J. Math. Kyoto Univ. (JMKYAZ)
40-2 (2000) 337-361

A Fleming–Viot process with unbounded selection

By

Stewart N. ETHIER and Tokuzo SHIGA

Abstract

Tachida (1991) proposed a discrete-time model of nearly neutral mutation in which the selection coefficient of a new mutant has a fixed normal distribution with mean 0. The usual diffusion approximation leads to a probability-measure-valued diffusion process, known as a Fleming–Viot process, with the unusual feature of an unbounded selection intensity function. Although the existence of such a diffusion has been proved by Overbeck *et al.* (1995) using Dirichlet forms, we can now characterize the process via the martingale problem. This leads to a limit theorem justifying the diffusion approximation, using a stronger than usual topology on the state space. Also established are existence, uniqueness, and reversibility of the stationary distribution of the Fleming–Viot process.

1. Introduction

Tachida's (1991) nearly neutral mutation model (or normal-selection model) is most easily described in terms of a Fleming–Viot process with house-of-cards (or parent-independent) mutation and haploid selection. In particular, the set of possible alleles, known as the type space, is a locally compact, separable metric space E, so the state space for the process is (a subset of) $\mathscr{P}(E)$, the set of Borel probability measures on E; the mutation operator A on $B(E)$, the space of bounded Borel functions on E, is given by

$$Af = \frac{1}{2}\theta(\langle f, \nu_0\rangle - f), \tag{1.1}$$

where $\theta > 0$, $\nu_0 \in \mathscr{P}(E)$, and $\langle f, \mu\rangle := \int_E f\, d\mu$; and the selection intensity (or scaled selection coefficient) for allele $x \in E$ is $h(x)$, where h is a Borel function on E. More specifically, Tachida's model effectively assumes that

$$E = \mathbf{R}, \qquad \nu_0 = N(0, \sigma_0^2), \qquad h(x) \equiv x, \tag{1.2}$$

where $\sigma_0^2 > 0$. In other words, the type of an individual is identified with its selection intensity, and that of a new mutant is taken to be normal with mean 0 and variance σ_0^2.

Ethier (1997) derived some properties of what was presumed to be the unique stationary distribution for this process, but a characterization of the process, as well as a proof of the uniqueness of the stationary distribution, were left as open

Communicated by Prof. K. Ueno, August 12, 1999

problems. In this paper we treat these and related problems. The difficulty, of course, is that the function h is unbounded. Overbeck *et al.* (1995) were able to prove the existence of Fleming–Viot processes with unbounded selection intensity functions using Dirichlet forms, but they did not address the issues of existence and uniqueness of solutions of the martingale problem. These issues were addressed by Albeverio and Röckner (1995) and Overbeck (1995), but only under conditions that are too restrictive for (1.2).

In a second paper, Tachida (1996) pointed out that there is no biological reason for assuming normality of ν_0, and considered instead a family of distributions on $\mathbf{R}$ symmetric about 0. In this paper we weaken (1.2) as follows. Let E, ν_0, and h be arbitrary, subject to the condition that there exist a continuous function $h_0 : E \mapsto [0, \infty)$ and a constant $\rho_0 \in (1, \infty]$ such that

$$(1.3) \qquad |h| \le h_0, \qquad \langle e^{\rho h_0}, \nu_0 \rangle < \infty \qquad \text{whenever } 0 < \rho < \rho_0.$$

The second condition in (1.3) is simply that $\nu_0 h_0^{-1}$ have a moment generating function that is finite on $(0, \rho_0)$ for some $\rho_0 > 1$. This assumption is in force throughout the paper.

Consider Tachida's (1996) family of distributions (with $E = \mathbf{R}$, $h(x) \equiv x$, and $h_0 = |h|$). Condition (1.3) is satisfied with $\rho_0 = \infty$ in all but one case. The exception is the symmetrized exponential distribution

$$(1.4) \qquad \nu_0(dx) = \frac{1}{\sqrt{2}\sigma_0} \exp(-\sqrt{2}|x|/\sigma_0)dx,$$

which is parametrized here by its standard deviation $\sigma_0 > 0$. If $\sigma_0 < \sqrt{2}$, then condition (1.3) holds with $\rho_0 = \sqrt{2}/\sigma_0$. We will return to this example in Section 4.

The generator of the Fleming–Viot process in question will be denoted by $\mathscr{L}_h$ to emphasize its dependence on the selection intensity function h. (Of course, it also depends on E, ν_0, and θ.) It acts on functions φ on $\mathscr{P}(E)$ of the form

$$(1.5) \qquad \varphi(\mu) = F(\langle f_1, \mu \rangle, \dots, \langle f_k, \mu \rangle) = F(\langle \mathbf{f}, \mu \rangle),$$

where $k \ge 1$, $f_1, \dots, f_k \in \bar{C}(E)$ (the space of bounded continuous functions on E), and $F \in C^2(\mathbf{R}^k)$, according to the formula

$$(1.6) \qquad (\mathscr{L}_h \varphi)(\mu) = \frac{1}{2} \sum_{i,j=1}^{k} (\langle f_i f_j, \mu \rangle - \langle f_i, \mu \rangle \langle f_j, \mu \rangle) F_{z_i z_j}(\langle \mathbf{f}, \mu \rangle) + \sum_{i=1}^{k} (\langle A f_i, \mu \rangle + \langle f_i h, \mu \rangle - \langle f_i, \mu \rangle \langle h, \mu \rangle) F_{z_i}(\langle \mathbf{f}, \mu \rangle).$$

This suffices if h is bounded (e.g. $h \equiv 0$), but if not, because $\langle f_i h, \mu \rangle$ and $\langle h, \mu \rangle$ appear in (1.6), we need to restrict the state space to a suitable subset of $\mathscr{P}(E)$. We take as our state space the set of Borel probability measures μ on E that satisfy the condition imposed on ν_0 in (1.3).

Let us therefore define

$$(1.7)\qquad \mathscr{P}^\circ(E) = \{\mu \in \mathscr{P}(E) : \langle e^{\rho h_0}, \mu\rangle < \infty \text{ for each } \rho \in (0, \rho_0)\}$$

and, for $\mu, \nu \in \mathscr{P}^\circ(E)$,

$$(1.8)\qquad d^\circ(\mu, \nu) = d(\mu, \nu) + \int_{(0,\rho_0)} \left(1 \wedge \sup_{0 \le \rho \le r} |\langle e^{\rho h_0}, \mu\rangle - \langle e^{\rho h_0}, \nu\rangle|\right) e^{-r}\, dr,$$

where d is a metric on $\mathscr{P}(E)$ that induces the topology of weak convergence. Then $(\mathscr{P}^\circ(E), d^\circ)$ is a complete separable metric space and $d^\circ(\mu_n, \mu) \to 0$ if and only if $\mu_n \Rightarrow \mu$ and $\sup_n \langle e^{\rho h_0}, \mu_n\rangle < \infty$ for each $\rho \in (0, \rho_0)$. (This is where we use the continuity of h_0.) Thus, the topology on $\mathscr{P}^\circ(E)$ is somewhat stronger than the topology of weak convergence (if h_0 is unbounded).

Section 2 establishes existence and uniqueness of solutions of the appropriate martingale problem for $\mathscr{L}_h$. Surprisingly, existence is more difficult than uniqueness. Section 3 gives a precise description of Tachida's (1991) model as a measure-valued Wright–Fisher model and proves a weak convergence result that justifies the diffusion approximation of that model by the Fleming–Viot process with generator $\mathscr{L}_h$. The idea of the proof is to show that the Girsanov formula for the Wright–Fisher model converges in some sense to that for the Fleming–Viot process. Section 4 establishes existence, uniqueness, and reversibility of the stationary distribution of the Fleming–Viot process.

Two obvious problems remain unresolved, namely (a) justification of the diffusion approximation of the stationary distribution of the Wright–Fisher model by that of the Fleming–Viot process, and (b) proof of the strong ergodicity of the Fleming–Viot process.

2. Characterization of the process

Let $\Omega := C_{(\mathscr{P}(E), d)}[0, \infty)$ have the topology of uniform convergence on compact sets, let $\mathscr{F}$ be the Borel σ-field, let $\{\mu_t, t \ge 0\}$ be the canonical coordinate process, and let $\{\mathscr{F}_t\}$ be the corresponding filtration.

We will need a lemma from Ethier (1997), which is essentially a result of Dawson (1978).

Lemma 2.1. *Let $h_1, h_2 \in B(E)$. If $P \in \mathscr{P}(\Omega)$ is a solution of the martingale problem for $\mathscr{L}_{h_1}$, then*

$$(2.1)\qquad R_t := \exp\left\{\langle h_2, \mu_t\rangle - \langle h_2, \mu_0\rangle - \int_0^t \left[\frac{1}{2}(\langle h_2^2, \mu_s\rangle - \langle h_2, \mu_s\rangle^2) \right.\right.$$
$$\left.\left. + \frac{1}{2}\theta(\langle h_2, \nu_0\rangle - \langle h_2, \mu_s\rangle) + \langle h_1 h_2, \mu_s\rangle - \langle h_1, \mu_s\rangle\langle h_2, \mu_s\rangle\right] ds\right\}$$

is a mean-one $\{\mathscr{F}_t\}$*-martingale on* $(\Omega, \mathscr{F}, P)$. *Furthermore, the measure* $Q \in \mathscr{P}(\Omega)$ *defined by*

$$dQ = R_t\, dP \qquad on\ \mathscr{F}_t, \qquad t \geq 0, \tag{2.2}$$

is a solution of the martingale problem for $\mathscr{L}_{h_1+h_2}$.

We now define

$$\Omega^\circ = C_{(\mathscr{P}^\circ(E), d^\circ)}[0, \infty) \subset \Omega = C_{(\mathscr{P}(E), d)}[0, \infty). \tag{2.3}$$

For each $\mu \in \mathscr{P}(E)$ we denote by $P_\mu \in \mathscr{P}(\Omega)$ the unique solution of the martingale problem for $\mathscr{L}_0$ (i.e., the distribution of the neutral model) starting at μ.

Lemma 2.2. *For each* $\mu \in \mathscr{P}^\circ(E)$, $T > 0$, $\rho \in (0, \rho_0)$, *and* $\lambda > \langle e^{\rho h_0}, \mu\rangle + \frac{1}{2}\theta T \langle e^{\rho h_0}, \nu_0\rangle$,

$$P_\mu\left\{\sup_{0\leq t\leq T} \langle e^{\rho h_0}, \mu_t\rangle > \lambda\right\} \leq \frac{(1 + \frac{1}{2}\theta T)(\langle e^{\rho h_0}, \mu\rangle \vee \langle e^{\rho h_0}, \nu_0\rangle)}{\lambda - \langle e^{\rho h_0}, \mu\rangle - \frac{1}{2}\theta T\langle e^{\rho h_0}, \nu_0\rangle}. \tag{2.4}$$

In particular, given $\mu \in \mathscr{P}^\circ(E)$, *we have* $\sup_{0\leq t\leq T}\langle e^{\rho h_0}, \mu_t\rangle < \infty$ P_μ*-a.s. for each* $\rho \in (0, \rho_0)$ *and* $T > 0$, *and therefore* $P_\mu(\Omega^\circ) = 1$.

