{-1}(x)) \partial_x \phi_t^{-1}(x) dx] \\ &= \int f(x) [a (G_t \delta_{x_0})(dx) + (1-a) (G_t v_0)(dx)] \end{aligned}$$

This gives  $(G_t \mu_0)(dx) = a (G_t \delta_{x_0})(dx) + (1-a) (G_t v_0)(dx)$ . However, this notation is misleading because implicit in the  $G_t$  operator is the function  $h(t)$ , the mean of the overall process over time. Here,  $h(t)$  involves both the delta measure and the density. The solution operator  $G_t$  is therefore not linear for this reason.

Nevertheless, we can still use this formula to obtain expressions for solutions. We illustrate this with a concrete example.

**Example 7.** Take  $x_0 = 0$  and  $v_0(x)$  the density function for  $\text{Beta}(\lambda - \alpha, \alpha)$  with  $\alpha \in (0, \lambda)$ . Using the solution formula and direct calculation, we obtain

$$\begin{aligned} \mu_t(dx) &= a w_t(0) \delta_0(dx) + (1-a) w_t(x) (v_0 \circ \phi_t^{-1})(x) \partial_x \phi_t^{-1}(x) dx \\ &= e^{-\lambda} \int_0^t h(z) dz \{a \delta_0(dx) + (1-a) e^{(\lambda-\alpha)t} v_0(x) dx\} \end{aligned}$$

Note in particular that  $\mu_t$  remains a linear combination of  $\delta_0$  and the Beta distribution. The Beta distribution ultimately dominates because  $\lambda > \alpha$ .

We now use lemma 5 to show that Beta distributions,  $\delta_0$ , and  $\delta_1$  are fixed points for the deterministic partial differential equation and thus provide a proof of lemma 2 announced earlier in this note.

**Proof of lemma 2.** Note that we could prove this lemma by substituting  $\delta_0$ ,  $\delta_1$ , and the Beta distribution into the deterministic partial differential equation (3) and showing the right-hand side equals zero. Instead, we will show that these distribution are fixed points of the solution operator. Let  $v$  be the density of the Beta distribution,

$$v(x) = \frac{1}{B(\lambda - \alpha, \alpha)} x^{\lambda - \alpha - 1} (1 - x)^{\alpha - 1}.$$

The mean of  $v$  is  $\frac{\lambda - \alpha}{\lambda}$ . Using (9)

$$(G_t v)(x) = v\left(\frac{x}{e^{-t} + x(1 - e^{-t})}\right) [e^{-t} + x(1 - e^{-t})]^{\lambda - 2} e^{(\lambda - 1)t - (\lambda - \alpha)t} = v(x)$$

$v$  is therefore a fixed point of the solution operator and hence is a fixed point of the deterministic partial differential equation.

For  $\delta_0$  and  $\delta_1$ , we use example 3 above to obtain  $(G_t \delta_{x_0})(dx) = \delta_{\phi_t(x_0)}(dx)$ . Since  $x_0 = 0$  and  $x_0 = 1$  are fixed points of  $\phi_t$ , it follows that  $\delta_0$  and  $\delta_1$  are fixed points of  $G_t$ .  $\square$

We now prove when the fixed points are stable. We begin with a lemma which gives more general conditions than those given in theorem 3 for the delta measure at zero to attract a given initial condition.

**Lemma 7.** *If for some  $\alpha \geq \lambda > 0$ ,*

$$\lim_{x \rightarrow 0} x^{-\alpha} \mu_0([1 - x, 1]) < \infty$$

*then  $\mu_t \rightarrow \delta_0$  as  $t \rightarrow \infty$ . In particular, this condition holds if  $\mu_0([1 - \varepsilon, 1]) = 0$  for some  $\varepsilon > 0$ .*

To prove this and subsequent results, we will need the following technical lemma.

**Lemma 8.** *Setting  $h(t) = \langle x, \mu_t \rangle$ , the following two implications hold:*

$$\begin{aligned} \int_0^\infty h(t) dt < \infty &\implies h(t) \rightarrow 0 \quad \text{as } t \rightarrow \infty. \\ \int_0^\infty [1 - h(t)] dt < \infty &\implies h(t) \rightarrow 1 \quad \text{as } t \rightarrow \infty. \end{aligned}$$

**Proof of lemma 8.** Since  $h(t) \geq 0$  and  $1 - h(t) \geq 0$ , the only obstruction to the implication is that  $h(t)$  (or  $1 - h(t)$ ) could have ever shorter and shorter intervals were they return to an order one value before returning to a value close to zero. This would require  $h(t)$  to have unbounded derivatives. However this is not possible since

$$\frac{dh}{dt}(t) = -(h - \langle x^2, \mu_t \rangle) + \lambda(\langle x^2, \mu_t \rangle - h^2)$$

from which one easily see that  $-1 \leq \frac{dh}{dt}(t) \leq \lambda$  since  $0 \leq h - \langle x^2, \mu_t \rangle \leq 1$  and  $0 \leq \langle x^2, \mu_t \rangle - h^2 \leq 1$ .

□

**Proof of lemma 7.** As usual let  $h(t) = \langle x, \mu_t \rangle$ . We begin by observing that if

$$\int_0^\infty h(t)dt < \infty$$

then  $h(t) \rightarrow 0$  as  $t \rightarrow \infty$  by lemma 8 and  $\mu_t \rightarrow \delta_0$  as we wish to prove. Thus, we henceforth assume that  $\int_0^\infty h(t)dt = \infty$ . Under this assumption, we will show that for any continuous function  $f$

$$\int_0^1 f(x)\mu_t(dx) \rightarrow f(0) \quad \text{as} \quad t \rightarrow \infty.$$

Since  $f$  is continuous, given any  $\varepsilon > 0$ , there exists a  $\delta > 0$  so that  $|f(x) - f(0)| < \varepsilon$  whenever  $x \leq \delta$ . Hence

$$\left| \int_0^1 f(x)\mu_t(dx) - f(0) \right| \leq \int_0^1 |f(x) - f(0)|\mu_t(dx) \leq \varepsilon + \int_{\delta}^1 |f(x) - f(0)|\mu_t(dx) \quad (13)$$

