数理解析研究所講究録
1089 巻 1999 年 126-143

A Fleming–Viot process with unbounded selection

志賀 徳造 (東京工業大学理工学研究科) (Tokuzo Shiga)

This is a joint work with S. N. Ethier (University of Utah).

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

$$(Af)(x) = \frac{\theta}{2}\int_E (f(y) - f(x))\,\nu_0(dy), \tag{1.1}$$

where $\theta > 0$ and $\nu_0 \in \mathcal{P}(E)$; 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 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 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) studied Fleming–Viot processes with unbounded selection intensity functions using Dirichlet forms, but were concerned mainly with support properties and did not address the issue of uniqueness of solutions of the martingale problem.

It involves almost no additional effort to weaken (1.2) as follows. Let E, ν_0, and h be arbitrary, subject to the condition that there exists a continuous function $h_0 : E \mapsto [0, \infty)$ such that

$$|h| \le h_0, \qquad \int_E e^{\rho h_0(x)}\,\nu_0(dx) < \infty, \quad \rho > 0. \tag{1.3}$$

The second requirement in (1.3) is simply that $\nu_0 h_0^{-1}$ have an everywhere-finite moment generating function. This assumption is in force throughout the paper. The generator of the Fleming–Viot process in question will be denoted by $\mathcal{L}_h$ 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 $\mathcal{P}(E)$ of the form

$$\varphi(\mu) = F(\langle f_1, \mu\rangle, \ldots, \langle f_k, \mu\rangle) = F(\langle \mathbf{f}, \mu\rangle), \tag{1.4}$$

where $k \ge 1$, $f_1, \ldots, f_k \in \overline{C}(E)$, $F \in C^2(\mathbf{R}^k)$, Note that because $\langle h, \mu\rangle$ appears in (1.5) and h may be unbounded, we need to restrict the state space to a suitable subset of $\mathcal{P}(E)$. Although other choices are possible, we take as our state space the set of Borel probability measures μ on E such that μh_0^{-1} has an everywhere-finite moment generating function.

Let us define

$$\mathcal{P}^\circ(E) = \{\mu \in \mathcal{P}(E) : \langle e^{\rho h_0}, \mu\rangle < \infty \text{ for all } \rho > 0\} \tag{1.6}$$

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

$$d^\circ(\mu,\nu) = d(\mu,\nu) + \int_0^\infty \Big(1 \wedge \sup_{0\le\rho\le r} |\langle e^{\rho h_0}, \mu\rangle - \langle e^{\rho h_0}, \nu\rangle|\Big) e^{-r}\, dr \tag{1.7}$$

where d is a metric on $\mathcal{P}(E)$ that induces the topology of weak convergence. Then $(\mathcal{P}^\circ(E), d^\circ)$ is a complete separable metric space and $d^\circ(\mu_n, \mu) \to 0$ if and only if $\mu_n \Rightarrow \mu$ and $e^{\rho h_0}$ is $\{\mu_n\}$-uniformly integrable for each $\rho > 0$. Thus, the topology on $\mathcal{P}^\circ(E)$ may be slightly stronger than the topology of weak convergence.

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

We will need four lemmas from Ethier (1997). All were proved under assumptions (1.2) but extend easily to (1.3).

LEMMA 2.1. Let h_1 and h_2 be bounded Borel functions on E. If $P \in \mathcal{P}(\Omega)$ is a solution of the martingale problem for $\mathcal{L}_{h_1}$, then

$$\begin{aligned} R_t &= \exp\left\{\langle h_2, \mu_t\rangle - \langle h_2, \mu_0\rangle - \int_0^t e^{-\langle h_2, \mu_s\rangle} \mathcal{L}_{h_1} e^{\langle h_2, \mu_s\rangle}\, ds\right\} \\ &= \exp\left\{\langle h_2, \mu_t\rangle - \langle h_2, \mu_0\rangle - \int_0^t \Big[\tfrac{1}{2}(\langle h_2^2, \mu_s\rangle - \langle h_2, \mu_s\rangle^2)\right. \\ &\qquad \left. + \tfrac{1}{2}\theta(\langle h_2, \nu_0\rangle - \langle h_2, \mu_s\rangle) + \langle h_1 h_2, \mu_s\rangle - \langle h_1, \mu_s\rangle\langle h_2, \mu_s\rangle\Big]\, ds\right\} \end{aligned} \tag{2.1}$$

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

$$dQ = R_t\, dP \quad \text{on} \quad \mathcal{F}_t, \qquad t \ge 0, \tag{2.2}$$

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

We now define

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

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

LEMMA 2.2. For each $\mu \in \mathcal{P}^\circ(E)$, $T > 0$, and $\rho > 0$,

$$\mathbf{E}^{P_\mu}\Big[\sup_{0\le t\le T} \langle e^{\rho h_0}, \mu_t\rangle^2\Big] \le (12T+3)\langle e^{2\rho h_0}, \mu\rangle + (12T + \tfrac{3}{4}\theta^2 T^2)\langle e^{2\rho h_0}, \nu_0\rangle. \tag{2.4}$$

In particular, recalling (1.3), $e^{\rho h_0}$ is $\{\mu_t,\ 0 \le t \le T\}$-uniformly integrable P_μ-a.s. for each $\rho > 0$ and $T > 0$, and therefore $P_\mu(\Omega^\circ) = 1$.