Remark. Ethier (1997) assumed in effect that $\rho_0 = \infty$ and used

$$\mathbf{E}^{P_\mu}\left[\sup_{0\leq t\leq T} \langle e^{\rho h_0}, \mu_t\rangle^2\right] \leq (12T + 3)\langle e^{2\rho h_0}, \mu\rangle + \left(12T + \frac{3}{4}\theta^2 T^2\right)\langle e^{2\rho h_0}, \nu_0\rangle \tag{2.5}$$

in place of (2.4). (Actually, the formulation in the previous paper contains a small error, and (2.5) is the corrected version.)

Proof. Fix $\mu \in \mathscr{P}(E)$ and $g \in B(E)$. Note first that

$$\begin{aligned} \mathbf{E}^{P_\mu}[\langle g, \mu_t\rangle] &= \langle U(t)g, \mu\rangle \\ &= e^{-\theta t/2}\langle g, \mu\rangle + (1 - e^{-\theta t/2})\langle g, \nu_0\rangle \leq \langle g, \mu\rangle \vee \langle g, \nu_0\rangle \end{aligned} \tag{2.6}$$

for all $t \geq 0$, where $\{U(t)\}$ is the semigroup on $B(E)$ with generator A as in (1.1). Assume that g is also continuous; then

$$Z^g(t) := \langle g, \mu_t\rangle - \langle g, \mu_0\rangle - \frac{1}{2}\theta\int_0^t (\langle g, \nu_0\rangle - \langle g, \mu_s\rangle)ds \tag{2.7}$$

is a continuous $\{\mathscr{F}_t\}$-martingale on $(\Omega, \mathscr{F}, P_\mu)$. Assume that g is nonnegative as well; then $\langle g, \mu_t\rangle \leq Z^g(t) + \langle g, \mu_0\rangle + \frac{1}{2}\theta t\langle g, \nu_0\rangle$ for all $t \geq 0$. Consequently, given $T > 0$ and $\lambda > \langle g, \mu\rangle + \frac{1}{2}\theta T\langle g, \nu_0\rangle$, we have

$$(2.8)\qquad P_\mu\left\{\sup_{0\le t\le T}\langle g,\mu_t\rangle>\lambda\right\}\le P_\mu\left\{\sup_{0\le t\le T}Z^g(t)>\lambda-\langle g,\mu\rangle-\frac{1}{2}\theta T\langle g,\nu_0\rangle\right\}$$
$$\le\frac{\mathbf{E}^{P_\mu}[Z^g(T)^+]}{\lambda-\langle g,\mu\rangle-\frac{1}{2}\theta T\langle g,\nu_0\rangle}$$
$$\le\frac{\mathbf{E}^{P_\mu}[\langle g,\mu_T\rangle+\frac{1}{2}\theta\int_0^T\langle g,\mu_s\rangle ds]}{\lambda-\langle g,\mu\rangle-\frac{1}{2}\theta T\langle g,\nu_0\rangle}$$
$$\le\frac{(1+\frac{1}{2}\theta T)(\langle g,\mu\rangle\vee\langle g,\nu_0\rangle)}{\lambda-\langle g,\mu\rangle-\frac{1}{2}\theta T\langle g,\nu_0\rangle},$$

where the last inequality uses (2.6).

If we now assume that $\mu\in\mathscr{P}^\circ(E)$ and let $g=e^{\rho h_0}\wedge K$ in (2.8), where $\rho\in(0,\rho_0)$, we obtain (2.4) by letting $K\to\infty$.

Let Ω° have the topology of uniform convergence on compact sets, let $\mathscr{F}^\circ$ be the Borel σ-field, let $\{\mu_t, t\ge 0\}$ be the canonical coordinate process on Ω°, and let $\{\mathscr{F}_t^\circ\}$ be the corresponding filtration. We do not distinguish notationally between the canonical coordinate process on Ω and that on Ω° (the latter is just the restriction to Ω° of the the former), between $P_\mu\in\mathscr{P}(\Omega)$ and its restriction to $\mathscr{F}^\circ$ (note that $\mathscr{F}^\circ\subset\mathscr{F}$), or between R_t of (2.1) and its restriction to Ω°. We temporarily denote R_t by $R_t^{h_1,h_1+h_2}$ to indicate its dependence on h_1 and h_2, which we now allow to be unbounded.

Lemma 2.3. *For each $\mu\in\mathscr{P}^\circ(E)$, $\{R_t^{0,h}, t\ge 0\}$ is a mean-one $\{\mathscr{F}_t^\circ\}$-martingale on $(\Omega^\circ,\mathscr{F}^\circ,P_\mu)$.*

Proof. It is enough to prove the existence of $\delta_0>0$ such that

$$(2.9)\qquad \mathbf{E}^{P_\mu}\left[\frac{R_{t+\delta}^{0,h}}{R_t^{0,h}}\,\middle|\,\mathscr{F}_t^\circ\right]=1$$

whenever $t\ge 0$ and $0<\delta<\delta_0$. For then, given $t>s\ge 0$, choose $s=t_0<t_1<\cdots<t_n=t$ with $\max_{1\le i\le n}(t_i-t_{i-1})<\delta_0$, and argue that

$$(2.10)\qquad \mathbf{E}^{P_\mu}\left[\frac{R_t^{0,h}}{R_s^{0,h}}\,\middle|\,\mathscr{F}_s^\circ\right]=\mathbf{E}^{P_\mu}\left[\prod_{i=1}^n\frac{R_{t_i}^{0,h}}{R_{t_{i-1}}^{0,h}}\,\middle|\,\mathscr{F}_s^\circ\right]=1$$

by successively conditioning on $\mathscr{F}_{t_{n-1}}^\circ,\ldots,\mathscr{F}_{t_1}^\circ$.

Define $h_K=(-K)\vee(h\wedge K)$, and note that, for fixed $t\ge 0$, (2.9) holds for all $\delta>0$ if h is replaced by h_K, and hence

$$(2.11)\qquad \mathbf{E}^{P_\mu}\left[\frac{R_{t+\delta}^{0,h_K}}{R_t^{0,h_K}}e^{\langle h_K,\mu_t\rangle+(1/2)\theta\delta\langle h_K,\nu_0\rangle}\,\middle|\,\mathscr{F}_t^\circ\right]=e^{\langle h_K,\mu_t\rangle+(1/2)\theta\delta\langle h_K,\nu_0\rangle}.$$

The integrand in (2.11) is bounded by

$$\exp\left\{\langle h_K,\mu_{t+\delta}\rangle+\frac{1}{2}\theta\int_t^{t+\delta}\langle h_K,\mu_s\rangle ds\right\} \tag{2.12}$$
$$\leq\exp\left\{\langle h_0,\mu_{t+\delta}\rangle+\frac{1}{2}\theta\int_t^{t+\delta}\langle h_0,\mu_s\rangle ds\right\},$$

so it will suffice by dominated convergence to show that the right side of (2.12) is integrable for δ sufficiently small.

Choose $p\in(1,p_0)$, let $q=p/(p-1)$, and define $\delta_0=2p/(\theta q)$. (This is where we use the assumption that $p_0>1$.) Then, by the Hölder and Jensen inequalities and (2.6),

$$\mathbf{E}^{P_\mu}\left[\exp\left\{\langle h_0,\mu_{t+\delta}\rangle+\frac{1}{2}\theta\int_t^{t+\delta}\langle h_0,\mu_s\rangle ds\right\}\right] \tag{2.13}$$
$$\leq\mathbf{E}^{P_\mu}[\exp\{p\langle h_0,\mu_{t+\delta}\rangle\}]^{1/p}\mathbf{E}^{P_\mu}\left[\exp\left\{\frac{1}{2}\theta q\int_t^{t+\delta}\langle h_0,\mu_s\rangle ds\right\}\right]^{1/q}$$
$$\leq\mathbf{E}^{P_\mu}[\langle e^{ph_0},\mu_{t+\delta}\rangle]^{1/p}\left(\frac{1}{\delta}\int_t^{t+\delta}\mathbf{E}^{P_\mu}[\langle e^{\theta q\delta h_0/2},\mu_s\rangle]ds\right)^{1/q}$$
$$\leq[\langle e^{ph_0},\mu\rangle\vee\langle e^{ph_0},\nu_0\rangle]^{1/p}[\langle e^{\theta q\delta h_0/2},\mu\rangle\vee\langle e^{\theta q\delta h_0/2},\nu_0\rangle]^{1/q}$$
$$\leq\langle e^{ph_0},\mu\rangle\vee\langle e^{ph_0},\nu_0\rangle$$

if $0<\delta<\delta_0$, and the proof is complete.

For each $\mu\in\mathscr{P}^\circ(E)$, Lemma 2.3 allows us to define $Q_\mu\in\mathscr{P}(\Omega^\circ)$ by

$$dQ_\mu=R_t^{0,h}\,dP_\mu\qquad\text{on }\mathscr{F}_t^\circ,\qquad t\geq 0. \tag{2.14}$$

We now show that Q_μ solves the Ω° martingale problem for $\mathscr{L}_h$ starting at μ. (The domain of $\mathscr{L}_h$ is the space of functions φ on $\mathscr{P}^\circ(E)$ of the form (1.5).)

Lemma 2.4. *Let $\mu\in\mathscr{P}^\circ(E)$. Then*

$$M_t^h:=\varphi(\mu_t)-\varphi(\mu_0)-\int_0^t(\mathscr{L}_h\varphi)(\mu_s)ds \tag{2.15}$$

is an $\{\mathscr{F}_t^\circ\}$-martingale on $(\Omega^\circ,\mathscr{F}^\circ,Q_\mu)$ for each $\varphi\in\mathscr{D}(\mathscr{L}_h)$.

Proof. Fix $\mu\in\mathscr{P}^\circ(E)$. Let h_K be as in the preceding proof, and define $Q_\mu^K\in\mathscr{P}(\Omega^\circ)$ as in (2.14) but with h replaced by h_K. Then, by Lemma 2.1, Q_μ^K solves the Ω martingale problem for $\mathscr{L}_{h_K}$ starting at μ. Let $\varphi\in\mathscr{D}(\mathscr{L}_h)$ be arbitrary, define

$$M_t^{h_K}=\varphi(\mu_t)-\varphi(\mu_0)-\int_0^t(\mathscr{L}_{h_K}\varphi)(\mu_s)ds, \tag{2.16}$$

and fix $t\geq 0$ and $\delta>0$. Then

$$\mathbf{E}^{Q_\mu^K}[M_{t+\delta}^{h_K}-M_t^{h_K}|\mathscr{F}_t]=0, \tag{2.17}$$

hence

$$\mathbf{E}^{P_\mu}[(M_{t+\delta}^{h_K} - M_t^{h_K})R_{t+\delta}^{0,h_K}|\mathscr{F}_t] = 0 \tag{2.18}$$

and

$$\mathbf{E}^{P_\mu}\left[(M_{t+\delta}^{h_K} - M_t^{h_K})\frac{R_{t+\delta}^{0,h_K}}{R_t^{0,h_K}} e^{\langle h_K,\mu_t\rangle + (1/2)\theta\delta\langle h_K,\nu_0\rangle}\middle|\mathscr{F}_t\right] = 0. \tag{2.19}$$

Note that the integrand in (2.19) is bounded by a constant multiple of

$$\left(1 + \int_t^{t+\delta}(1 + \langle h_0,\mu_s\rangle)ds\right)\exp\left\{\langle h_0,\mu_{t+\delta}\rangle + \frac{1}{2}\theta\int_t^{t+\delta}\langle h_0,\mu_s\rangle ds\right\}. \tag{2.20}$$