Now setting

$$\begin{aligned} \int_{\delta}^1 |f(x) - f(0)|\mu_t(dx) &= \int_{\phi_t^{-1}(\delta)}^1 |(f \circ \phi_t)(x) - f(0)| (w_t \circ \phi_t)(x) \mu_0(dx) \\ &\leq 2\|f\|_\infty \int_{\phi_t^{-1}(\delta)}^1 (w_t \circ \phi_t)(y) \mu_0(dy). \end{aligned}$$

Since for all  $y \in [\phi_t^{-1}(\delta), 1]$  and  $t > 0$ , we have

$$(w_t \circ \phi_t)(y) \leq e^{\lambda t - \lambda \int_0^t h(s)ds}$$

we see that

$$\int_{\delta}^1 |f(x) - f(0)|\mu_t(dx) \leq 2\|f\|_\infty e^{\lambda t - \lambda \int_0^t h(s)ds} \mu_0([\phi_t^{-1}(\delta), 1]).$$

Now using the assumptions on  $\mu_0$  and that  $\phi_t^{-1}(\delta) \geq 1 - D e^{-t}$  for some  $D > 0$  and all  $t > 0$ , one has that

$$e^{\lambda t - \lambda \int_0^t h(s)ds} \mu_0([\phi_t^{-1}(\delta), 1]) \leq \hat{D} e^{-(\alpha - \lambda)t - \lambda \int_0^t h(s)ds}$$

for some constant  $\hat{D}$  and all  $t > 0$ . Since  $\alpha \geq \lambda$  and  $\int_0^\infty h(s)ds = \infty$ , this bound converges to zero as  $t \rightarrow \infty$  and the proof is complete as the  $\varepsilon$  in (13) was arbitrary. □

**Proof of theorem 3.** We start with the setting when  $\mu_0(\{1\}) > 0$  and begin by writing  $\mu_t(dx) = a_t \delta_1(dx) + (1 - a_t) \nu_t(dx)$  for some time dependent process  $a_t \in [0, 1]$  with  $a_0 > 0$  and some probability measure valued process  $\nu_t(dx)$ . As usual we define  $h(t) = \langle x, \mu_t \rangle$  and using the representation given in (8), one sees that  $a_t$  solves

$$\frac{da_t}{dt} = \lambda a_t(1 - h(t)) \implies a_t = a_0 \exp\left(\lambda \int_0^t [1 - h(s)]ds\right).$$

Since  $1 - h(t) \geq 0$ , we know that  $\int_0^t [1 - h(s)]ds$  converges as  $t \rightarrow \infty$ . If it converges to  $\infty$  then  $a_t$  also converges to  $\infty$  since  $a_0 > 0$ . However this is impossible since  $a_t \in [0, 1]$  for all  $t \geq 0$ . Thus, we conclude that  $\int_0^t [1 - h(s)]ds < \infty$ . Then lemma 8 implies that  $h(t) \rightarrow 1$  which in turn implies that  $\mu_t \rightarrow \delta_1$  as  $t \rightarrow \infty$ .

We now turn to the setting when  $x^{-\alpha} \mu_0([1 - x, 1]) \rightarrow C > 0$  as  $x \rightarrow 0$ . The case when  $\lambda \leq \alpha$  is already handled by lemma 7 leaving only the case when  $\lambda > \alpha > 0$  to be proven. For  $x \in [0, 1]$ , define  $U(x) = \mu_0([0, x])$ . Since  $\mu_0$  is a probability measure we know that  $U$  has finite variation and is regular in the sense that both the right limit  $U(x^+)$  and the left limit  $U(x^-)$  exist, where  $U(x^\pm) = \lim U(y)$  as  $y \rightarrow^\pm x$ . At the extreme points, only the limit obtained by staying in  $[0, 1]$  is defined.

Now for any smooth function  $f$  of  $[0, 1]$ , we have from (8) that

$$\int_0^1 f(x) \mu_t(dx) = Z_t \int_0^1 f(x) g_t(x) (\mu_0 \phi_t^{-1})(dx) = Z_t \int_0^1 [(fg_t) \circ \phi_t](x) \mu_0(dx)$$

where  $w_t(x)$  has been written as the product of  $g_t(x) = (e^{-t} + x(1 - e^{-t}))^\lambda$  and  $Z_t$  some positive, time dependent normalizing constant. It is enough to show that for some time positive, dependent constant  $K_t$ ,

$$K_t \int_0^1 [(fg_t) \circ \phi_t](x) \mu_0(dx) \rightarrow \int_0^1 f(x) x^{\lambda - \alpha - 1} (1 - x)^{\alpha - 1} dx \quad \text{as } t \rightarrow \infty. \quad (14)$$

Since  $x \mapsto f(x)g_t(x)$  is continuous on  $[0, 1]$ , even if  $U(x)$  has discontinuities the integration by parts formula for Lebesgue–Stieltjes integrals produces

$$\begin{aligned} \int_0^1 [(fg_t) \circ \phi_t](x) \mu_0(dx) &= (fg_t U)(1^-) - (fg_t U)(0^+) - \int_0^1 \partial_x [(fg_t) \circ \phi_t](x) U(x) dx \\ &= [fg_t(U - 1)](1^-) + [fg_t(1 - U)](0^+) + \int_0^1 \partial_x [(fg_t) \circ \phi_t] \\ &\quad \times (x) [1 - U](x) dx. \end{aligned}$$

Here we have used that  $\phi_t$  is continuous with  $\phi_t(1) = 1$  and  $\phi_t(0) = 0$ .

First observe that  $1 - U(1^-) = 0$  since  $\mu_0([1 - x, 1]) \rightarrow 0$  as  $x \rightarrow 0$  by assumption and that  $g_t(0^+) = e^{-\lambda t}$ . Hence

$$[fg_t(1 - U)](1^-) + [fg_t(U - 1)](0^+) = [U(0^+) - 1]f(0)e^{-\lambda t}. \quad (15)$$

Now turning to the integral term, applying the chain rule and changing variables to  $y = \phi_t(x)$  produces

$$\begin{aligned} \int_0^1 \partial_x [(fg_t) \circ \phi_t](x) [1 - U](x) dx &= \int_0^1 [\partial_x (fg_t) \circ \phi_t](x) [1 - U](x) (\partial_x \phi_t)(x) dx \\ &= \int_0^1 [\partial_x (fg_t)](y) [(1 - U) \circ \phi_t^{-1}](y) dy \end{aligned}$$