Remark. The corresponding result in Ethier (1997) contains a small error, so here we provide a corrected proof.

Proof. Fix $\mu \in \mathcal{P}^\circ(E)$, $T > 0$, and $\rho > 0$. For each $g \in \overline{C}(E)$,

$$Z^g(t) \equiv \langle g, \mu_t\rangle - \langle g, \mu_0\rangle - \tfrac{1}{2}\theta \int_0^t (\langle g, \nu_0\rangle - \langle g, \mu_s\rangle)\, ds \tag{2.5}$$

is a continuous $\{\mathcal{F}_t\}$-martingale on $(\Omega, \mathcal{F}, P_\mu)$ with quadratic variation process

$$\langle Z^g\rangle_t = \int_0^t (\langle g^2, \mu_s\rangle - \langle g, \mu_s\rangle^2)\, ds. \tag{2.6}$$

If, in addition, g is nonnegative, then $\langle g, \mu_t\rangle \le Z^g(t) + \langle g, \mu_0\rangle + \frac{1}{2}\theta t\langle g, \nu_0\rangle$ for all $t \ge 0$, so

$$\mathbf{E}^{P_\mu}\Big[\sup_{0\le t\le T} \langle g, \mu_t\rangle^2\Big] \le 3\mathbf{E}^{P_\mu}\Big[\sup_{0\le t\le T} Z^g(t)^2\Big] + 3\langle g, \mu\rangle^2 + \tfrac{3}{4}\theta^2 T^2 \langle g, \nu_0\rangle^2, \tag{2.7}$$

and

$$\begin{aligned} \mathbf{E}^{P_\mu}\Big[\sup_{0\le t\le T} Z^g(t)^2\Big] &\le 4\mathbf{E}^{P_\mu}[Z^g(T)^2] \\ &= 4\int_0^T \mathbf{E}^{P_\mu}[\langle g^2, \mu_s\rangle - \langle g, \mu_s\rangle^2]\, ds \\ &\le 4\int_0^T \langle U(s) g^2, \mu\rangle\, ds \\ &\le 4T(\langle g^2, \mu\rangle + \langle g^2, \nu_0\rangle), \end{aligned} \tag{2.8}$$

where $\{U(t)\}$ is the semigroup on $\overline{C}(E)$ with generator $Af = \frac{1}{2}\theta(\langle f, \nu_0\rangle - f)$; it is given by

$$U(t)f = e^{-\theta t/2} f + (1 - e^{-\theta t/2})\langle f, \nu_0\rangle. \tag{2.9}$$

Now let $g = e^{\rho h_0} \wedge K$ in (2.7) and (2.8), and noting that

$$\sup_{0\le t\le T} \langle e^{\rho h_0}, \mu_t\rangle^2 = \lim_{K\to\infty} \sup_{0\le t\le T} \langle e^{\rho h_0} \wedge K, \mu_t\rangle^2, \tag{2.10}$$

we obtain (2.4).

Let Ω° have the topology of uniform convergence on compact sets, let $\mathcal{F}^\circ$ be the Borel σ-field, let $\{\mu_t,\ t \ge 0\}$ be the canonical coordinate process on Ω°, and let $\{\mathcal{F}_t^\circ\}$ be the corresponding filtration. We do not distinguish notationally between the canonical coordinate process on Ω and that on Ω°, between $P_\mu \in \mathcal{P}(\Omega)$ and its restriction to $\mathcal{F}^\circ$ (note that $\mathcal{F}^\circ \subset \mathcal{F}$), or between R_t of (1.9) 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.

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

This is proved using the estimate of Lemma 2.2. For each $\mu \in \mathcal{P}^\circ(E)$, Lemma 2.3 allows us to define $Q_\mu \in \mathcal{P}(\Omega^\circ)$ by

$$dQ_\mu = R_t^{0,h}\, dP_\mu \quad \text{on} \quad \mathcal{F}_t^\circ, \qquad t \geq 0. \tag{2.11}$$

Lemma 2.2 can again be used to show that Q_μ solves the Ω° martingale problem for $\mathcal{L}_h$ starting at μ.