With p, q, and δ_0 as in (2.13), choose $\alpha > 1$ such that $\alpha p < p_0$, and put $\beta = \alpha/(\alpha - 1)$. Then, by Hölder's inequality and the argument used for (2.13), the P_μ-expectation of (2.20) is at most

$$\begin{aligned} &\mathbf{E}^{P_\mu}\left[\left(1 + \delta + \int_t^{t+\delta}\langle h_0,\mu_s\rangle ds\right)^\beta\right]^{1/\beta} \\ &\cdot \mathbf{E}^{P_\mu}\left[\exp\left\{\alpha\langle h_0,\mu_{t+\delta}\rangle + \frac{1}{2}\theta\alpha\int_t^{t+\delta}\langle h_0,\mu_s\rangle ds\right\}\right]^{1/\alpha} \\ &\le \{2^{\beta-1}(1+\delta_0)^\beta + 2^{\beta-1}\delta_0^\beta(\langle h_0^\beta,\mu\rangle \vee \langle h_0^\beta,\nu_0\rangle)\}^{1/\beta} \\ &\cdot \{\langle e^{\alpha p h_0},\mu\rangle \vee \langle e^{\alpha p h_0},\nu_0\rangle\}^{1/\alpha} \end{aligned} \tag{2.21}$$

if $0 < \delta < \delta_0$. For such δ, we conclude that

$$\mathbf{E}^{P_\mu}\left[(M_{t+\delta}^{h} - M_t^{h})\frac{R_{t+\delta}^{0,h}}{R_t^{0,h}} e^{\langle h,\mu_t\rangle + (1/2)\theta\delta\langle h,\nu_0\rangle}\middle|\mathscr{F}_t^\circ\right] = 0. \tag{2.22}$$

We would now like to factor out the $\mathscr{F}_t^\circ$-measurable factors to obtain

$$\mathbf{E}^{P_\mu}[(M_{t+\delta}^{h} - M_t^{h})R_{t+\delta}^{0,h}|\mathscr{F}_t^\circ] = 0 \tag{2.23}$$

and

$$\mathbf{E}^{Q_\mu}[M_{t+\delta}^{h} - M_t^{h}|\mathscr{F}_t^\circ] = 0, \tag{2.24}$$

but first we must show that the integrands in (2.23) and (2.24) are integrable.

Observe that $\{M_t^{h_K}, t \ge 0\}$ is a continuous square-integrable $\{\mathscr{F}_t\}$-martingale on $(\Omega, \mathscr{F}, Q_\mu^K)$ with increasing process $\langle\langle M^{h_K}\rangle\rangle_t = \int_0^t \psi(\mu_s)ds$, where

$$\begin{aligned} 0 \le \psi(\mu_s) &= \sum_{i,j=1}^k (\langle f_i f_j,\mu_s\rangle - \langle f_i,\mu_s\rangle\langle f_j,\mu_s\rangle)F_{z_i}(\langle \mathbf{f},\mu_s\rangle)F_{z_j}(\langle \mathbf{f},\mu_s\rangle) \\ &\le \left(\sum_{i=1}^k \|f_i\|_\infty C_i\right)^2 =: C \end{aligned} \tag{2.25}$$

if φ is as in (1.5) and $C_i := \sup_{\nu \in \mathscr{P}(E)} |F_{z_i}(\langle \mathbf{f}, \nu \rangle)|$. It follows that

$$\mathbf{E}^{P_\mu}[(M_t^{h_K})^2 R_t^{0,h_K}] = \mathbf{E}^{Q_\mu^K}[(M_t^{h_K})^2] = \mathbf{E}^{Q_\mu^K}[\langle\langle M^{h_K} \rangle\rangle_t] \le Ct, \qquad t \ge 0. \tag{2.26}$$

Now apply Fatou's lemma, obtaining

$$\mathbf{E}^{P_\mu}[(M_t^h)^2 R_t^{0,h}] \le Ct, \qquad t \ge 0, \tag{2.27}$$

hence

$$\mathbf{E}^{Q_\mu}[(M_t^h)^2] \le Ct, \qquad t \ge 0. \tag{2.28}$$

This provides the needed justification for (2.23) and (2.24).

As in the preceding proof, once we have (2.24) whenever $t \ge 0$ and $0 < \delta < \delta_0$, we have it for all $t \ge 0$ and $\delta > 0$.

Theorem 2.5. *For each $\mu \in \mathscr{P}^\circ(E)$, the Ω° martingale problem for $\mathscr{L}_h$ starting at μ has one and only one solution.*

Proof. It remains to prove uniqueness. Given $\mu \in \mathscr{P}^\circ(E)$, let $Q_\mu \in \mathscr{P}(\Omega^\circ)$ be a solution of the Ω° martingale problem for $\mathscr{L}_h$ starting at μ. Then $\{R_t^{h,0}, t \ge 0\}$ is an $\{\mathscr{F}_t^\circ\}$ local martingale on $(\Omega^\circ, \mathscr{F}^\circ, Q_\mu)$. In fact, if we define

$$\tau_N = \inf\{t \ge 0 : \langle h_0^2, \mu_t \rangle \ge N\}, \tag{2.29}$$

then $\{R_{t\wedge\tau_N}^{h,0}, t \ge 0\}$ is a mean-one $\{\mathscr{F}_{t\wedge\tau_N}^\circ\}$-martingale on $(\Omega^\circ, \mathscr{F}^\circ, Q_\mu)$. Using essentially Theorem 1.3.5 of Stroock and Varadhan (1979), there exists for each $N \ge 1$ a probability measure P_μ^N on $(\Omega^\circ, \mathscr{F}_{\tau_N}^\circ)$ such that

$$dP_\mu^N = R_{t\wedge\tau_N}^{h,0}\, dQ_\mu \qquad \text{on } \mathscr{F}_{t\wedge\tau_N}^\circ, \qquad t \ge 0. \tag{2.30}$$

Furthermore, by the argument that was used to prove Lemma 2.1,

$$\varphi(\mu_{t\wedge\tau_N}) - \varphi(\mu_0) - \int_0^{t\wedge\tau_N} (\mathscr{L}_0\varphi)(\mu_s)ds \tag{2.31}$$

is an $\{\mathscr{F}_{t\wedge\tau_N}^\circ\}$-martingale on $(\Omega^\circ, \mathscr{F}_{\tau_N}^\circ, P_\mu^N)$ for every $\varphi \in \mathscr{D}(\mathscr{L}_h)$. Again we apply Theorem 1.3.5 of Stroock and Varadhan (1979) to deduce the existence of a probability measure P_μ° on $(\Omega^\circ, \mathscr{F}^\circ)$ such that

$$P_\mu^\circ = P_\mu^N \qquad \text{on } \mathscr{F}_{\tau_N}^\circ, \qquad N \ge 1. \tag{2.32}$$

We claim that

$$\varphi(\mu_t) - \varphi(\mu_0) - \int_0^t (\mathscr{L}_0\varphi)(\mu_r)dr \tag{2.33}$$

is an $\{\mathscr{F}_t^\circ\}$-martingale on $(\Omega^\circ, \mathscr{F}^\circ, P_\mu^\circ)$ for every $\varphi \in \mathscr{D}(\mathscr{L}_h)$. To see this, fix such a φ, let H be a bounded continuous function on $\mathscr{P}^\circ(E)^m$, where $m \ge 1$, and let $0 < s_1 < \cdots < s_m \le s < t$. Then

$$\mathbf{E}^{P_\mu^\circ}\left[\left(\varphi(\mu_{t\wedge\tau_N}) - \varphi(\mu_{s\wedge\tau_N}) - \int_{s\wedge\tau_N}^{t\wedge\tau_N} (\mathscr{L}_0\varphi)(\mu_r)dr\right) H(\mu_{s_1\wedge\tau_N}, \ldots, \mu_{s_m\wedge\tau_N})\right] = 0 \tag{2.34}$$

for each $N \geq 1$, hence

$$(2.35) \qquad \mathbf{E}^{P_\mu^\circ}\left[\left(\varphi(\mu_t) - \varphi(\mu_s) - \int_s^t (\mathscr{L}_0\varphi)(\mu_r)dr\right)H(\mu_{s_1}, \dots, \mu_{s_m})\right] = 0.$$

This proves the claim, and so P_μ°, extended to $(\Omega, \mathscr{F})$ in the obvious way, is a solution of the Ω martingale problem for $\mathscr{L}_0$ starting at μ, and must therefore equal P_μ.

Finally, from

$$(2.36) \qquad dP_\mu = R^{h,0}_{t\wedge\tau_N}\, dQ_\mu \qquad \text{on } \mathscr{F}^\circ_{t\wedge\tau_N}, \qquad t \geq 0,$$

we obtain

$$(2.37) \qquad dQ_\mu = R^{0,h}_{t\wedge\tau_N}\, dP_\mu \qquad \text{on } \mathscr{F}^\circ_{t\wedge\tau_N}, \qquad t \geq 0,$$

and in particular that for each $N \geq 1$, $\mathbf{E}^{Q_\mu}[\varphi(\mu_{t\wedge\tau_N})]$ is uniquely determined for every $\varphi \in \bar{C}(\mathscr{P}^\circ(E))$ and $t \geq 0$, hence the same is true of $\mathbf{E}^{Q_\mu}[\varphi(\mu_t)]$. Thus, the Q_μ-distribution of μ_t is uniquely determined for every $t \geq 0$, implying that the Ω° martingale problem for $\mathscr{L}_h$ starting at μ has a unique solution.

3. Diffusion approximation of the Wright–Fisher model

The motivation for the Fleming–Viot process characterized in Section 2 is that for large populations it approximates Tachida's (1991) model, which was originally formulated as a Wright–Fisher model. In this section we provide a justification for this diffusion approximation. It does not follow from existing results (such as Ethier and Kurtz (1987)) because of the unboundedness of h.

We begin by formulating a Wright–Fisher model that is general enough to include Tachida's model. It depends on several parameters, some of which have already been introduced:

- E (a locally compact, separable metric space) is the set of possible alleles, and is known as the type space.
- M (a positive integer) is the haploid population size.
- u (in $[0,1]$) is the mutation rate (i.e., probability) per gene per generation.
- ν_0 (in $\mathscr{P}(E)$) is the distribution of the type of a new mutant; this is the house-of-cards assumption.
- $w(x)$ (a positive Borel function defined for each $x \in E$) is the fitness of allele x.

The Wright–Fisher model is a Markov chain describing the evolution of the composition of the population of types $(x_1, \dots, x_M) \in E^M$ or, since the order of the types is unimportant, $M^{-1}\sum_{i=1}^M \delta_{x_i} \in \mathscr{P}(E)$. (Here $\delta_x \in \mathscr{P}(E)$ denotes the unit mass at $x \in E$.) Thus, the state space for the process is

$$(3.1) \qquad \mathscr{P}_M(E) := \left\{\frac{1}{M}\sum_{i=1}^M \delta_{x_i} \in \mathscr{P}(E) : (x_1, \dots, x_M) \in E^M\right\}$$

with the topology of weak convergence. Time is discrete and measured in generations. The transition mechanism is specified by

$$\mu := \frac{1}{M}\sum_{i=1}^{M}\delta_{x_i} \mapsto \frac{1}{M}\sum_{i=1}^{M}\delta_{Y_i}, \tag{3.2}$$

where

$$Y_1, \dots, Y_M \text{ are i.i.d. } \mu^{**} \qquad \text{[random sampling]}, \tag{3.3}$$

$$\mu^{**} = (1-u)\mu^* + u\nu_0 \qquad \text{[house-of-cards mutation]}, \tag{3.4}$$

$$\mu^*(\Gamma) = \int_\Gamma w(x)\mu(dx)/\langle w, \mu\rangle \qquad \text{[haploid selection]}. \tag{3.5}$$

(Integrability in (3.5) is not an issue, because μ has finite support.) This suffices to describe the Wright–Fisher model in terms of the parameters listed above.