For any fixed  $x \in (0, 1)$  by direct calculation and use of the assumption on  $\mu_0$ , one sees that

$$\left. \begin{aligned} \partial_x(fg_t)(x) &\rightarrow \partial_x(x^\lambda f)(x) \\ e^{\alpha t}(1 - U(\phi_t^{-1}(x))) &= e^{\alpha t}\mu_0([\phi_t^{-1}(x), 1]) \rightarrow C\left(\frac{1-x}{x}\right)^\alpha \end{aligned} \right\} \text{ as } t \rightarrow \infty.$$

Combining these facts with (15) and the fact that  $e^{-(\lambda-\alpha)t} \rightarrow 0$  as  $t \rightarrow \infty$  since  $\lambda > \alpha$  produces

$$e^{\alpha t} \int_0^1 [(fg_t) \circ \phi_t](x) \mu_0(dx) \rightarrow C \int_0^1 \partial_x(x^\lambda f)(x) \left(\frac{1-x}{x}\right)^\alpha dx \text{ as } t \rightarrow \infty.$$

for some new positive constant  $C$ . Now since integration by parts implies that

$$\frac{1}{\alpha} \int_0^1 \partial_x(x^\lambda f)(x) \left(\frac{1-x}{x}\right)^\alpha dx = \int_0^1 f(x) x^{\lambda-\alpha-1} (1-x)^{\alpha-1} dx$$

the last part of the proof is complete.  $\square$

#### 4. Proofs of weak convergence

The proofs of theorems 1 and 4 follow a standard procedure [3, 12, 13]. Both proofs require: (i) tightness of the sequence of stochastic processes—which implies a subsequential limit, and (ii) uniqueness of this limit. For the tightness of  $\{\mu_t^{m,n}\}_{m,n}$  on  $D([0, T], \mathcal{P}([0, 1]))$ , it is sufficient, by theorem 14.26 in Kallenberg [14] to show that  $\{\langle f, \mu_t^{m,n} \rangle\}$  is tight on  $D([0, T], \mathbb{R})$  for any test function  $f$  from a countably dense subset of continuous, positive functions on  $[0, 1]$ . For the uniqueness of solutions to the partial differential equation in theorem 1, we apply Gronwall's inequality. For uniqueness of solutions to the martingale problem in theorem 4, we apply a Girsanov theorem by Dawson [5].

##### 4.1. Semimartingale property of multilevel selection process

It will be useful for what follows to treat  $\langle f, \mu_t^{m,n} \rangle$  as a semimartingale. Below,  $D_x^+ f$  is the first order difference quotient of  $f$  taken from the right,  $D_x^- f$  is the first order difference quotient of  $f$  taken from the left, and  $D_{xx} f$  is the second order difference quotient.

**Lemma 9.** *For  $f \in C^2([0, 1])$  and  $\mu_t^{m,n}$  with generator  $L^{m,n}$  defined in (1),*

$$\langle f, \mu_t^{m,n} \rangle - \langle f, \mu_0^{m,n} \rangle = A_t^{m,n}(f) + M_t^{m,n}(f) \quad (16)$$

where  $A_t^{m,n}(f)$  is a process of finite variation,  $A_t^{m,n}(f) := \int_0^t a_z^{m,n}(f) dz$ , with

$$\begin{aligned} a_t^{m,n}(f) &= \sum_i \mu_t^{m,n}\left(\frac{i}{n}\right) \frac{i}{n} \left(1 - \frac{i}{n}\right) \left[ \frac{1}{n} D_{xx} f\left(\frac{i}{n}\right) - s D_x^- f\left(\frac{i}{n}\right) \right] \\ &+ wr \left\{ \sum_j \mu_t^{m,n}\left(\frac{j}{n}\right) \frac{j}{n} f\left(\frac{j}{n}\right) - \sum_i \mu_t^{m,n}\left(\frac{i}{n}\right) f\left(\frac{i}{n}\right) \sum_j \mu_t^{m,n}\left(\frac{j}{n}\right) \frac{j}{n} \right\} \end{aligned} \quad (17)$$

and  $M_t^{m,n}(f)$  is a càdlàg martingale with (conditional) quadratic variation

$$\begin{aligned} \langle M^{m,n}(f) \rangle_t &= \frac{1}{m} \int_0^t \left\{ \frac{1}{n} \sum_i \mu_z^{m,n} \left( \frac{i}{n} \right) \frac{i}{n} \left( 1 - \frac{i}{n} \right) \left[ \left( D_x^+ f \left( \frac{i}{n} \right) \right)^2 + (1+s) \left( D_x^- f \left( \frac{i}{n} \right) \right)^2 \right] \right. \\ &\quad \left. + w \sum_{i,j} \mu_z^{m,n} \left( \frac{i}{n} \right) \mu_z^{m,n} \left( \frac{j}{n} \right) \left( 1 + r \frac{j}{n} \right) \left( f \left( \frac{i}{n} \right) - f \left( \frac{j}{n} \right) \right)^2 \right\} dz \end{aligned} \quad (18)$$

**Proof.** By Dynkin's formula (see, for example, lemma 17.21 in [14]),

$$\psi(\mu_t^{m,n}) - \psi(\mu_0^{m,n}) - \int_0^t (L^{m,n} \psi)(\mu_s^{m,n}) ds$$

where  $\psi \in \text{dom}(L^{m,n})$ , is a càdlàg martingale. In particular, this is true for

$$\psi(\mu_t^{m,n}) = F(\langle f, \mu_t^{m,n} \rangle)$$

where  $f \in C^2([0, 1])$  and  $F : \mathbb{R} \rightarrow \mathbb{R}$ . Setting  $F(x) = x$  and plugging this  $f$  into (1):

$$\begin{aligned} (L^{m,n} \langle f, \cdot \rangle)(v) &= \sum_i v \left( \frac{i}{n} \right) \frac{i}{n} \left( 1 - \frac{i}{n} \right) \left[ \frac{1}{n} D_{xx} f \left( \frac{i}{n} \right) - s D_x^- f \left( \frac{i}{n} \right) \right] \\ &\quad + wr \left\{ \sum_j v \left( \frac{j}{n} \right) \frac{j}{n} f \left( \frac{j}{n} \right) - \sum_i v \left( \frac{i}{n} \right) f \left( \frac{i}{n} \right) \sum_j v \left( \frac{j}{n} \right) \frac{j}{n} \right\} \end{aligned}$$

Thus,

$$\langle f, \mu_t^{m,n} \rangle - \langle f, \mu_0^{m,n} \rangle - \int_0^t a_z^{m,n}(f) dz = M_t^{m,n}(f) \quad (19)$$

where  $M_t^{m,n}(f)$  is some martingale and  $a_t^{m,n}(f) = (L^{m,n} \langle f, \cdot \rangle)(\mu_t^{m,n})$ .  $A_t(f)$  is a process of finite variation because for a given  $f$ ,  $a_t^{m,n}(f)$  is uniformly bounded in  $t$ .