LEMMA 2.4. Let $\mu \in \mathcal{P}^\circ(E)$. Then

$$\varphi(\mu_t) - \int_0^t (\mathcal{L}_h \varphi)(\mu_s)\, ds \tag{2.12}$$

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

These results from Ethier (1997) proved the existence of solutions of the Ω° martingale problem for $\mathcal{L}_h$. We now complete the characterization.

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

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

$$\tau_N = \inf\{t \geq 0 : \langle h_0^2, \mu_t \rangle \geq N\}, \tag{2.13}$$

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

$$dP_\mu^N = R_{t\wedge\tau_N}^{h,0}\, dQ_\mu \quad \text{on} \quad \mathcal{F}_{t\wedge\tau_N}^\circ, \qquad t \geq 0. \tag{2.14}$$

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

$$\varphi(\mu_{t\wedge\tau_N}) - \int_0^{t\wedge\tau_N} (\mathcal{L}_0 \varphi)(\mu_s)\, ds \tag{2.15}$$

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

$$P_\mu^\circ = P_\mu^N \quad \text{on} \quad \mathcal{F}_{\tau_N}^\circ, \qquad N \geq 1. \tag{2.16}$$

We claim that

$$\varphi(\mu_t) - \int_0^t (\mathcal{L}_0 \varphi)(\mu_r)\, dr \tag{2.17}$$

is an $\{\mathcal{F}_t^\circ\}$-martingale on $(\Omega^\circ, \mathcal{F}^\circ, P_\mu^\circ)$ for every $\varphi \in \mathcal{D}(\mathcal{L}_0)$. To see this, fix such a φ, let H be a bounded continuous function on $\mathcal{P}^\circ(E)^m$, where $m \geq 1$, and let $0 < s_1 < \cdots < s_m \leq s < t$. Then

$$\mathbf{E}^{P_\mu^\circ}\left[\left(\varphi(\mu_{t\wedge\tau_N}) - \varphi(\mu_{s\wedge\tau_N}) - \int_{s\wedge\tau_N}^{t\wedge\tau_N} (\mathcal{L}_0\varphi)(\mu_r)\,dr\right) H(\mu_{s_1\wedge\tau_N}, \ldots, \mu_{s_m\wedge\tau_N})\right] = 0 \tag{2.18}$$

for each $N \geq 1$, hence

$$\mathbf{E}^{P_\mu^\circ}\left[\left(\varphi(\mu_t) - \varphi(\mu_s) - \int_s^t (\mathcal{L}_0\varphi)(\mu_r)\,dr\right) H(\mu_{s_1}, \ldots, \mu_{s_m})\right] = 0. \tag{2.19}$$

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

Finally, from

$$dP_\mu = R_{t\wedge\tau_N}^{h,0}\,dQ_\mu \quad \text{on} \quad \mathcal{F}_{t\wedge\tau_N}^\circ, \qquad t \geq 0, \tag{2.20}$$

we obtain

$$dQ_\mu = R_{t\wedge\tau_N}^{0,h}\,dP_\mu \quad \text{on} \quad \mathcal{F}_{t\wedge\tau_N}^\circ, \qquad t \geq 0, \tag{2.21}$$

and in particular that for each $N \geq 1$, $\mathbf{E}^{Q_\mu}[\varphi(\mu_{t\wedge\tau_N})]$ is uniquely determined for every bounded continuous φ 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 $\mathcal{L}_h$ 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 type space or set of possible alleles.
- M (a positive integer) is the haploid population size, or $M = 2N$ is the number of gametes in a diploid population of size N.
- u (in [0,1]) is the mutation rate per generation per gene.
- ν_0 (in $\mathcal{P}(E)$) 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 gametes $(x_1, \ldots, x_M) \in E^M$ or, since the order of the gametes is unimportant, $M^{-1}\sum_{i=1}^M \delta_{x_i} \in \mathcal{P}(E)$. (Here $\delta_x \in \mathcal{P}(E)$ is the unit mass at $x \in E$.) Thus, the state space for the process is

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

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

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

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

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

(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 $\mathcal{P}^\circ(E)$ of 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 $\mathcal{L}_h$ as in (1.4)–(1.5). 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. These two results require two simple lemmas concerning Markov chains, whose proofs can be left to the interested reader.