However, since we are interested in a diffusion approximation, we further assume that

$$u = \frac{\theta}{2M}, \qquad w(x) = \exp\left\{\frac{h(x)}{M}\right\}, \tag{3.6}$$

where θ is a positive constant and h is as in (1.3). (Note the use of the exponential in (3.6). This ensures that $w(x)$ is always positive, in contrast to the more conventional and asymptotically equivalent $w(x) = 1 + h(x)/M$.)

The aim here is to prove, assuming the continuity of h, that convergence in $\mathscr{P}^\circ(E)$ of the initial distributions implies convergence in distribution in Ω° of the sequence of rescaled and linearly interpolated Wright–Fisher models to a Fleming–Viot process with generator $\mathscr{L}_h$. We postpone a careful statement of the result to the end of the section.

The proof requires a moment estimate on the neutral ($h \equiv 0$) Wright–Fisher model that is analogous to Lemma 2.2 for the neutral diffusion model, as well as a Girsanov-type formula for the Wright–Fisher model that is a bit different from Lemmas 2.3 and 2.4 for the diffusion model. First we need a simple lemma concerning Markov chains, whose proof can be left to the interested reader.

Let S be a separable metric space, and let $\{X_n, n = 0, 1, \dots\}$ denote the canonical coordinate process on $\Xi := S^{\mathbf{Z}_+}$, which has the product topology.

Lemma 3.1. *Let $(P_x)_{x\in S}$ and $(Q_x)_{x\in S}$ be (time-homogeneous) Markovian families of probability measures on $(\Xi, \mathscr{B}(\Xi))$, and suppose there exists a Borel function $V : S \times S \mapsto [0, \infty)$ satisfying*

$$\mathbf{E}^{Q_x}[f(X_1)] = \mathbf{E}^{P_x}[f(X_1)V(X_0, X_1)] \tag{3.7}$$

for all $f \in B(S)$ and $x \in S$. If we define $R_0 \equiv 1$ and

$$R_n = \prod_{i=1}^{n} V(X_{i-1}, X_i), \qquad n \geq 1, \tag{3.8}$$

then

$$\mathbf{E}^{Q_x}[f(X_0, X_1, \ldots, X_n)] = \mathbf{E}^{P_x}[f(X_0, X_1, \ldots, X_n)R_n] \tag{3.9}$$

for all $f \in B(S^{n+1})$ *and* $x \in S$. *In particular*, $\{R_n, n = 0, 1, \ldots\}$ *is a mean-one* $\{\mathscr{F}_n^X\}$*-martingale on* $(\Xi, \mathscr{B}(\Xi), P_x)$ *for each* $x \in S$, *and* $Q_x|\mathscr{F}_n^X \ll P_x|\mathscr{F}_n^X$ *with Radon–Nikodym derivative* R_n *for each* $n \geq 0$ *and* $x \in S$.

Let $\Xi_M := \mathscr{P}_M(E)^{\mathbf{Z}_+}$ have the product topology, let $\mathscr{F}$ be the Borel σ-field, let $\{\mu_n, n = 0, 1, \ldots\}$ be the canonical coordinate process, and let $\{\mathscr{F}_n\}$ be the corresponding filtration. For $\mu \in \mathscr{P}_M(E)$ we denote by $P_\mu^{(M)} \in \mathscr{P}(\Xi_M)$ the distribution of the neutral Wright–Fisher model starting at μ.

Lemma 3.2. *For each* $\mu \in \mathscr{P}_M(E)$, $T > 0$, $\rho \in (0, \rho_0)$, *and* $\lambda > \langle e^{\rho h_0}, \mu\rangle + \frac{1}{2}\theta T\langle e^{\rho h_0}, \nu_0\rangle$,

$$P_\mu^{(M)}\left\{\max_{0 \leq n \leq [MT]} \langle e^{\rho h_0}, \mu_n\rangle > \lambda\right\} \leq \frac{(1 + \frac{1}{2}\theta T)(\langle e^{\rho h_0}, \mu\rangle \vee \langle e^{\rho h_0}, \nu_0\rangle)}{\lambda - \langle e^{\rho h_0}, \mu\rangle - \frac{1}{2}\theta T\langle e^{\rho h_0}, \nu_0\rangle}. \tag{3.10}$$

Remark. The analogue of (2.5), namely

$$\mathbf{E}^{P_\mu^{(M)}}\left[\max_{0 \leq n \leq [MT]} \langle e^{\rho h_0}, \mu_n\rangle^2\right] \leq (12T + 3)\langle e^{2\rho h_0}, \mu\rangle + \left(12T + \frac{3}{4}\theta^2 T^2\right)\langle e^{2\rho h_0}, \nu_0\rangle, \tag{3.11}$$

also holds.

Proof. Fix $\mu \in \mathscr{P}_M(E)$ and $g \in B(E)$. Note first that

$$\begin{aligned} \mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_1\rangle] - \langle g, \mu\rangle &= \mathbf{E}\left[\left\langle g, \frac{1}{M}\sum_{i=1}^M \delta_{Y_i}\right\rangle\right] - \langle g, \mu\rangle \\ &= \langle g, (1-u)\mu + u\nu_0\rangle - \langle g, \mu\rangle \\ &= \frac{\theta}{2M}(\langle g, \nu_0\rangle - \langle g, \mu\rangle), \end{aligned} \tag{3.12}$$

where $Y_1, \ldots, Y_M$ are i.i.d. $(1-u)\mu + u\nu_0$; expectations without superscripts refer to unspecified probability spaces. Also,

$$\begin{aligned} \mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_k\rangle] &= \mathbf{E}^{P_\mu^{(M)}}[\mathbf{E}^{P_{\mu_{k-1}}^{(M)}}[\langle g, \mu_1\rangle]] \\ &= (1-u)\mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_{k-1}\rangle] + u\langle g, \nu_0\rangle \\ &= (1-u)^k\langle g, \mu\rangle + [1 - (1-u)^k]\langle g, \nu_0\rangle \\ &\leq \langle g, \mu\rangle \vee \langle g, \nu_0\rangle \end{aligned} \tag{3.13}$$

for all $k \geq 1$.

By the Markov property and (3.12),

$$Z_n^g := \langle g, \mu_n \rangle - \langle g, \mu_0 \rangle - \frac{\theta}{2M} \sum_{k=0}^{n-1} (\langle g, \nu_0 \rangle - \langle g, \mu_k \rangle) \tag{3.14}$$

is an $\{\mathscr{F}_n\}$-martingale on $(\Xi_M, \mathscr{F}, P_\mu^{(M)})$. Assume that g is also nonnegative; then $\langle g, \mu_n \rangle \le Z_n^g + \langle g, \mu_0 \rangle + (2M)^{-1}\theta n \langle g, \nu_0 \rangle$ for all $n \ge 0$. Consequently, given $T > 0$ and $\lambda > \langle g, \mu \rangle + \frac{1}{2}\theta T \langle g, \nu_0 \rangle$,

$$\begin{aligned} &P_\mu^{(M)}\left\{ \max_{0 \le n \le [MT]} \langle g, \mu_n \rangle > \lambda \right\} \\ &\le P_\mu^{(M)}\left\{ \max_{0 \le n \le [MT]} Z_n^g > \lambda - \langle g, \mu \rangle - \frac{1}{2}\theta T \langle g, \nu_0 \rangle \right\} \\ &\le \frac{\mathbf{E}^{P_\mu^{(M)}}[(Z_{[MT]}^g)^+]}{\lambda - \langle g, \mu \rangle - \frac{1}{2}\theta T \langle g, \nu_0 \rangle} \\ &\le \frac{\mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_{[MT]} \rangle + (2M)^{-1}\theta \sum_{k=0}^{[MT]-1} \langle g, \mu_k \rangle]}{\lambda - \langle g, \mu \rangle - \frac{1}{2}\theta T \langle g, \nu_0 \rangle} \\ &\le \frac{(1 + \frac{1}{2}\theta T)(\langle g, \mu \rangle \vee \langle g, \nu_0 \rangle)}{\lambda - \langle g, \mu \rangle - \frac{1}{2}\theta T \langle g, \nu_0 \rangle}, \end{aligned} \tag{3.15}$$

where the last inequality uses (3.13).

As in the proof of Lemma 2.2, we apply (3.15) with $g = e^{\rho h_0} \wedge K$, where $\rho \in (0, \rho_0)$, and (3.10) follows by letting $K \to \infty$.

We define the map $\Phi_M : \Xi_M \mapsto \Omega^\circ$ by

$$\Phi_M(\mu_0, \mu_1, \dots)_t = (1 - (Mt - [Mt]))\mu_{[Mt]} + (Mt - [Mt])\mu_{[Mt]+1}. \tag{3.16}$$

This transformation maps a discrete-time process to a continuous-time one with continuous piecewise-linear sample paths, rescaling time by a factor of M. For each $\mu \in \mathscr{P}_M(E)$, let $P_\mu^{(M)} \in \mathscr{P}(\Xi_M)$ denote the distribution of the neutral Wright–Fisher model starting at μ, and, for each $\mu \in \mathscr{P}^\circ(E)$, let $P_\mu \in \mathscr{P}(\Omega^\circ)$ denote the distribution of the neutral Fleming–Viot process starting at μ.

The next lemma shows that the neutral Wright–Fisher model, with time rescaled appropriately, converges in distribution in Ω° (not just Ω) to the neutral Fleming–Viot process.