Next, setting  $F(x) = x^2$  and plugging this  $\psi$  into (1):

$$\begin{aligned} (L^{m,n} \langle f, \cdot \rangle^2)(v) &= 2 \langle f, v \rangle a_t^{m,n}(f) + \frac{1}{mn} \sum_i v \left( \frac{i}{n} \right) \frac{i}{n} \left( 1 - \frac{i}{n} \right) \left[ \left( D_x^+ f \left( \frac{i}{n} \right) \right)^2 + (1+s) \left( D_x^- f \left( \frac{i}{n} \right) \right)^2 \right] \\ &\quad + \frac{w}{m} \sum_{i,j} v \left( \frac{i}{n} \right) v \left( \frac{j}{n} \right) \left( 1 + r \frac{j}{n} \right) \left( f \left( \frac{i}{n} \right) - f \left( \frac{j}{n} \right) \right)^2 \end{aligned}$$

Thus,

$$\langle f, \mu_t^{m,n} \rangle^2 - \langle f, \mu_0^{m,n} \rangle^2 - \int_0^t c_z^{m,n}(f) dz = \text{martingale} \quad (20)$$

where  $c_t^{m,n}(f) = (L^{m,n} \langle f, \cdot \rangle^2)(\mu_t^{m,n})$ .

Alternatively, take  $Y_t = \langle f, \mu_t^{m,n} \rangle$  and apply Ito's formula (for example, p78 in [18]) to  $Y_t^2$  to obtain

$$\langle f, \mu_t^{m,n} \rangle^2 - \langle f, \mu_0^{m,n} \rangle^2 = 2 \int_0^t \langle f, \mu_z \rangle a_z^{m,n}(f) dz + [M^{m,n}(f)]_t + \text{martingale} \quad (21)$$

where  $[M^{m,n}(f)]_t$  is the quadratic variation process of  $M_t^{m,n}$ . Since  $\langle M^{m,n}(f) \rangle_t$  is the compensator of  $[M^{m,n}(f)]_t$ ,

$$[M^{m,n}(f)]_t - \langle M^{m,n}(f) \rangle_t$$

is a martingale. Thus,

$$\langle f, \mu_t^{m,n} \rangle^2 - \langle f, \mu_0^{m,n} \rangle^2 - 2 \int_0^t \langle f, \mu_z^{m,n} \rangle a_z^{m,n}(f) dz - \langle M^{m,n}(f) \rangle_t = \text{martingale} \quad (22)$$

The compensator  $\langle M^{m,n}(f) \rangle_t$  is a predictable process of finite variation (see p118 in [18]). By the Doob–Meyer inequality (p103 in [18]), the martingale in (22) is the same as the martingale in (20). Equating these martingale parts we obtain

$$2 \int_0^t \langle f, \mu_z^{m,n} \rangle a_z^{m,n}(f) dz + \langle M^{m,n}(f) \rangle_t = \int_0^t c_z^{m,n}(f) dz. \quad (23)$$

Substituting in the expressions for  $a_z^{m,n}$  and  $c_z^{m,n}$  then gives the explicit expression for the conditional quadratic variation (18) in the statement of the lemma.  $\square$

#### 4.2. Proof of deterministic limit

To prove theorem 1, we need the two following lemmas. The first uses criteria in Billingsley [2] to show tightness of the sequence of processes  $\langle f, \mu_t^{m,n} \rangle$ . The second uses Gronwall's inequality to show uniqueness of solutions to the limiting system.

**Lemma 10.** *The processes  $\langle f, \mu_t^{m,n} \rangle$ , as a sequence in  $\{(m, n)\}$ , is tight for all positive-valued test functions  $f \in C^1([0, 1])$ .*

**Proof.** By theorem 13.2 in [2], a sequence of probability measures  $\{P_n\}$  on  $D([0, T], \mathbb{R}^+)$  is tight if and only if (i) for all  $\eta > 0$ , there exists  $a$  such that

$$P_n \left( x : \sup_{t \in [0, T]} |x(t)| \geq a \right) \leq \eta \text{ for } n \geq 1$$

and (ii) for all  $\varepsilon > 0$  and  $\eta > 0$ , there exists  $\delta \in (0, 1)$  and  $n_0$  such that

$$P_n(x : w'_x(\delta) \geq \varepsilon) \leq \eta \text{ for all } n > n_0$$

where  $w'$  is the modulus of continuity for càdlàg processes and is defined

$$w'_x(\delta) := \inf_{\{t_i\}} \max_{1 \leq i \leq v} \sup_{s, t \in [t_{i-1}, t_i]} |x(s) - x(t)|$$

where  $\{t_i\}$  is a partition of  $[0, T]$  such that  $\max_i \{t_i - t_{i-1}\} \leq \delta$  and  $x \in D([0, T], \mathbb{R}^+)$  is distributed according to  $P_n$ .

First, note that since  $\mu_t^{m,n}$  is a probability measure, we have

$$|\langle f, \mu_t^{m,n} \rangle| \leq \|f\|_\infty$$

for all  $t, m$ , and  $n$ . Thus, (i) holds.