LEMMA 3.1. Let $\{X_n,\ n = 0, 1, \ldots\}$ be a Markov chain in a separable metric space S with transition function $\pi(x, dy)$, and define the operator P on $B(S)$ by $(Pf)(x) = \int_S f(y)\,\pi(x, dy)$. Then, for each $f \in B(S)$,

$$M_n \equiv f(X_n) - f(X_0) - \sum_{k=0}^{n-1}(Pf - f)(X_k) \tag{3.7}$$

is a zero-mean $\{\mathcal{F}_n^X\}$-martingale, as is $M_n^2 - A_n$, where

$$A_n = \sum_{k=0}^{n-1}\{P(f^2) - (Pf)^2\}(X_k). \tag{3.8}$$

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

LEMMA 3.2. Let $(P_x)_{x\in S}$ and $(Q_x)_{x\in S}$ be (time-homogeneous) Markovian families of probability measures on $(\Xi, \mathcal{B}(\Xi))$, and suppose there exists a Borel function $V : S \times S \mapsto [0, \infty)$ satisfying

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

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

$$R_n = \prod_{i=1}^{n} V(X_{i-1}, X_i), \qquad n \geq 1, \tag{3.10}$$

then

$$\mathbf{E}^{Q_x}[f(X_0, X_1, \ldots, X_n)] = \mathbf{E}^{P_x}[f(X_0, X_1, \ldots, X_n)R_n] \tag{3.11}$$

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

We begin by applying Lemma 3.1 to the neutral Wright–Fisher model. As in Section 2 it will be convenient to use the canonical coordinate process

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

LEMMA 3.3. For each $\mu \in \mathcal{P}_M(E)$, $T > 0$, and $\rho > 0$,

$$\mathbf{E}^{P_\mu^{(M)}}\left[\max_{0\leq n\leq [MT]} \langle e^{pho}, \mu_n\rangle^2\right] \leq (12T + 3)\langle e^{2pho}, \mu\rangle + (12T + \tfrac{3}{4}\theta^2T^2)\langle e^{2pho}, \nu_0\rangle. \tag{3.12}$$

Remark. Note that the right side of (3.12) is identical to that of (2.4).

Proof. Let $g \in \overline{C}(E)$. Note first that, for each $\mu \in \mathcal{P}_M(E)$,

$$\begin{aligned} \mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_1\rangle] - \langle g, \mu\rangle &= \mathbf{E}\left[\left\langle g, \frac{1}{M}\sum_{i=1}^{M}\delta_{Y_i}\right\rangle\right] - \langle g, \mu\rangle \\ &= \langle g, (1-u)\mu + u\nu_0\rangle - \langle g, \mu\rangle = \frac{\theta}{2M}(\langle g, \nu_0\rangle - \langle g, \mu\rangle), \end{aligned} \tag{3.13}$$

where $Y_1, \ldots, Y_M$ are i.i.d. $(1-u)\mu + u\nu_0$,

$$\begin{aligned} \mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_1\rangle^2] - (\mathbf{E}^{P_\mu^{(M)}}[\langle g, \mu_1\rangle])^2 &= \mathbf{Var}\left(\left\langle g, \frac{1}{M}\sum_{i=1}^{M}\delta_{Y_i}\right\rangle\right) \\ &= \frac{1}{M^2}\sum_{i=1}^{M}\mathbf{Var}(g(Y_i)) \leq \frac{1}{M}\{(1-u)\langle g^2, \mu\rangle + u\langle g^2, \nu_0\rangle\}, \end{aligned} \tag{3.14}$$

and

$$\mathbf{E}^{P_\mu^{(M)}}[\langle g^2,\mu_k\rangle] = \mathbf{E}^{P_\mu^{(M)}}[\mathbf{E}^{P_{\mu_{k-1}}^{(M)}}[\langle g^2,\mu_1\rangle]] \\ = (1-u)\mathbf{E}^{P_\mu^{(M)}}[\langle g^2,\mu_{k-1}\rangle] + u\langle g^2,\nu_0\rangle \\ = (1-u)^k\langle g^2,\mu\rangle + [1-(1-u)^k]\langle g^2,\nu_0\rangle \tag{3.15}$$

for all $k \geq 1$.

By Lemma 3.1,

$$Z_n^g \equiv \langle g,\mu_n\rangle - \langle g,\mu_0\rangle - \frac{\theta}{2M}\sum_{k=0}^{n-1}(\langle g,\nu_0\rangle - \langle g,\mu_k\rangle) \tag{3.16}$$

is an $\{\mathcal{F}_n\}$-martingale on $(\Xi,\mathcal{F},P_\mu^{(M)})$ with

$$\mathbf{E}^{P_\mu^{(M)}}[(Z_n^g)^2] \leq \frac{1}{M}\sum_{k=0}^{n-1}\{(1-u)\mathbf{E}^{P_\mu^{(M)}}[\langle g^2,\mu_k\rangle] + u\langle g^2,\nu_0\rangle\} \\ = \frac{1}{M}\sum_{k=0}^{n-1}\{(1-u)^{k+1}\langle g^2,\mu\rangle + [1-(1-u)^{k+1}]\langle g^2,\nu_0\rangle\} \\ \leq \frac{n}{M}(\langle g^2,\mu\rangle + \langle g^2,\nu_0\rangle) \tag{3.17}$$