Lemma 3.3. *Let $\{\mu^{(M)}\} \subset \mathscr{P}_M(E) \subset \mathscr{P}^\circ(E)$ and $\mu \in \mathscr{P}^\circ(E)$ satisfy $d^\circ(\mu^{(M)}, \mu) \to 0$. For simplicity of notation, denote $P_{\mu^{(M)}}^{(M)}$ by just $P^{(M)}$. Then $P^{(M)}\Phi_M^{-1} \Rightarrow P_\mu$ on Ω°.*

Proof. First, we verify the compact containment condition (Ethier and Kurtz (1986)) in Ω°. Let $\varepsilon > 0$ and $T > 0$ be given. It is well known that $P^{(M)}\Phi_M^{-1} \Rightarrow$

P_μ on Ω. In particular, $\{P^{(M)}\Phi_M^{-1}\}$ satisfies the compact containment condition in Ω, so there exists a compact set $K \subset \mathscr{P}(E)$ such that $P^{(M)}\Phi_M^{-1}\{\mu_t \in K$ for $0 \le t \le T\} \ge 1 - \frac{\varepsilon}{2}$ for all M. Fix a sequence $0 < r_1 < r_2 < \cdots$ with $r_k \to \rho_0$. For each positive integer k, define the constant

$$C_k = \sup_M \left\{ \langle e^{r_k h_0}, \mu^{(M)} \rangle + \frac{1}{2}\theta T \langle e^{r_k h_0}, \nu_0 \rangle + \varepsilon^{-1} 2^{k+1} \left(1 + \frac{1}{2}\theta T\right)(\langle e^{r_k h_0}, \mu^{(M)} \rangle \vee \langle e^{r_k h_0}, \nu_0 \rangle) \right\}. \tag{3.17}$$

Then

$$\hat{K} := K \cap \bigcap_{k=1}^{\infty} \{\mu \in \mathscr{P}(E) : \langle e^{r_k h_0}, \mu \rangle \le C_k\} \tag{3.18}$$

is compact in $\mathscr{P}^\circ(E)$, and

$$\begin{aligned} P^{(M)}\Phi_M^{-1}\{\mu_t \in \hat{K} \text{ for } 0 \le t \le T\} &\ge 1 - P^{(M)}\Phi_M^{-1}\{\mu_t \notin K \text{ for some } t \in [0, T]\} \\ &\quad - P^{(M)}\left(\bigcup_{k=1}^{\infty} \left\{ \max_{0 \le n \le [MT]} \langle e^{r_k h_0}, \mu_n \rangle > C_k \right\} \right) \\ &\ge 1 - \frac{\varepsilon}{2} - \sum_{k=1}^{\infty} \frac{(1 + \frac{1}{2}\theta T)(\langle e^{r_k h_0}, \mu^{(M)} \rangle \vee \langle e^{r_k h_0}, \nu_0 \rangle)}{C_k - \langle e^{r_k h_0}, \mu^{(M)} \rangle - \frac{1}{2}\theta T \langle e^{r_k h_0}, \nu_0 \rangle} \\ &\ge 1 - \varepsilon \end{aligned} \tag{3.19}$$

for all M.

For completeness, we prove here convergence of the generators, though the argument is essentially as in Ethier and Kurtz (1986), Section 10.4. For functions φ on $\mathscr{P}^\circ(E)$ of the form

$$\varphi(\mu) = \langle f_1, \mu \rangle \cdots \langle f_n, \mu \rangle, \tag{3.20}$$

where $n \ge 1$ and $f_1, \ldots, f_n \in \bar{C}(E)$, define $\mathscr{L}_0^{(M)}\varphi$ on $\mathscr{P}_M(E)$ by

$$(\mathscr{L}_0^{(M)}\varphi)(\mu) = M\{\mathbf{E}^{P_\mu^{(M)}}[\varphi(\mu_1)] - \varphi(\mu)\}. \tag{3.21}$$

Letting $\pi(n, k)$ denote the set of partitions β of $\{1, \ldots, n\}$ into k unordered subsets $\beta_1, \ldots, \beta_k$ (with $\min \beta_1 < \cdots < \min \beta_k$), and letting $Y_1, \ldots, Y_M$ be i.i.d. $\mu^{**} := (1 - u)\mu + u\nu_0$, we have

$$
\begin{aligned}
(3.22)\qquad \mathbf{E}^{P_\mu^{(M)}}[\varphi(\mu_1)] &= \mathbf{E}\left[\left\langle f_1, \frac{1}{M}\sum_{i=1}^M \delta_{Y_i}\right\rangle \cdots \left\langle f_n, \frac{1}{M}\sum_{i=1}^M \delta_{Y_i}\right\rangle\right] \\
&= \frac{1}{M^n}\mathbf{E}\left[\left(\sum_{i=1}^M f_1(Y_i)\right)\cdots\left(\sum_{i=1}^M f_n(Y_i)\right)\right] \\
&= \frac{1}{M^n}\sum_{k=1}^n \frac{M!}{(M-k)!}\sum_{\beta\in\pi(n,k)}\prod_{j=1}^k\left\langle \prod_{i\in\beta_j} f_i, \mu^{**}\right\rangle
\end{aligned}
$$

for all $\mu \in \mathscr{P}_M(E)$. Consequently,

$$
\begin{aligned}
(3.23)\qquad & (\mathscr{L}_0^{(M)}\varphi)(\mu) \\
&= M\left\{\frac{1}{M^n}\frac{M!}{(M-n)!}\prod_{j=1}^n \langle f_j, \mu^{**}\rangle \right. \\
&\quad + \frac{1}{M^n}\frac{M!}{(M-n+1)!}\sum_{1\le i<j\le n}\langle f_i f_j, \mu^{**}\rangle \prod_{l:l\ne i,j}\langle f_l, \mu^{**}\rangle \\
&\quad \left. + O(M^{-2}) - \prod_{j=1}^n \langle f_j, \mu\rangle\right\} \\
&= M\left\{\left(1 - \frac{\binom{n}{2}}{M}\right)\prod_{j=1}^n \langle f_j, \mu^{**}\rangle \right. \\
&\quad \left. + \frac{1}{M}\sum_{1\le i<j\le n}\langle f_i f_j, \mu^{**}\rangle \prod_{l:l\ne i,j}\langle f_l, \mu^{**}\rangle - \prod_{j=1}^n \langle f_j, \mu\rangle\right\} + O(M^{-1}) \\
&= \sum_{1\le i<j\le n}(\langle f_i f_j, \mu^{**}\rangle - \langle f_i, \mu^{**}\rangle\langle f_j, \mu^{**}\rangle)\prod_{l:l\ne i,j}\langle f_l, \mu^{**}\rangle \\
&\quad + \sum_{i=1}^n \langle Af_i, \mu\rangle \prod_{j:j<i}\langle f_j, \mu\rangle \prod_{j:j>i}\langle f_j, \mu^{**}\rangle + O(M^{-1}) \\
&= \sum_{1\le i<j\le n}(\langle f_i f_j, \mu\rangle - \langle f_i, \mu\rangle\langle f_j, \mu\rangle)\prod_{l:l\ne i,j}\langle f_l, \mu\rangle \\
&\quad + \sum_{i=1}^n \langle Af_i, \mu\rangle \prod_{j:j\ne i}\langle f_j, \mu\rangle + O(M^{-1}) \\
&= (\mathscr{L}_0\varphi)(\mu) + O(M^{-1}),
\end{aligned}
$$

uniformly in $\mu \in \mathscr{P}_M(E)$. Thus, the lemma follows from several results in Ethier and Kurtz (1986) (Theorems 3.9.1 and 3.9.4, Proposition 3.10.4, and Corollary 4.8.13).

For the next two lemmas we require the infinitely-many-alleles assumption that every mutant is of a type that has not previously appeared. Mathematically, this amounts to

$$\nu_0(\{x\}) = 0, \qquad x \in E. \tag{3.24}$$

This of course includes (1.2).

For each $\mu \in \mathscr{P}_M(E)$, we denote by $P_\mu^{(M)}$ and $Q_\mu^{(M)}$ in $\mathscr{P}(\Xi_M)$ the distributions of the neutral and selective Wright–Fisher models, respectively, starting at μ.

Lemma 3.4. *Assume* (3.24). *Then, for each* $\mu \in \mathscr{P}_M(E)$,

$$dQ_\mu^{(M)} = R_n^{(M)}\, dP_\mu^{(M)} \qquad on\ \mathscr{F}_n, \qquad n \geq 0, \tag{3.25}$$

where

$$R_n^{(M)} = \exp\left\{\sum_{k=1}^{n} \langle h 1_{\operatorname{supp}\mu_{k-1}}, \mu_k\rangle - \sum_{k=1}^{n} \langle 1_{\operatorname{supp}\mu_{k-1}}, \mu_k\rangle M \log\langle e^{h/M}, \mu_{k-1}\rangle\right\}. \tag{3.26}$$

Proof. Let $\varphi \in B(\mathscr{P}_M(E))$ and $\mu \in \mathscr{P}_M(E)$. Then

$$\begin{aligned}
\mathbf{E}^{Q_\mu^{(M)}}[\varphi(\mu_1)] &= \int_E \cdots \int_E \varphi\left(\frac{1}{M}\sum_{j=1}^{M} \delta_{y_j}\right) \mu^{**}(dy_1)\cdots\mu^{**}(dy_M) \\
&= \sum_{I \subset \{1,2,\ldots,M\}} (1-u)^{|I|} u^{M-|I|} \int_E \cdots \\
&\qquad \int_E \varphi\left(\frac{1}{M}\sum_{j=1}^{M} \delta_{y_j}\right) \prod_{i \in I} \mu^*(dy_i) \prod_{i \in I^c} \nu_0(dy_i) \\
&= \sum_{I \subset \{1,2,\ldots,M\}} (1-u)^{|I|} u^{M-|I|} \int_E \cdots \\
&\qquad \int_E \varphi\left(\frac{1}{M}\sum_{j=1}^{M} \delta_{y_j}\right) \frac{\prod_{i\in I} w(y_i)}{\langle w, \mu\rangle^{|I|}} \prod_{i \in I} \mu(dy_i) \prod_{i \in I^c} \nu_0(dy_i) \\
&= \int_E \cdots \int_E \varphi\left(\frac{1}{M}\sum_{j=1}^{M} \delta_{y_j}\right) \frac{\prod_{1\leq i\leq M:\, y_i \in \operatorname{supp}\mu} w(y_i)}{\langle w, \mu\rangle^{|\{1\leq i\leq M:\, y_i\in\operatorname{supp}\mu\}|}} \\
&\qquad \times \prod_{i=1}^{M} ((1-u)\mu + u\nu_0)(dy_i) \\
&= \mathbf{E}^{P_\mu^{(M)}}[\varphi(\mu_1) V^{(M)}(\mu_0, \mu_1)],
\end{aligned} \tag{3.27}$$

where, if $\mu_1 = M^{-1}\sum_{j=1}^{M} \delta_{y_j}$,

$$V^{(M)}(\mu_0, \mu_1) := \frac{\prod_{1 \le i \le M: y_i \in \operatorname{supp} \mu_0} w(y_i)}{\langle w, \mu_0 \rangle^{|\{1 \le i \le M: y_i \in \operatorname{supp} \mu_0\}|}} \tag{3.28}$$

$$= \frac{\exp\{\langle M(\log w) 1_{\operatorname{supp} \mu_0}, \mu_1 \rangle\}}{\langle w, \mu_0 \rangle^{M \langle 1_{\operatorname{supp} \mu_0}, \mu_1 \rangle}}$$

$$= \exp\{\langle h 1_{\operatorname{supp} \mu_0}, \mu_1 \rangle - \langle 1_{\operatorname{supp} \mu_0}, \mu_1 \rangle M \log \langle e^{h/M}, \mu_0 \rangle\}.$$

The next-to-last equality in (3.27) uses (3.24). The result now follows from Lemma 3.1.