For (ii), we have by Markov's inequality:

$$P_{m,n}(w'(\delta) \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}_{m,n}(w'(\delta)) \quad (24)$$

where  $w'(\delta) := w'_{\langle f, \mu_t^{m,n} \rangle}(\delta)$ . We will use the fact that  $\langle f, \mu_t^{m,n} \rangle$  is a pure jump process to bound the right-hand side. The process  $\langle f, \mu_t^{m,n} \rangle$  has two types of jumps: nearest-neighbor, and occupied-site jumps. Nearest-neighbor jumps occur at rate

$$\sum_i m \mu_t^{m,n} \left( \frac{i}{n} \right) i \left( 1 - \frac{i}{n} \right) (2 + s) \leq \frac{mn}{4} (2 + s)$$

and have magnitude

$$|\langle f, \mu_t^{m,n} \rangle - \langle f, \mu_t^{m,n} \rangle| = \left| \left\langle f, \mu_t^{m,n} + \frac{1}{m} \left( \delta_{\frac{i \pm 1}{n}} - \delta_{\frac{i}{n}} \right) \right\rangle - \langle f, \mu_t^{m,n} \rangle \right| \leq \frac{1}{mn} \max_i |D_x^- f(\frac{i}{n})|$$

Occupied-site jumps occur at rate

$$\sum_{i,j} m \mu_t^{m,n} \left( \frac{i}{n} \right) \mu_t^{m,n} \left( \frac{j}{n} \right) \left( 1 + r \frac{j}{n} \right) \leq m(1 + r)$$

and have magnitude

$$|\langle f, \mu_t^{m,n} \rangle - \langle f, \mu_t^{m,n} \rangle| = \left| \left\langle f, \mu_t^{m,n} + \frac{1}{m} \left( \delta_j - \delta_i \right) \right\rangle - \langle f, \mu_t^{m,n} \rangle \right| \leq \frac{2}{m} \|f\|_\infty$$

Putting this together,

$$\begin{aligned} \mathbb{E}_{m,n}(w'(\delta)) &\leq \mathbb{E}_{m,n}[\text{number of nearest-neighbor jumps in time } \delta] \cdot \frac{1}{mn} \max_i |D_x^- f(\frac{i}{n})| \\ &\quad + \mathbb{E}_{m,n}[\text{number of occupied-site jumps in time } \delta] \cdot 2 \frac{1}{m} \|f\|_\infty \\ &\leq \frac{mn}{4} (2 + s) \delta \frac{1}{mn} \max_i |D_x^- f(\frac{i}{n})| + m(1 + r) \delta \frac{2}{m} \|f\|_\infty \\ &= \left\{ \frac{2 + s}{4} \max_i |D_x^- f(\frac{i}{n})| + 2(1 + r) \|f\|_\infty \right\} \delta \end{aligned}$$

Because  $f \in C^1([0, 1])$ , the expression in curly brackets is uniformly bounded by  $C_f$ , a constant that depends on  $f$  but not on  $m$  nor  $n$ . Substituting the above into (24) we get that for  $\delta < \frac{\varepsilon \eta}{C_f}$ ,

$$P_{m,n}(w'(\delta) \geq \varepsilon) \leq \eta$$

for all  $m$  and  $n$ . Thus, both conditions for tightness are satisfied and  $\langle f, \mu_t^{m,n} \rangle$  is tight.  $\square$

**Lemma 11.** *The integro-partial differential equation (3) in theorem 1 has a unique solution.*

**Proof.** Suppose  $\mu_t$  satisfies (3). Fix  $t \geq 0$  and let  $\psi_t(x)$  be a function of time  $t$  and space  $x$ . By the chain rule and the differential equation (3),

$$\begin{aligned} \frac{d}{dt} \langle \psi_t, \mu_t \rangle &= \frac{d}{dz} \langle \psi_z, \mu_t \rangle \Big|_{z=t} + \frac{d}{dz} \langle \psi_t, \mu_z \rangle \Big|_{z=t} \\ &= \langle \frac{\partial}{\partial t} \psi_t, \mu_t \rangle - \langle sx(1-x) \frac{\partial \psi_t}{\partial x}, \mu_t \rangle + wr [\langle x\psi_t, \mu_t \rangle - \langle \psi_t, \mu_t \rangle \langle x, \mu_t \rangle] \\ \langle \psi_t, \mu_t \rangle &= \langle \psi_0, \mu_0 \rangle + \int_0^t \langle \frac{\partial}{\partial z} \psi_z(x) + G\psi_z(x), \mu_z \rangle dz \\ &\quad + wr \int_0^t \langle x\psi_z, \mu_z \rangle - \langle \psi_z, \mu_z \rangle \langle x, \mu_z \rangle dz \end{aligned} \quad (25)$$

where  $Gf = -sx(1-x) \frac{\partial}{\partial x} f$ . Let  $P_t$  be the semigroup operator associated with  $G$ . In fact, using the method of characteristics (or lemma 5 with  $\lambda = 0$ ),

$$P_t f = f \left( \frac{x e^{-st}}{1 - x + x e^{-st}} \right) \quad (26)$$

Now, set  $\psi_z(x) = P_{t-z}f(x)$  for  $0 \leq z \leq t$ , where  $f \in C^1([0, 1])$  is some test function. Substituting this into (25), we have

$$\begin{aligned} \langle P_0 f, \mu_t \rangle &= \langle Pf, \mu_0 \rangle + \int_0^t \left\langle \frac{\partial}{\partial z} P_{t-z} f(x) + G P_{t-z} f(x), \mu_z \right\rangle dz \\ &\quad + \int_0^t wr [\langle x P_{t-z} f, \mu_z \rangle - \langle P_{t-z} f, \mu_z \rangle \langle x, \mu_z \rangle] dz \\ \langle f, \mu_t \rangle &= \langle Pf, \mu_0 \rangle + \int_0^t wr [\langle x P_{t-z} f, \mu_z \rangle - \langle P_{t-z} f, \mu_z \rangle \langle x, \mu_z \rangle] dz \end{aligned} \quad (27)$$

since  $\frac{\partial}{\partial z} P_{t-z} f = -G P_{t-z} f$ . Thus, any  $\mu_t$  that satisfies (3) also satisfies (27). We show that (27) has a unique solution, which in turn implies that (3) has a unique solution.