for all $n \geq 1$ and $\mu \in \mathcal{P}_M(E)$. If, in addition, g is nonnegative, then $\langle g,\mu_n\rangle \leq Z_n^g + \langle g,\mu_0\rangle + (2M)^{-1}\theta n\langle g,\nu_0\rangle$ for all $n \geq 0$, so, for each $\mu \in \mathcal{P}_M(E)$,

$$\mathbf{E}^{P_\mu^{(M)}}\Big[\max_{0\leq n\leq [MT]}\langle g,\mu_n\rangle^2\Big] \leq 3\mathbf{E}^{P_\mu^{(M)}}\Big[\max_{0\leq n\leq [MT]}(Z_n^g)^2\Big] + 3\langle g,\mu\rangle^2 + \tfrac{3}{4}\theta^2T^2\langle g,\nu_0\rangle^2, \tag{3.18}$$

for all $T > 0$, and

$$\mathbf{E}^{P_\mu^{(M)}}\Big[\max_{0\leq n\leq [MT]}(Z_n^g)^2\Big] \leq 4\mathbf{E}^{P_\mu^{(M)}}[(Z_{[MT]}^g)^2] \leq 4T(\langle g^2,\mu\rangle + \langle g^2,\nu_0\rangle). \tag{3.19}$$

As in the proof of Lemma 2.2, we apply (3.18) and (3.19) with $g = e^{\rho h_0} \wedge K$, and the result follows by letting $K \to \infty$.

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

$$\Phi_M(\mu_0,\mu_1,\ldots)_t = (1-(Mt-[Mt]))\mu_{[Mt]} + (Mt-[Mt])\mu_{[Mt]+1}. \tag{3.20}$$

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 $\mu \in \mathcal{P}_M(E)$, let $P_\mu^{(M)} \in \mathcal{P}(\Xi_M)$ denote the distribution of the neutral Wright–Fisher model starting at μ, and let $P_\mu \in \mathcal{P}(\Omega^\circ)$ 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 3.4. Let $\{\mu^{(M)}\} \subset \mathcal{P}_M(E) \subset \mathcal{P}^\circ(E)$ and $\mu \in \mathcal{P}^\circ(E)$ satisfy $d^\circ(\mu^{(M)}, \mu) \to 0$. For simplicity of notation, denote $P^{(M)}_{\mu^{(M)}}$ 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)). Let $\varepsilon > 0$ and $T > 0$ be given. For each positive integer r, define the constant

$$C_r = \varepsilon^{-1} 2^r \sup_M \{(12T+3)\langle e^{2rh_0}, \mu^{(M)}\rangle + (12T + \tfrac{3}{4}\theta^2 T^2)\langle e^{2rh_0}, \nu_0\rangle\}^{1/2}. \tag{3.21}$$

Then

$$K \equiv \bigcap_{r=1}^{\infty} \{\mu \in \mathcal{P}(E) : \langle e^{rh_0}, \mu\rangle \le C_r\} \tag{3.22}$$

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

$$\begin{aligned} &P^{(M)}\Phi_M^{-1}\{\mu_t \in K \text{ for } 0 \le t \le T\} \\ &= 1 - P^{(M)}\left(\bigcup_{r=1}^{\infty}\left\{\max_{0\le n\le [MT]} \langle e^{rh_0}, \mu_n\rangle > C_r\right\}\right) \\ &\ge 1 - \sum_{r=1}^{\infty} C_r^{-1}\mathbf{E}^{P^{(M)}}\left[\max_{0\le n\le [MT]}\langle e^{rh_0}, \mu_n\rangle\right] \\ &\ge 1 - \varepsilon \end{aligned} \tag{3.23}$$

for all M, where the last inequality uses Lemma 3.3 and (3.21).

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 $\mathcal{P}^\circ(E)$ of the form

$$\varphi(\mu) = \langle f_1, \mu\rangle \cdots \langle f_n, \mu\rangle, \tag{3.24}$$

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

$$(\mathcal{L}_0^{(M)}\varphi)(\mu) = M\{\mathbf{E}^{P_\mu^{(M)}}[\varphi(\mu_1)] - \varphi(\mu)\}. \tag{3.25}$$