We next show that the Girsanov-type formula for the Wright–Fisher model converges is some sense to the one for the Fleming–Viot process. First, we need a bit of notation. Define $\hat{R}_t^{(M)}$ on Ω° for all $t \ge 0$ so as to satisfy

$$\hat{R}_t^{(M)} \circ \Phi_M = R_{[Mt]}^{(M)} \qquad \text{on } \Xi_M, \qquad t \ge 0, \tag{3.29}$$

where $R_\mu^{(M)}$ is as in Lemma 3.4. Specifically, we take

$$\hat{R}_t^{(M)} = \exp\left\{ \sum_{k=1}^{[Mt]} \langle h 1_{\operatorname{supp} \mu_{(k-1)/M}}, \mu_{k/M} \rangle \right.$$
$$\left. - \sum_{k=1}^{[Mt]} \langle 1_{\operatorname{supp} \mu_{(k-1)/M}}, \mu_{k/M} \rangle M \log \langle e^{h/M}, \mu_{(k-1)/M} \rangle \right\}. \tag{3.30}$$

We also define R_t on Ω° for all $t \ge 0$ to be what we called $R_t^{0,h}$ in Section 2, namely,

(3.31)

$$R_t = \exp\left\{ \langle h, \mu_t \rangle - \langle h, \mu_0 \rangle - \int_0^t \left[\frac{1}{2}(\langle h^2, \mu_s \rangle - \langle h, \mu_s \rangle^2) + \frac{1}{2}\theta(\langle h, \nu_0 \rangle - \langle h, \mu_s \rangle) \right] ds \right\}.$$

Lemma 3.5. *Assume that h is continuous and* (3.24) *holds, let $T > 0$ be arbitrary, and let $P^{(M)}$ be as in Lemma* 3.3. *Then there exist Borel functions $F_M, G_M : \Omega^\circ \mapsto (0, \infty)$, a continuous function $F : \Omega^\circ \mapsto (0, \infty)$, and a positive constant G such that*

$$\hat{R}_T^{(M)} = F_M G_M, \qquad R_T = FG, \tag{3.32}$$

$F_M \to F$ uniformly on compact subsets of Ω°, and $G_M \to G$ in $P^{(M)}\Phi_M^{-1}$-probability.

Proof. Let

$$\log F_M = \sum_{k=1}^{[MT]} \langle h, \mu_{k/M} \rangle - \sum_{k=1}^{[MT]} M \log \langle e^{h/M}, \mu_{(k-1)/M} \rangle$$
$$+ \frac{1}{2}\theta \sum_{k=1}^{[MT]} \log \langle e^{h/M}, \mu_{(k-1)/M} \rangle, \tag{3.33}$$

(3.34) $$\log G_M = \sum_{k=1}^{[MT]} \left(M\langle 1_{(\operatorname{supp} \mu_{(k-1)/M})^c}, \mu_{k/M}\rangle - \frac{1}{2}\theta \right) \log\langle e^{h/M}, \mu_{(k-1)/M}\rangle$$
$$- \sum_{k=1}^{[MT]} \langle h 1_{(\operatorname{supp} \mu_{(k-1)/M})^c}, \mu_{k/M}\rangle,$$

(3.35) $$\log F = \langle h, \mu_T\rangle - \langle h, \mu_0\rangle - \int_0^T \frac{1}{2}(\langle h^2, \mu_t\rangle - \langle h, \mu_t\rangle^2)dt + \int_0^T \frac{1}{2}\theta\langle h, \mu_t\rangle dt,$$

and

(3.36) $$\log G = -\frac{1}{2}\theta T\langle h, \nu_0\rangle,$$

and note that (3.32) holds. Then, pathwise on Ω°,

(3.37) $$\log\langle e^{h/M}, \mu_{(k-1)/M}\rangle$$
$$= \log\left(1 + \frac{\langle h, \mu_{(k-1)/M}\rangle}{M} + \frac{\langle h^2, \mu_{(k-1)/M}\rangle}{2M^2} + O(M^{-3})\right)$$
$$= \frac{\langle h, \mu_{(k-1)/M}\rangle}{M} + \frac{\frac{1}{2}(\langle h^2, \mu_{(k-1)/M}\rangle - \langle h, \mu_{(k-1)/M}\rangle^2)}{M^2} + O(M^{-3}),$$

so

(3.38) $$\log F_M = \langle h, \mu_{[MT]/M}\rangle - \langle h, \mu_0\rangle - \frac{1}{M}\sum_{k=1}^{[MT]} \frac{1}{2}(\langle h^2, \mu_{(k-1)/M}\rangle - \langle h, \mu_{(k-1)/M}\rangle^2)$$
$$+ \frac{1}{M}\sum_{k=1}^{[MT]} \frac{1}{2}\theta\langle h, \mu_{(k-1)/M}\rangle + O(M^{-1})$$
$$= \log F + o(1).$$

To show that these results hold uniformly on compact subsets of Ω° requires a more careful analysis, which we illustrate with an example.

Consider the problem of showing that, for fixed $T > 0$,

(3.39) $$\frac{1}{M}\sum_{k=1}^{[MT]} \langle h, \mu_{(k-1)/M}\rangle \to \int_0^T \langle h, \mu_t\rangle dt$$

uniformly on compact subsets of Ω°. This requires several observations. First, note that, for each $\omega \in \Omega^\circ$, $t \mapsto \langle h, \omega_t\rangle$ is continuous since h is continuous and $|h| \le h_0$. (Recall the topology on $\mathscr{P}^\circ(E)$.) Second, we claim that, if $\{\omega^{(n)}\} \subset \Omega^\circ$, $\omega \in \Omega^\circ$, and $\omega^{(n)} \to \omega$, then $\langle h, \omega_t^{(n)}\rangle \to \langle h, \omega_t\rangle$ uniformly on compact t-intervals. Of course, $\omega^{(n)} \to \omega$ means that $d^\circ(\omega_t^{(n)}, \omega_t) \to 0$ uniformly on compact t-intervals, hence $d^\circ(\omega_{t_n}^{(n)}, \omega_t) \to 0$ whenever $t_n \to t$, hence $\langle h, \omega_{t_n}^{(n)}\rangle \to \langle h, \omega_t\rangle$ whenever $t_n \to t$, and this is equivalent to our assertion. Third, it follows that $\omega \mapsto \int_0^T \langle h, \omega_t \rangle dt$ is continuous on Ω°. This argument, incidentally, leads to the conclusion that F is continuous on Ω°. Finally, it therefore suffices to show that, if $\{\omega^{(K)}\} \subset \Omega^\circ$, $\omega \in \Omega^\circ$, and $\omega^{(K)} \to \omega$, then

$$\frac{1}{M} \sum_{k=1}^{[MT]} \langle h, \omega^{(K)}_{(k-1)/M} \rangle \to \int_0^T \langle h, \omega_t \rangle dt. \tag{3.40}$$

But by the second observation, $\langle h, \omega_t^{(K)} \rangle \to \langle h, \omega_t \rangle$ uniformly on compact t-intervals, and therefore, using the first observation, (3.40) follows. The rest of the proof that $F_M \to F$ uniformly on compact subsets of Ω° is handled in the same way.

Next, because of (3.24), the $P^{(M)}\Phi_M^{-1}$-distribution of the second sum in $\log G_M$ is the distribution of

$$\frac{1}{M} \sum_{k=1}^{[MT]} \sum_{l=1}^{X_k} h(\xi_{kl}), \tag{3.41}$$

where $X_1, X_2, \ldots$ are independent binomial$(M, \theta/(2M))$ random variables and ξ_{kl} $(k, l = 1, 2, \ldots)$ are i.i.d. ν_0 and independent of $X_1, X_2, \ldots$. This converges in L^2 to $\frac{1}{2}\theta T \langle h, \nu_0 \rangle$, since

$$\begin{aligned} \mathbf{E}\left[\left(\frac{1}{M}\sum_{k=1}^{[MT]}\sum_{l=1}^{X_k} h(\xi_{kl}) - \frac{1}{2}\theta\frac{[MT]}{M}\langle h, \nu_0\rangle\right)^2\right] \\ = \mathbf{E}\left[\left(\frac{1}{M}\sum_{k=1}^{[MT]}\sum_{l=1}^{X_k}\{h(\xi_{kl}) - \langle h, \nu_0\rangle\} + \frac{1}{M}\sum_{k=1}^{MT}\left(X_k - \frac{1}{2}\theta\right)\langle h, \nu_0\rangle\right)^2\right] \\ = \frac{1}{M^2}\sum_{k=1}^{[MT]}\mathbf{E}\left[\left(\sum_{l=1}^{X_k}\{h(\xi_{kl}) - \langle h, \nu_0\rangle\}\right)^2\right] + \frac{1}{M^2}\sum_{k=1}^{MT}\mathbf{Var}(X_k)\langle h, \nu_0\rangle^2 \\ = \frac{1}{M^2}\sum_{k=1}^{[MT]}\mathbf{E}[X_k](\langle h^2, \nu_0\rangle - \langle h, \nu_0\rangle^2) + \frac{1}{M^2}\sum_{k=1}^{MT}\mathbf{Var}(X_k)\langle h, \nu_0\rangle^2 \\ \le \frac{[MT]}{M^2}\frac{1}{2}\theta\langle h^2, \nu_0\rangle. \end{aligned} \tag{3.42}$$

Finally, using (3.24) once again, the $P^{(M)}\Phi_M^{-1}$-distribution of the first sum in $\log G_M$ has second moment

$$\sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}\left[\left(M\langle 1_{(\mathrm{supp}\,\mu_{k-1})^c}, \mu_k\rangle - \frac{1}{2}\theta\right)^2\right] \mathbf{E}^{P^{(M)}}[(\log\langle e^{h/M}, \mu_{k-1}\rangle)^2] \tag{3.43}$$

by virtue of the fact that $M\langle 1_{(\mathrm{supp}\,\mu_{k-1})^c}, \mu_k\rangle$ is independent of μ_{k-1} and distributed binomial$(M, \theta/(2M))$ under $P^{(M)}$. But (3.43) is bounded by

$$(3.44)\qquad \sum_{k=1}^{[MT]} \frac{1}{2}\theta \mathbf{E}^{P^{(M)}}[(\log\langle e^{h/M}, \mu_{k-1}\rangle)^2] \le \frac{1}{2}\theta \sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[(\log\langle e^{h_0/M}, \mu_{k-1}\rangle)^2]$$

$$\le \frac{1}{2}\theta \sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[\langle e^{h_0/M} - 1, \mu_{k-1}\rangle^2]$$

$$\le \frac{1}{2}\theta \frac{1}{M^2} \sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[\langle h_0 e^{h_0/M}, \mu_{k-1}\rangle^2]$$

$$= O(M^{-1}),$$

using (3.11). To see the first inequality in (3.44), note that

$$(3.45)\qquad \log\langle e^{-h_0/M}, \mu\rangle \le \log\langle e^{h/M}, \mu\rangle \le \log\langle e^{h_0/M}, \mu\rangle$$

and therefore

$$(3.46)\qquad |\log\langle e^{h/M}, \mu\rangle| \le \max\{\log\langle e^{h_0/M}, \mu\rangle, -\log\langle e^{-h_0/M}, \mu\rangle\} = \log\langle e^{h_0/M}, \mu\rangle,$$

where the last identity uses Jensen's inequality. This proves the lemma.

Our last lemma is a simple result about weak convergence.

Lemma 3.6. *Let S be a separable metric space, let $f_n, g_n : S \mapsto [0, \infty)$ $(n \ge 1)$ be Borel functions, let $f : S \mapsto [0, \infty)$ be continuous (but not necessarily bounded), let g be a positive constant, and let $H : S \mapsto \mathbf{R}$ be bounded and continuous. Assume that $f_n \to f$ uniformly on compact sets. Let P_n $(n \ge 1)$ and P be Borel probability measures on S such that $P_n \Rightarrow P$, $g_n \to g$ in P_n-probability, and $\int_S f_n g_n \, dP_n = \int_S fg \, dP = 1$ for all $n \ge 1$. Then $\int_S f_n g_n H \, dP_n \to \int_S fgH \, dP$.*

Proof. By Theorem 5.5 of Billingsley (1968), $P_n f_n^{-1} \Rightarrow Pf^{-1}$ and $P_n(f_n H)^{-1} \Rightarrow P(fH)^{-1}$. Since $P_n g_n^{-1} \Rightarrow \delta_g$, it follows that $P_n(f_n g_n)^{-1} \Rightarrow P(fg)^{-1}$ and $P_n(f_n g_n H)^{-1} \Rightarrow P(fgH)^{-1}$. By Theorem 5.4 of Billingsley, this together with the assumptions that $f_n g_n \ge 0$, $fg \ge 0$, and $\int_S f_n g_n \, dP_n = \int_S fg \, dP = 1$ for all $n \ge 1$ imply that $\{f_n g_n\}$ is $\{P_n\}$-uniformly integrable. Since H is bounded, $\{f_n g_n H\}$ is also $\{P_n\}$-uniformly integrable. This, together with $P_n(f_n g_n H)^{-1} \Rightarrow P(fgH)^{-1}$ proved just above, gives the desired conclusion.