Suppose  $\mu_t$  and  $\nu_t$  both satisfy (27), with  $\mu_0 = \nu_0$ . Let  $t \geq 0$ .

$$\begin{aligned} \|\mu_t - \nu_t\|_{TV} &= \sup_{\|f\|_\infty \leq 1} |\langle f, \mu_t \rangle - \langle f, \nu_t \rangle| \\ &= \sup_{\|f\|_\infty \leq 1} \left\{ \int_0^t wr [\langle x P_{t-z} f, \mu_z - \nu_z \rangle + wr [\langle x, \mu_z \rangle \langle P_{t-z} f, \mu_z \rangle - \langle x, \nu_z \rangle \langle P_{t-z} f, \nu_z \rangle]] dz \right\} \end{aligned} \quad (28)$$

We can bound the first term in the integrand by

$$wr |\langle x P_{t-z} f, \mu_z - \nu_z \rangle| \leq wr \|\mu_z - \nu_z\|_{TV}$$

because  $\|x P_{t-z} f\|_\infty \leq \|P_{t-z} f\|_\infty \leq \|f\|_\infty \leq 1$ , where the first inequality follows from  $x \in [0, 1]$  and the second from (26). For the second term in the integrand of (28), add and subtract  $\langle x, \nu_z \rangle \langle P_{t-z} f, \mu_z \rangle$ :

$$\begin{aligned}
wr|\langle x, \mu_z \rangle \langle P_{t-z}f, \mu_z \rangle - \langle x, \nu_z \rangle \langle P_{t-z}f, \nu_z \rangle| &= wr|\langle x, \mu_z - \nu_z \rangle \langle P_{t-z}f, \mu_z \rangle + \langle x, \nu_z \rangle \langle P_{t-z}f, \mu_z - \nu_z \rangle| \\
&\leq wr(\|P_{t-z}f\|_\infty \|\mu_z - \nu_z\|_{TV} + \|\mu_z - \nu_z\|_{TV}) \\
&\leq wr(\|f\|_\infty + 1) \|\mu_z - \nu_z\|_{TV}
\end{aligned}$$

again, the inequalities follow from  $x \in [0, 1]$ ,  $\|P_t f\|_\infty \leq \|f\|_\infty$  and also that  $\mu_z$  and  $\nu_z$  are probability measures. Substituting this back into (28),

$$\|\mu_t - \nu_t\|_{TV} \leq \int_0^t 3wr \|\mu_z - \nu_z\|_{TV} dz$$

By Gronwall's inequality,  $\|\mu_t - \nu_t\|_{TV} = 0$ , so we have uniqueness.  $\square$

**Proof of theorem 1.** The uniqueness of the limit is given by lemma 11 and the tightness of the process by lemma 10. It remains to show that  $\{\langle f, \mu_t^{m,n} \rangle\}_{m,n}$  converges to the solution of (3). Recall from lemma 9 that

$$\langle f, \mu_t^{m,n} \rangle - \langle f, \mu_0^{m,n} \rangle = A_t^{m,n}(f) + M_t^{m,n}(f)$$

Since tightness implies relative compactness (Prohorov's theorem), there exists a subsequence of  $\mu_t^{m,n}$  that converges to a limit, call it  $\mu_t$ . Thus,  $\langle f, \mu_t^{m,n} \rangle \rightarrow \langle f, \mu_t \rangle$ . We also have  $\langle f, \mu_0^{m,n} \rangle \rightarrow \langle f, \mu_0 \rangle$  by assumption. In addition,

$$\begin{aligned}
A_t^{m,n}(f) &= \int_0^t \left\{ \sum_i \mu_z^{m,n} \left( \frac{i}{n} \right) \frac{i}{n} \left( 1 - \frac{i}{n} \right) \left[ \frac{1}{n} D_{xx} f \left( \frac{i}{n} \right) - s D_x^- f \left( \frac{i}{n} \right) \right] \right. \\
&\quad \left. + wr \left[ \sum_j \mu_z^{m,n} \left( \frac{j}{n} \right) \frac{j}{n} f \left( \frac{j}{n} \right) - \sum_i \mu_z^{m,n} \left( \frac{i}{n} \right) f \left( \frac{i}{n} \right) \sum_j \mu_z^{m,n} \left( \frac{j}{n} \right) \frac{j}{n} \right] \right\} dz \\
&\rightarrow \int_0^t \left\{ \langle -x(1-x)s \frac{df}{dx}, \mu_z \rangle + wr [\langle xf(x), \mu_z \rangle - \langle f(x), \mu_z \rangle \langle x, \mu_z \rangle] \right\} dz \\
&=: A_t(f)
\end{aligned}$$

The factor of  $\frac{1}{m}$  in the quadratic variation (18) implies that  $M_t^{m,n} \rightarrow 0$  as  $m, n \rightarrow \infty$ . Therefore,

$$\langle f, \mu_t \rangle - \langle f, \mu_0 \rangle = A_t(f)$$

or,

$$\frac{d}{dt} \langle f, \mu_t \rangle = \langle -x(1-x)s \frac{df}{dx}, \mu_t \rangle + wr [\langle xf(x), \mu_t \rangle - \langle f(x), \mu_t \rangle \langle x, \mu_t \rangle]$$

$\square$

#### 4.3. Proof of Fleming–Viot limit

The elementary proof for tightness in theorem 1 does not easily carry over for the case of theorem 4. We thus use a criterion by Aldous [1] to prove tightness for the martingale part of the stochastic process.

First, consider the semimartingale formulation of  $\langle f, \mu_t^{m,n} \rangle$  (16) with the rescaled parameters  $s = \frac{\sigma}{n}$  and  $\rho = \frac{r}{m}$ . Let  $E_t^{m,n}(f) := \int_0^t e_t^{m,n}(f) dz$  and  $N_t^{m,n}(f)$  denote the drift and martingale parts of  $\langle f, \nu_t^{m,n} \rangle$ , the rescaled process. Then

$$\begin{aligned} E_t^{m,n}(f) &= \int_0^t \sum_i \nu_z^{m,n}\left(\frac{i}{n}\right) \frac{i}{n} \left(1 - \frac{i}{n}\right) \left[ D_{xx} f\left(\frac{i}{n}\right) - \sigma D_x^- f\left(\frac{i}{n}\right) \right] \\ &\quad + w\rho \frac{n}{m} \left\{ \sum_j \nu_z^n\left(\frac{j}{n}\right) \frac{j}{n} f\left(\frac{j}{n}\right) - \sum_i \nu_z^n\left(\frac{i}{n}\right) f\left(\frac{i}{n}\right) \sum_j \nu_z^n\left(\frac{j}{n}\right) \frac{j}{n} \right\} dz \quad (29) \end{aligned}$$

and

$$\begin{aligned} \langle N^{m,n}(f) \rangle_t &= \int_0^t \left\{ \frac{n}{m^2} \sum_i \nu_z^{m,n}\left(\frac{i}{n}\right) \frac{i}{n} \left(1 - \frac{i}{n}\right) \left[ \left( D_x^+ f\left(\frac{i}{n}\right) \right)^2 + \left(1 + \frac{\sigma}{n}\right) \left( D_x^- f\left(\frac{i}{n}\right) \right)^2 \right] \right. \\ &\quad \left. + w \frac{n}{m} \sum_{i,j} \nu_z^{m,n}\left(\frac{i}{n}\right) \nu_t^{m,n}\left(\frac{j}{n}\right) \left(1 + \frac{\rho}{m} \frac{j}{n}\right) \left( f\left(\frac{i}{n}\right) - f\left(\frac{j}{n}\right) \right)^2 \right\} dz \quad (30) \end{aligned}$$