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. $\mu^{**} \equiv (1-u)\mu + u\nu_0$, we have

$$\begin{aligned} \mathbf{E}^{P_\mu^{(M)}}[\varphi(\mu_1)] &= \mathbf{E}\left[\left\langle f_1, \frac{1}{M}\sum_{i=1}^{M}\delta_{Y_i}\right\rangle \cdots \left\langle f_n, \frac{1}{M}\sum_{i=1}^{M}\delta_{Y_i}\right\rangle\right] \\ &= \frac{1}{M^n}\mathbf{E}\left[\left(\sum_{i=1}^{M} f_1(Y_i)\right)\cdots\left(\sum_{i=1}^{M} f_n(Y_i)\right)\right] \\ &= \frac{1}{M^n}\sum_{k=1}^{n}\frac{M!}{(M-k)!}\sum_{\beta\in\pi(n,k)}\prod_{j=1}^{k}\left\langle \prod_{i\in\beta_j} f_i, \mu^{**}\right\rangle \end{aligned} \tag{3.26}$$

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

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

uniformly in $\mu \in \mathcal{P}_M(E)$. 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 is,

$$
\nu_0(\{x\}) = 0, \qquad x \in E. \tag{3.28}
$$

This of course includes (1.2).

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

LEMMA 3.5. Assume (3.28). Then, for each $\mu \in \mathcal{P}_M(E)$,

$$
dQ_\mu^{(M)} = R_n^{(M)}\,dP_\mu^{(M)} \quad \text{on} \quad \mathcal{F}_n^\circ, \qquad n \ge 0, \tag{3.29}
$$

where

$$
R_n^{(M)} = \exp\left\{\sum_{k=1}^n\langle h1_{\operatorname{supp}\mu_{k-1}},\mu_k\rangle - \sum_{k=1}^n\langle 1_{\operatorname{supp}\mu_{k-1}},\mu_k\rangle M\log\langle e^{h/M},\mu_{k-1}\rangle\right\}. \tag{3.30}
$$

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

$$
\mathbf{E}^{Q_\mu^{(M)}}[\varphi(\mu_1)] = \int_E\cdots\int_E\varphi\left(\frac{1}{M}\sum_{i=1}^M\delta_{y_i}\right)\mu^{**}(dy_1)\cdots\mu^{**}(dy_M) \tag{3.31}
$$

$$= \sum_{I\subset\{1,2,\ldots,M\}} (1-u)^{|I|}u^{M-|I|} \int_E \cdots \int_E \varphi\Big(\frac{1}{M}\sum_{i=1}^M \delta_{y_i}\Big) \prod_{i\in I} \mu^*(dy_i) \prod_{i\in I^c} \nu_0(dy_i)$$

$$= \sum_{I\subset\{1,2,\ldots,M\}} (1-u)^{|I|}u^{M-|I|} \int_E \cdots \int_E \varphi\Big(\frac{1}{M}\sum_{i=1}^M \delta_{y_i}\Big) \frac{\prod_{i\in I} w(y_i)}{\langle w,\mu\rangle^{|I|}} \prod_{i\in I} \mu(dy_i) \prod_{i\in I^c} \nu_0(dy_i)$$

$$= \mathbf{E}^{P_\mu^{(M)}}[\varphi(\mu_1)V^{(M)}(\mu_0,\mu_1)],$$

where, if $\mu_1 = M^{-1}\sum_{i=1}^M \delta_{y_i}$ and $I = \{1 \le i \le M : y_i \in \operatorname{supp}\mu_0\}$,

$$\begin{aligned} V^{(M)}(\mu_0,\mu_1) &= \frac{\prod_{i\in I} w(y_i)}{\langle w,\mu_0\rangle^{|I|}} \\ &= \frac{\exp\{\langle M(\log w)1_{\operatorname{supp}\mu_0},\mu_1\rangle\}}{\langle w,\mu_0\rangle^{M\langle 1_{\operatorname{supp}\mu_0},\mu_1\rangle}} \\ &= \exp\{\langle h1_{\operatorname{supp}\mu_0},\mu_1\rangle - \langle 1_{\operatorname{supp}\mu_0},\mu_1\rangle M\log\langle e^{h/M},\mu_0\rangle\}. \end{aligned} \tag{3.32}$$

The first equality in (3.32) uses (3.28). The result now follows from Lemma 3.2.

We next show that the Girsanov-type formula for the Wright–Fisher model converges in 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

$$\hat{R}_t^{(M)} \circ \Phi_M = R_{[Mt]}^{(M)} \quad \text{on} \quad \Xi_M, \qquad t \ge 0, \tag{3.33}$$

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

$$\begin{aligned} \hat{R}_t^{(M)} = \exp\Bigg\{ &\sum_{k=1}^{[Mt]} \langle h1_{\operatorname{supp}\mu_{(k-1)/M}},\mu_{k/M}\rangle \\ &- \sum_{k=1}^{[Mt]} \langle 1_{\operatorname{supp}\mu_{(k-1)/M}},\mu_{k/M}\rangle M\log\langle e^{h/M},\mu_{(k-1)/M}\rangle\Bigg\}. \end{aligned} \tag{3.34}$$