For each $\mu \in \mathscr{P}_M(E)$, let $Q_\mu^{(M)} \in \mathscr{P}(\Xi_M)$ denote the distribution of the selective Wright–Fisher model starting at μ, and for each $\mu \in \mathscr{P}^\circ(E)$, let $Q_\mu \in \mathscr{P}(\Omega^\circ)$ denote the distribution of the selective Fleming–Viot process starting at μ.

We have now done almost all the work required to prove the main result of this section.

Theorem 3.7. *Assume that h is continuous. Let $\{\mu^{(M)}\} \subset \mathscr{P}_M(E) \subset \mathscr{P}^\circ(E)$ and $\mu \in \mathscr{P}^\circ(E)$ satisfy $d^\circ(\mu^{(M)}, \mu) \to 0$. For simplicity of notation, denote $Q_{\mu^{(M)}}^{(M)}$ by just $Q^{(M)}$. Then $Q^{(M)}\Phi_M^{-1} \Rightarrow Q_\mu$ on Ω°.*

Proof. First, we prove the theorem under the additional assumption (3.24). Let $T > 1$ be arbitrary. We apply Lemma 3.6 with $S = \Omega^\circ$, $(f_n, g_n, f, g) = (F_M, G_M, F, G)$ from Lemma 3.5, H an arbitrary bounded continuous $\mathscr{F}_{T-1}$-measurable function on Ω°, and $(P_n, P) = (P^{(M)}\Phi_M^{-1}, P_\mu)$ from Lemma 3.3. Lemma 3.5 gives the required convergence of $\{f_n\}$ and $\{g_n\}$ and the continuity of f. Lemma 3.3 gives $P_n \Rightarrow P$. The requirement that $\int_S f_n g_n \, dP_n = 1$ for all n follows from

$$\int_{\Omega^\circ} \hat{R}_T^{(M)} dP^{(M)}\Phi_M^{-1} = \int_{\Xi_M} \hat{R}_T^{(M)} \circ \Phi_M \, dP^{(M)} = \int_{\Xi_M} R_{[MT]}^{(M)} \, dP^{(M)} = 1, \tag{3.47}$$

which uses (3.29), and of course $\int_S fg \, dP = 1$ because $\int_{\Omega^\circ} R_T \, dP_\mu = 1$. Thus, Lemma 3.6 implies that

$$\int_{\Omega^\circ} H \, dQ^{(M)}\Phi_M^{-1} = \int_{\Omega^\circ} H\hat{R}_T^{(M)} \, dP^{(M)}\Phi_M^{-1} \to \int_{\Omega^\circ} HR_T \, dP_\mu = \int_{\Omega^\circ} H \, dQ_\mu. \tag{3.48}$$

(We assumed H to be $\mathscr{F}_{T-1}$-measurable so that it would be $\mathscr{F}_{[MT]/M}$-measurable for every M.) Since the collection of all such H (as T varies) is convergence determining, $Q^{(M)}\Phi_M^{-1} \Rightarrow Q_\mu$.

Finally, we need to remove assumption (3.24). Given arbitrary E, ν_0, and h (satisfying (1.3) of course) with h continuous, define

$$\tilde{E} = E \times [0, 1], \qquad \tilde{\nu}_0 = \nu_0 \times \lambda, \qquad \tilde{h}(x, v) \equiv h(x), \tag{3.49}$$

where λ is Lebesgue measure, and apply the theorem under (3.24), which we have just proved. The initial distributions $\mu^{(M)}$ and μ can be replaced by $\mu^{(M)} \times \delta_0$ and $\mu \times \delta_0$, and the distributions $Q^{(M)}$ and Q_μ as well as the mapping Φ_M will be distinguished from the original ones with tildes. Letting $\pi : \tilde{E} \mapsto E$ denote projection onto the first coordinate, the mapping $\Lambda : C_{\mathscr{P}^\circ(\tilde{E})}[0, \infty) \mapsto \Omega^\circ$ given by $\Lambda(\tilde{\omega}) = \{\tilde{\omega}_t \pi^{-1}, t \geq 0\}$ is continuous, and hence

$$Q^{(M)}\Phi_M^{-1} = \tilde{Q}^{(M)}\tilde{\Phi}_M^{-1}\Lambda^{-1} \Rightarrow \tilde{Q}_{\mu \times \delta_0}\Lambda^{-1} = Q_\mu, \tag{3.50}$$

as required.

4. Characterization of the stationary distribution

If h is bounded, then it is known that the Fleming–Viot process in $\mathscr{P}(E)$ with generator $\mathscr{L}_h$ has a unique stationary distribution $\Pi_h \in \mathscr{P}(\mathscr{P}(E))$, is strongly ergodic, and is reversible. In fact,

$$\Pi_0(\cdot) = \mathbf{P}\left\{\sum_{i=1}^{\infty} \rho_i \delta_{\xi_i} \in \cdot\right\}, \tag{4.1}$$

where $\xi_1, \xi_2, \ldots$ are i.i.d. ν_0 and $(\rho_1, \rho_2, \ldots)$ is Poisson–Dirichlet with parameter θ and independent of $\xi_1, \xi_2, \ldots$. Furthermore,

$$(4.2)\qquad \Pi_h(d\mu) = e^{2\langle h,\mu\rangle}\Pi_0(d\mu) \Big/ \int_{\mathscr{P}(E)} e^{2\langle h,\nu\rangle}\Pi_0(d\nu).$$

These results can be found in Ethier and Kurtz (1994, 1998).

The finiteness of the normalizing constant in (4.2) is precisely the condition needed in the work of Overbeck *et al.* (1995). Notice that

$$(4.3)\qquad \int_{\mathscr{P}(E)} e^{2\langle h,\nu\rangle}\Pi_0(d\nu) = \mathbf{E}\left[\exp\left\{2\sum_{i=1}^{\infty}\rho_i h(\xi_i)\right\}\right] = \mathbf{E}\left[\prod_{i=1}^{\infty}\langle e^{2\rho_i h}, \nu_0\rangle\right].$$

A sufficient condition for this to be finite is $\langle e^{2h_0}, \nu_0\rangle < \infty$. At least when $E = \mathbf{R}$, $h(x) \equiv x$, $h_0 = |h|$, and ν_0 is symmetric on $\mathbf{R}$ (as in Tachida (1996)), a necessary condition for the finiteness of (4.3) is $\langle e^{\rho h_0}, \nu_0\rangle < \infty$ for all $\rho < 2$.

In this section we impose a slightly stronger condition: E, ν_0, and h are arbitrary, subject to the condition that there exist a continuous function $h_0 : E \mapsto [0,\infty)$ and a constant $\rho_0 \in (2,\infty]$ such that (1.3) holds. In other words, we now require $\rho_0 > 2$.

Recalling the example in (1.4), we have seen that we can characterize the process in that case if $\sigma_0 < \sqrt{2}$, and, as we will show below, we can characterize the stationary distribution as well if $\sigma_0 < \sqrt{2}/2$. The construction of Overbeck *et al.* (1995) requires $\sigma_0 < \sqrt{2}/2$ or both $\sigma_0 = \sqrt{2}/2$ and $\theta > 1$. (To see this, use (4.3) and Watterson and Guess (1977), Eq. (3.2.10).)

The following lemma was proved by Ethier (1997) under (1.2) and extends (with essentially the same proof) to (1.3).

Lemma 4.1. *Assume* (1.3) *with* $\rho_0 > 2$. *Then* $\Pi_0(\mathscr{P}^\circ(E)) = 1$ *and* $e^{2\langle h_0,\cdot\rangle} \in L^1(\Pi_0)$. *In addition,* Π_h, *defined by* (4.2), *is such that* $\mathscr{L}_h$ *is a symmetric linear operator on* $L^2(\Pi_h)$.

However, it does not immediately follow that Π_h is a reversible stationary distribution for the Fleming–Viot process with generator $\mathscr{L}_h$. The theorems of Fukushima and Stroock (1986) and Echeverria (1982) do not apply, again because of the unboundedness of h.

We can now state the main result of this section.

Theorem 4.2. *Assume* (1.3) *with* $\rho_0 > 2$. *Then* Π_h, *defined by* (4.2), *is a reversible stationary distribution for the Fleming–Viot process with generator* $\mathscr{L}_h$, *and it is the unique stationary distribution for this process.*

Proof. Reversibility is equivalent to

$$(4.4)\qquad \int_{\mathscr{P}^\circ(E)} \varphi(\mu)\mathscr{T}_h(t)\psi(\mu)\Pi_h(d\mu) = \int_{\mathscr{P}^\circ(E)} \psi(\mu)\mathscr{T}_h(t)\varphi(\mu)\Pi_h(d\mu)$$

for all $\varphi, \psi \in B(\mathscr{P}^\circ(E))$ and $t \geq 0$, where $\{\mathscr{T}_h(t)\}$ is the semigroup corresponding to $\mathscr{L}_h$. Using Lemma 2.3 and the notation of Section 2, as well as (4.2), we see that (4.4) is equivalent to

$$\int_{\mathscr{P}^\circ(E)} \varphi(\mu)\mathbf{E}^{P_\mu}[\psi(\mu_t)R_t^{0,h}]e^{2\langle h,\mu\rangle}\Pi_0(d\mu) = \int_{\mathscr{P}^\circ(E)} \psi(\mu)\mathbf{E}^{P_\mu}[\varphi(\mu_t)R_t^{0,h}]e^{2\langle h,\mu\rangle}\Pi_0(d\mu) \tag{4.5}$$

for all $\varphi, \psi \in B(\mathscr{P}^\circ(E))$ and $t \geq 0$. But we can rewrite (4.5) as

$$\int_{\mathscr{P}^\circ(E)} \mathbf{E}^{P_\mu}\left[\varphi(\mu_0)\psi(\mu_t)\exp\left\{\langle h,\mu_t\rangle + \langle h,\mu_0\rangle - \int_0^t \gamma(\mu_s)ds\right\}\right]\Pi_0(d\mu) = \int_{\mathscr{P}^\circ(E)} \mathbf{E}^{P_\mu}\left[\psi(\mu_0)\varphi(\mu_t)\exp\left\{\langle h,\mu_t\rangle + \langle h,\mu_0\rangle - \int_0^t \gamma(\mu_s)ds\right\}\right]\Pi_0(d\mu), \tag{4.6}$$

where $\gamma(\mu) := \frac{1}{2}(\langle h^2,\mu\rangle - \langle h,\mu\rangle^2) + \frac{1}{2}\theta(\langle h,\nu_0\rangle - \langle h,\mu\rangle)$. Now the neutral model is known to be reversible (Ethier (1990), Shiga (1990)), so $\{\mu_s, 0 \leq s \leq t\}$ and $\{\mu_{t-s}, 0 \leq s \leq t\}$ are equal in distribution under $\int_{\mathscr{P}^\circ(E)} P_\mu(\cdot)\Pi_0(d\mu)$, implying that (4.6) holds, and therefore we have the reversibility (hence stationarity) of Π_h.