**Lemma 12.** *The processes  $\langle f, \nu_t^{m,n} \rangle$ , as a sequence in  $\{(m, n)\}$ , is tight for all  $f \in C^2([0, 1])$ .*

**Proof.** Since  $\langle f, \nu_t^{m,n} \rangle = E_t^{m,n}(f) + N_t^{m,n}(f)$ , it suffices, by the triangle inequality applied to Billingsley's tightness criterion (theorem 13.2 in [2]), to show tightness of  $E_t^{m,n}(f)$  and  $N_t^{m,n}(f)$  separately.

For the tightness of the finite variation term  $E_t^{m,n}(f)$ :

$$\begin{aligned} |e_t^{m,n}(f)| &\leq \frac{1}{4} \sum_i \nu_z^{m,n}\left(\frac{i}{n}\right) \left[ \left| D_{xx} f\left(\frac{i}{n}\right) \right| + \sigma \left| D_x^- f\left(\frac{i}{n}\right) \right| \right] \\ &\quad + w\rho \frac{n}{m} \left\{ \sum_j \nu_z^n\left(\frac{j}{n}\right) \frac{j}{n} |f\left(\frac{j}{n}\right)| + \sum_i \nu_z^n\left(\frac{i}{n}\right) |f\left(\frac{i}{n}\right)| \sum_j \nu_z^n\left(\frac{j}{n}\right) \frac{j}{n} \right\} \end{aligned}$$

For a given  $\gamma > 0$ , we can choose  $n$  and  $m$  sufficiently large such that  $\frac{n}{m} \in (\theta - \gamma, \theta + \gamma)$ ,  $|D_{xx} f(\frac{i}{n})| \leq \|f''\|_\infty + \gamma$ , and  $|D_x^- f(\frac{i}{n})| \leq \|f'\|_\infty + \gamma$ . We thus obtain

$$|e_t^{m,n}(f)| \leq \frac{1}{4} [\|f''\|_\infty + \gamma + \sigma(\|f'\|_\infty + \gamma)] + 2w\rho(\theta + \gamma) \|f\|_\infty$$

There are only a finite number of  $m$  and  $n$  for which this condition is not satisfied. Taking the maximum of the right-hand side of the above equation with the value of  $|e_t^{m,n}(f)|$  for such  $m$  and  $n$ , we obtain that for all  $m$  and  $n$ ,

$$|e_t^{m,n}(f)| \leq G_f$$

and therefore

$$\sup_{t \in [0, T]} |E_t^{m,n}(f)| \leq G_f T$$

where  $G_f$  is a constant that depends on  $f$ . Using the same conditions for tightness as in the proof of theorem 1, condition (i) is satisfied because  $E_t^{m,n}(f)$  is bounded uniformly in  $t$ ,  $m$ , and  $n$ . Condition (ii) is satisfied because  $|E_{t+\delta}^{m,n} - E_t^{m,n}| \leq \delta G_f$  for all  $t$ ,  $m$ , and  $n$  and therefore we can always choose  $\delta$  to be sufficiently small so that  $|E_{t+\delta}^{m,n} - E_t^{m,n}| \leq \varepsilon$  for some prescribed  $\varepsilon$ .

We will show tightness for the martingale part  $\langle N_t^{m,n}(f) \rangle_t$ , using Aldous' tightness condition (we use the result as stated in [9]). First, note that by equation (30),

$$\langle N_t^{m,n}(f) \rangle_t \leq J_f t$$

for  $f \in C^2([0, 1])$ , where  $J_f$  is a constant that depends on  $f$ . Thus for fixed  $t$ ,

$$\begin{aligned} P_{m,n}(|N_t^{m,n}(f)| > a) &\leq \frac{1}{a} \mathbb{E}_{m,n} |N_t^{m,n}(f)| \\ &\leq \frac{1}{a} (\mathbb{E}_{m,n} (N_t^{m,n}(f))^2)^{1/2} = \frac{1}{a} (\mathbb{E}_{m,n} \langle N_t^{m,n}(f) \rangle_t)^{1/2} \leq \frac{\sqrt{J_f t}}{a} \end{aligned}$$

Given  $\varepsilon > 0$ , choose  $a > \frac{\sqrt{J_f t}}{\varepsilon}$  and we have that  $N_t^{m,n}(f)$  is tight for each  $t$ . Next, let  $\tau$  be a stopping time, bounded by  $T$ , and let  $\varepsilon > 0$ . For  $\kappa > 0$ ,

$$P_{m,n}(|N_{\tau+\kappa}^{m,n}(f) - N_{\tau}^{m,n}(f)| \geq \varepsilon) \leq \frac{1}{\varepsilon} \mathbb{E}_{m,n} |N_{\tau+\kappa}^{m,n}(f) - N_{\tau}^{m,n}(f)|$$

Now (suppressing subscripts on expected value for clarity),

$$\begin{aligned} \mathbb{E} |N_{\tau+\kappa}^{m,n}(f) - N_{\tau}^{m,n}(f)| &\leq [\mathbb{E} (N_{\tau+\kappa}^{m,n}(f) - N_{\tau}^{m,n}(f))^2]^{1/2} \\ &= [\mathbb{E} (N_{\tau+\kappa}^{m,n}(f)^2 - N_{\tau}^{m,n}(f)^2 + 2N_{\tau}^{m,n}(f)(N_{\tau}^{m,n}(f) - N_{\tau+\kappa}^{m,n}(f)))]^{1/2} \\ &= [\mathbb{E} (\langle N^{m,n}(f) \rangle_{\tau+\kappa} - \langle N^{m,n}(f) \rangle_{\tau})]^{1/2} \\ &\leq \sqrt{J_f \kappa} \end{aligned}$$