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,

$$\begin{aligned} R_t = \exp\Bigg\{\langle h,\mu_t\rangle - \langle h,\mu_0\rangle - \int_0^t [\tfrac{1}{2}(\langle h^2,\mu_s\rangle - \langle h,\mu_s\rangle^2) \\ + \tfrac{1}{2}\theta(\langle h,\nu_0\rangle - \langle h,\mu_s\rangle)]\,ds\Bigg\}. \end{aligned} \tag{3.35}$$

LEMMA 3.6. Assume that h is continuous and (3.28) holds, let $T > 0$ be arbitrary, and let $P^{(M)}$ be as in Lemma 3.4. 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

$$\hat{R}_T^{(M)} = F_M G_M, \qquad R_T = FG, \tag{3.36}$$

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

Proof. Let

$$\log F_M = \sum_{k=1}^{[MT]} \langle h, \mu_{(k-1)/M} \rangle - \sum_{k=1}^{[MT]} M \log\langle e^{h/M}, \mu_{(k-1)/M} \rangle + \tfrac{1}{2}\theta \sum_{k=1}^{[MT]} \log\langle e^{h/M}, \mu_{(k-1)/M} \rangle, \tag{3.37}$$

$$\log G_M = -\sum_{k=1}^{[MT]} (M\langle 1_{(\operatorname{supp} \mu_{(k-1)/M})^c}, \mu_{k/M} \rangle - \tfrac{1}{2}\theta) \log\langle e^{h/M}, \mu_{(k-1)/M} \rangle - \sum_{k=1}^{[MT]} \langle h 1_{(\operatorname{supp} \mu_{(k-1)/M})^c}, \mu_{k/M} \rangle, \tag{3.38}$$

$$\log F = \langle h, \mu_T \rangle - \langle h, \mu_0 \rangle - \int_0^T \tfrac{1}{2}(\langle h^2, \mu_t \rangle - \langle h, \mu_t \rangle^2)\, dt + \int_0^T \tfrac{1}{2}\theta \langle h, \mu_t \rangle\, dt, \tag{3.39}$$

and

$$\log G = -\tfrac{1}{2}\theta T \langle h, \nu_0 \rangle, \tag{3.40}$$

and note that (3.36) holds. Then, pathwise on Ω°,

$$\begin{aligned} \log\langle e^{h/M}, \mu_{(k-1)/M} \rangle &= \log\left(1 + \frac{\langle h, \mu_{(k-1)/M} \rangle}{M} + \frac{\langle h^2, \mu_{(k-1)/M} \rangle}{2M^2} + O(M^{-3})\right) \\ &= \frac{\langle h, \mu_{(k-1)/M} \rangle}{M} + \frac{\tfrac{1}{2}(\langle h^2, \mu_{(k-1)/M} \rangle - \langle h, \mu_{(k-1)/M} \rangle^2)}{M^2} + O(M^{-3}), \end{aligned} \tag{3.41}$$

so

$$\begin{aligned} \log F_M &= \langle h, \mu_{[MT]/M} \rangle - \langle h, \mu_0 \rangle - \frac{1}{M} \sum_{k=1}^{[MT]} \tfrac{1}{2}(\langle h^2, \mu_{(k-1)/M} \rangle - \langle h, \mu_{(k-1)/M} \rangle^2) \\ &\quad + \frac{1}{M} \sum_{k=1}^{[MT]} \tfrac{1}{2}\theta \langle h, \mu_{(k-1)/M} \rangle + O(M^{-1}) \\ &= \log F + o(1). \end{aligned} \tag{3.42}$$

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 $T > 0$,

$$\frac{1}{M} \sum_{k=1}^{[MT]} \langle h, \mu_{(k-1)/M} \rangle \to \int_0^T \langle h, \mu_t \rangle\, dt \tag{3.43}$$

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 $\mathcal{P}^\circ(E)$, in which convergence entails a uniform integrability condition.) 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

$$\frac{1}{M} \sum_{k=1}^{[MT]} \langle h, \omega^{(K)}_{(k-1)/M} \rangle \to \int_0^T \langle h, \omega_t \rangle\, dt. \tag{3.44}$$