For the uniqueness of Π_h, we can apply essentially the argument used by Ethier and Kurtz (1998) in the case of bounded h. There is one additional step needed, so we provide the details.

Suppose the conclusion fails. Then by Lemma 5.3 of Ethier and Kurtz (1998) there exist mutually singular stationary distributions $\Pi_1, \Pi_2 \in \mathscr{P}(\mathscr{P}^\circ(E))$. We will show that this leads to a contradiction.

Let $\mathscr{P}(E \times E)$ have the topology of weak convergence, let $\tilde{\Omega} := C_{\mathscr{P}(E\times E)}[0,\infty)$ have the topology of uniform convergence on compact sets, let $\tilde{\mathscr{F}}$ be the Borel σ-field, let $\{\tilde{\mu}_t, t \geq 0\}$ be the canonical coordinate process, and let $\{\tilde{\mathscr{F}}_t\}$ be the corresponding filtration.

Define the operator $\tilde{A}$ on $B(E \times E)$ by

$$(\tilde{A}f)(x_1,x_2) = \frac{1}{2}\theta\int_E (f(y,y) - f(x_1,x_2))\nu_0(dy) \tag{4.7}$$

and the functions $\tilde{h}_1$ and $\tilde{h}_2$ on $E \times E$ by

$$\tilde{h}_i(x_1,x_2) = h(x_i). \tag{4.8}$$

Let $P \in \mathscr{P}(\tilde{\Omega})$ be (the distribution of) a neutral Fleming–Viot process with type space $E \times E$, mutation operator $\tilde{A}$, and initial distribution $\Gamma \in \mathscr{P}(\mathscr{P}(E \times E))$ given by

$$\Gamma(B) = \int_{\mathscr{P}^\circ(E)}\int_{\mathscr{P}^\circ(E)} 1_B(\mu_1 \times \mu_2)\Pi_1(d\mu_1)\Pi_2(d\mu_2). \tag{4.9}$$

With the projections $\pi_1, \pi_2 : E \times E \mapsto E$ defined by $\pi_i(x_1,x_2) = x_i$, observe that, on $(\tilde{\Omega}, \tilde{\mathscr{F}}, P)$, $\{\tilde{\mu}_t\pi_1^{-1}, t \geq 0\}$ and $\{\tilde{\mu}_t\pi_2^{-1}, t \geq 0\}$ are Fleming–Viot processes with generator $\mathscr{L}_0$ and initial distributions Π_1 and Π_2, and that they couple, that is, there is a stopping time $\tau < \infty$ P-a.s. such that $\tilde{\mu}_t\pi_1^{-1} = \tilde{\mu}_t\pi_2^{-1}$ for all $t \geq \tau$ P-a.s.

Let us define

$$\mathscr{P}^\circ(E \times E) = \{\mu \in \mathscr{P}(E \times E) : \mu\pi_i^{-1} \in \mathscr{P}^\circ(E) \text{ for } i = 1, 2\} \tag{4.10}$$

and, for $\mu, \nu \in \mathscr{P}^\circ(E \times E)$,

$$\begin{aligned} \tilde{d}^\circ(\mu, \nu) \\ = \tilde{d}(\mu, \nu) + \sum_{i=1}^{2} \int_{(0,\rho_0)} \left(1 \wedge \sup_{0 \le \rho \le r} |\langle e^{\rho h_0}, \mu\pi_i^{-1} \rangle - \langle e^{\rho h_0}, \nu\pi_i^{-1} \rangle| \right) e^{-r}\, dr, \end{aligned} \tag{4.11}$$

where $\tilde{d}$ is a metric on $\mathscr{P}(E \times E)$ that induces the topology of weak convergence. Then $(\mathscr{P}^\circ(E \times E), \tilde{d}^\circ)$ is a complete separable metric space and $\tilde{d}^\circ(\mu_n, \mu) \to 0$ if and only if $\mu_n \Rightarrow \mu$ and $\sup_n \langle e^{\rho h_0}, \mu_n \pi_i^{-1} \rangle < \infty$ for $i = 1, 2$ and each $\rho \in (0, \rho_0)$. We now define

$$\tilde{\Omega}^\circ = C_{(\mathscr{P}^\circ(E \times E), \tilde{d}^\circ)}[0, \infty) \subset \tilde{\Omega} = C_{(\mathscr{P}(E \times E), \tilde{d})}[0, \infty). \tag{4.12}$$

Let $\tilde{\Omega}^\circ$ have the topology of uniform convergence on compact sets, let $\tilde{\mathscr{F}}^\circ$ be the Borel σ-field, let $\{\tilde{\mu}_t, t \ge 0\}$ be the canonical coordinate process on $\tilde{\Omega}^\circ$, and let $\{\tilde{\mathscr{F}}_t^\circ\}$ be the corresponding filtration.

Then, exactly as in Lemma 2.3,

$$\begin{aligned} \tilde{R}_t^{(i)} = \exp\biggl\{ \langle \tilde{h}_i, \tilde{\mu}_t \rangle - \langle \tilde{h}_i, \tilde{\mu}_0 \rangle - \int_0^t \biggl[\frac{1}{2}(\langle \tilde{h}_i^2, \tilde{\mu}_s \rangle - \langle \tilde{h}_i, \tilde{\mu}_s \rangle^2) \\ + \frac{1}{2}\theta(\langle h, \nu_0 \rangle - \langle \tilde{h}_i, \tilde{\mu}_s \rangle) \biggr] ds \biggr\} \end{aligned} \tag{4.13}$$

is a mean-one $\{\tilde{\mathscr{F}}_t^\circ\}$-martingale on $(\tilde{\Omega}^\circ, \tilde{\mathscr{F}}^\circ, P)$ for $i = 1, 2$. Thus, we can define Q_1 and Q_2 in $\mathscr{P}(\tilde{\Omega}^\circ)$ by

$$dQ_i = \tilde{R}_t^{(i)}\, dP \qquad \text{on } \tilde{\mathscr{F}}_t^\circ, \qquad t \ge 0, \qquad i = 1, 2, \tag{4.14}$$

and exactly as in Lemma 2.4 we conclude that, for $i = 1, 2$, Q_i is a solution of the $\tilde{\Omega}^\circ$ martingale problem for $\mathscr{L}_{\tilde{h}_i}$ with initial distribution Γ. It follows that the Q_i-distribution of $\{\tilde{\mu}_t \pi_i^{-1}, t \ge 0\}$ is a solution of the Ω° martingale problem for $\mathscr{L}_h$ with initial distribution Π_i, hence it is a stationary solution. Letting

$$\tau_N = \inf\{t \ge 0 : \langle h_0^2, \tilde{\mu}_t \pi_1^{-1} \rangle + \langle h_0^2, \tilde{\mu}_t \pi_2^{-1} \rangle \ge N\} \tag{4.15}$$

there is a constant $c_N(T) > 0$ such that

$$\tilde{R}_t^{(i)} \ge c_N(T), \qquad 0 \le t \le T \wedge \tau_N, \qquad i = 1, 2. \tag{4.16}$$

Consequently, for $i = 1, 2$,

$$\begin{aligned} \Pi_i(G) = Q_i\{\tilde{\mu}_T \pi_i^{-1} \in G\} &\ge c_N(T) P\{\tilde{\mu}_T \pi_i^{-1} \in G, \tau_N > T\} \\ &\ge c_N(T) P\{\tilde{\mu}_T \pi_i^{-1} \in G, \tau_N > T, \tau \le T\} \end{aligned} \tag{4.17}$$

for all Borel sets G. But the right side of (4.17) does not depend on i and is a nonzero measure in G if first T is chosen large enough and then N (depending on T) is chosen large enough. This contradicts the assumed mutual singularity of Π_1 and Π_2 and completes the proof.

Acknowledgment. SNE would like to thank the Department of Applied Physics of the Tokyo Institute of Technology for their hospitality during the preparation of this paper.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH

DEPARTMENT OF MATHEMATICS
TOKYO INSTITUTE OF TECHNOLOGY

References

Albeverio, S. and Röckner, M. (1995), Dirichlet form methods for uniqueness of martingale problems and applications, Proc. Sympos. Pure Math. **57**, 513–528.

Billingsley, P. (1968), Convergence of Probability Measures, Wiley, New York.

Dawson, D. A. (1978), Geostochastic calculus. Canadian J. Statist. **6**, 143–168.

Echeverria, P. E. (1982), A criterion for invariant measures of Markov processes, Z. Wahrsch. verw. Gebiete **61**, 1–16.

Ethier, S. N. (1990), The infinitely-many-neutral-alleles diffusion model with ages, Adv. Appl. Probab. **22**, 1–24.

Ethier, S. N. (1997), On the normal-selection model. In Progress in Population Genetics and Human Evolution, P. Donnelly and S. Tavaré, eds. IMA Volumes in Math. and its Appl. **87**, 309–320, Springer, New York.

Ethier, S. N. and Kurtz, T. G. (1986), Markov Processes: Characterization and Convergence, Wiley, New York.

Ethier, S. N. and Kurtz, T. G. (1987), The infinitely-many-alleles model with selection as a measure-valued diffusion. In Stochastic Models in Biology, M. Kimura, G. Kallianpur, and T. Hida, eds. Lecture Notes in Biomathematics **70**, 72–86, Springer-Verlag, Berlin.

Ethier, S. N. and Kurtz, T. G. (1993), Fleming–Viot processes in population genetics, SIAM J. Control Opt. **31**, 345–386.

Ethier, S. N. and Kurtz, T. G. (1994), Convergence to Fleming–Viot processes in the weak atomic topology, Stochastic Process. Appl. **54**, 1–27.

Ethier, S. N. and Kurtz, T. G. (1998), Coupling and ergodic theorems for Fleming–Viot processes, Ann. Probab. **26**, 533–561.

Fukushima, M. and Stroock, D. W. (1986), Reversibility of solutions to martingale problems, Adv. Math. Suppl. Stud. **9**, 107–123, Academic Press, Orlando.

Overbeck, L. (1995), Uniqueness of the Fleming–Viot process with selection. In Dirichlet Forms and Stochastic Processes, Z. M. Ma, M. Röckner, and J. A. Yan, eds. 329–334, de Gruyter, Berlin.

Overbeck, L., Röckner, M., and Schmuland, B. (1995), An analytic approach to Fleming–Viot processes with interactive selection, Ann. Probab. **23**, 1–36.

Shiga, T. (1990), A stochastic equation based on a Poisson system for a class of measure-valued diffusion processes, J. Math. Kyoto Univ. **30**, 245–279.

Stroock, D. W. and Varadhan, S. R. S. (1979), Multidimensional Diffusion Processes, Springer, Berlin.

Tachida, H. (1991), A study on a nearly neutral mutation model in finite populations, Genetics **128**, 183–192.

Tachida, H. (1996), Effects of the shape of the distribution of mutant effect in nearly neutral mutation models, J. Genetics **75**, 33–48.

Watterson, G. A. and Guess, H. A. (1977), Is the most frequent allele the oldest?, Theoret. Popn. Biol. **11**, 141–160.