Hence,

$$P_{m,n}(|N_{\tau+\kappa}^{m,n}(f) - N_{\tau}^{m,n}(f)| \geq \varepsilon) \leq \frac{1}{\varepsilon} \sqrt{J_f \kappa}$$

By taking  $\kappa < \frac{\varepsilon^4}{\sqrt{J_f}}$ , we satisfy the conditions of Aldous' stopping criterion.  $\square$

**Lemma 13.** *The martingale problem (6) and (7) has a unique solution.*

**Proof.** The martingale problem with  $V(t, \nu, x) = 0$  corresponds to a neutral Fleming–Viot with linear mutation operator. Its uniqueness has previously been established (see for example [5]). To show uniqueness for nontrivial  $V$ , we use a Girsanov-type transform by Dawson [5]. It suffices to check that

$$\sup_{t, \mu, x} |V(t, \mu, x)| \leq V_0 \text{ (a constant)} \quad (31)$$

In our case,  $V(t, \mu, x) = x$  and since  $x \in [0, 1]$ , the condition is satisfied and the martingale problem has a unique solution.  $\square$

**Proof of theorem 4.** The uniqueness of the limit is given by lemma 13 and the tightness of the process by lemma 12. To see that the limit is the martingale problem stated in theorem 4, note that for a fixed  $t$ ,

$$\begin{aligned} E_t^{m,n}(f) &\longrightarrow \int_0^t \int_0^1 x(1-x) \left[ \frac{\partial^2}{\partial x^2} f(x) - \sigma \frac{\partial}{\partial s} f(x) \right] \nu_z(dx) \\ &\quad \times w \rho \theta \left\{ \int_0^1 x f(x) \nu_z(dx) - \int_0^1 f(x) \nu_z(dx) \int_0^1 x \nu_z(dx) \right\} dz \end{aligned}$$

as  $n, m \rightarrow \infty$  and

$$\langle N^{m,n}(f) \rangle_t \longrightarrow \int_0^t w \theta \int_0^1 \int_0^1 (f(x) - f(y))^2 \nu_z(dx) \nu_z(dy) dz$$

Finally, notice that

$$\begin{aligned} &\int_0^1 \int_0^1 (f(x) - f(y))^2 \nu_z(dx) \nu_z(dy) \\ &= 2 \int_0^1 \int_0^1 f(x)^2 \nu_z(dx) \nu_z(dy) - 2 \int_0^1 \int_0^1 f(x) f(y) \nu_z(dx) \nu_z(dy) \\ &= 2 \int_0^1 \int_0^1 f(x) f(y) \nu_z(dx) [\delta_x(dy) - \nu_z(dy)] \end{aligned}$$

and

$$\int_0^1 x f(x) \nu_z(dx) - \int_0^1 f(x) \nu_z(dx) \int_0^1 x \nu_z(dx) = \int_0^1 \int_0^1 f(x) y \nu_z(dx) [\delta_x(dy) - \nu_z(dy)]$$

satisfying the form of the martingale problem in the theorem.  $\square$

## Acknowledgments

The authors would like to thank Mike Reed and Katia Koelle for their roles in the collaboration out of which this paper's central model grew. We would also like to thank Rick Durrett of a number of useful discussions. SL would further like to thank Don Dawson, who sent her an early version of his notes on this topic [5], and Sylvie Méléard, who took time out of a conference to explain to her the elements of the proof for weak convergence. JCM would like to thank the NSF for its support though DMS-08-54879. SL would like to thank support from the NSF (grants NSF-EF-08-27416 and DMS-0942760), NIH (grant R01-GM094402), and the Simons Institute for the Theory of Computing.

## References

- [1] Aldous D 1978 Stopping times and tightness *Ann. Probab.* **6** 335–40
- [2] Billingsley P 1999 *Convergence of Probability Measures* 2nd edn (New York: Wiley-Interscience)
- [3] Champagnat N, Ferrière R and Méléard S 2006 Unifying evolutionary dynamics: from individual stochastic processes to macroscopic models *Theor. Population Biol.* **69** 297–321

- [4] Dawson D A and Greven A 1999 Hierarchically interacting Fleming–Viot processes with selection and mutation: multiple space time scale analysis and quasi-equilibria *Electron. J. Probab.* **4** 1–81
- [5] Dawson D A 2010 Introductory lecture on stochastic population systems *Technical Report Series of the Laboratory for Research in Statistics and Probability Technical Report* 451, Carleton University—University of Ottawa
- [6] Dawson D A and Hochberg K J 1991 A multilevel branching model *Adv. Appl. Probab.* **23** 701–15
- [7] Donnelly P and Kurtz T G 1999 Genealogical processes for Fleming–Viot models with selection and recombination *Ann. Appl. Probab.* **9** 1091–148
- [8] Durrett R 2008 *Probability Models for DNA Sequence Evolution* 2nd edn (New York: Springer)
- [9] Etheridge A 2000 *An Introduction to Superprocesses* (Providence, RI: American Mathematical Society)
- [10] Ethier S N and Kurtz T G 1993 Fleming–Viot processes in population genetics *SIAM J. Control Optim.* **31** 345–86
- [11] Fleming W H and Viot M 1979 Some measure-valued markov processes in population genetics theory *Indiana Univ. Math. J.* **28** 817–43
- [12] Fournier N and Méléard S 2004 A microscopic probabilistic description of a locally regulated population and macroscopic approximations *Ann. Appl. Probab.* **14** 1880–919
- [13] Joffe A and Metivier M 1986 Weak convergence of sequences of semimartingales with applications to multitype branching processes *Adv. Appl. Probab.* **18** 20–65
- [14] Kallenberg O 1997 *Foundations of Modern Probability* 1st edn (New York: Springer)
- [15] Luo S 2014 A unifying framework reveals key properties of multilevel selection *J. Theor. Biol.* **341** 41–52
- [16] Méléard S and Roelly S 2011 A host-parasite multilevel interacting process and continuous approximations (arXiv:1101.4015)
- [17] Pinchover Y and Rubinstein J 2005 *An Introduction to Partial Differential Equations* (Cambridge: Cambridge University Press)
- [18] Protter P E 2004 *Stochastic Integration and Differential Equations* 2nd edn (New York: Springer)