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, (3.44) 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 (3.28), the $P^{(M)}\Phi_M^{-1}$-distribution of the second sum in $\log G_M$ is the distribution of

$$\frac{1}{M} \sum_{k=1}^{[MT]} \sum_{l=1}^{X_k} h(\xi_{kl}), \tag{3.45}$$

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

$$\begin{aligned}
&\mathbf{E}\left[\left(\frac{1}{M} \sum_{k=1}^{[MT]} \sum_{l=1}^{X_k} h(\xi_{kl}) - \tfrac{1}{2}\theta \frac{[MT]}{M} \langle h, \nu_0 \rangle\right)^2\right] \\
&= \mathbf{E}\left[\left(\frac{1}{M} \sum_{k=1}^{[MT]} \sum_{l=1}^{X_k} \{h(\xi_{kl}) - \langle h, \nu_0 \rangle\} + \frac{1}{M} \sum_{k=1}^{[MT]} (X_k - \tfrac{1}{2}\theta) \langle h, \nu_0 \rangle\right)^2\right] \\
&= \frac{1}{M^2} \mathbf{E}\left[\sum_{k=1}^{[MT]} \left(\sum_{l=1}^{X_k} \{h(\xi_{kl}) - \langle h, \nu_0 \rangle\}\right)^2\right] + \frac{1}{M^2} \sum_{k=1}^{[MT]} \mathbf{Var}(X_k) \langle h, \nu_0 \rangle^2 \\
&= \frac{1}{M^2} \sum_{k=1}^{[MT]} \mathbf{E}[X_k] (\langle h^2, \nu_0 \rangle - \langle h, \nu_0 \rangle^2) + \frac{1}{M^2} \sum_{k=1}^{[MT]} \mathbf{Var}(X_k) \langle h, \nu_0 \rangle^2 \\
&\le \frac{[MT]}{M^2} \tfrac{1}{2}\theta \langle h^2, \nu_0 \rangle.
\end{aligned} \tag{3.46}$$

Finally, using (3.28) once again, the $P^{(M)}\Phi_M^{-1}$-distribution of the first sum in $\log G_M$ has second moment

$$\sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[(M \langle 1_{(\operatorname{supp} \mu_{k-1})^c}, \mu_k \rangle - \tfrac{1}{2}\theta)^2] \mathbf{E}^{P^{(M)}}[(\log \langle e^{h/M}, \mu_{k-1} \rangle)^2] \tag{3.47}$$

by virtue of the fact that $M\langle 1_{(\operatorname{supp}\mu_{k-1})^c}, \mu_k\rangle$ is independent of μ_{k-1} and distributed binomial$(M, \theta/(2M))$ under $P^{(M)}$. But (3.47) is bounded by

$$
\begin{aligned}
(3.48) \quad \sum_{k=1}^{[MT]} \tfrac{1}{2}\theta \mathbf{E}^{P^{(M)}}[(\log\langle e^{h/M}, \mu_{k-1}\rangle)^2] &\le \tfrac{1}{2}\theta \sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[(\log\langle e^{h_0/M}, \mu_{k-1}\rangle)^2] \\
&\le \tfrac{1}{2}\theta \sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[\langle e^{h_0/M} - 1, \mu_{k-1}\rangle^2] \\
&\le \tfrac{1}{2}\theta \frac{1}{M^2} \sum_{k=1}^{[MT]} \mathbf{E}^{P^{(M)}}[\langle h_0 e^{h_0/M}, \mu_{k-1}\rangle^2] \\
&= O(M^{-1}),
\end{aligned}
$$

using Lemma 3.3. To see the first inequality in (3.48), note that

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

and therefore

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

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

Our last lemma is a simple result about weak convergence.

LEMMA 3.7. 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 $\mu \in \mathcal{P}_M(E)$, let $Q_\mu^{(M)} \in \mathcal{P}(\Xi_M)$ denote the distribution of the selective Wright–Fisher model starting at μ, and for each $\mu \in \mathcal{P}^\circ(E)$, let $Q_\mu \in \mathcal{P}(\Omega^\circ)$ 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 3.8. Assume that h is continuous. Let $\{\mu^{(M)}\} \subset \mathcal{P}_M(E) \subset \mathcal{P}^\circ(E)$ and $\mu \in \mathcal{P}^\circ(E)$ 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 (3.28). Let $T > 1$ be arbitrary. We apply Lemma 3.7 with $S = \Omega^\circ$, $(f_n, g_n, f, g) = (F_M, G_M, F, G)$

from Lemma 3.6, H an arbitrary bounded continuous $\mathcal{F}_{T-1}$-measurable function on Ω°, and $(P_n, P) = (P^{(M)}\Phi_M^{-1}, P_\mu)$ from Lemma 3.4. Lemma 3.6 gives the required convergence of $\{f_n\}$ and $\{g_n\}$ and the continuity of f. Lemma 3.4 gives $P_n \Rightarrow P$. The requirement that $\int_S f_n g_n \, dP_n = 1$ for all n follows from

$$(3.51) \qquad \int_{\Omega^\circ} \hat{R}_T^{(M)} \, dP^{(M)}\Phi_M^{-1} = \int_{\Xi_M} \hat{R}_T^{(M)} \circ \Phi_M \, dP^{(M)} = \int_{\Xi_M} R_{[MT]}^{(M)} \, dP^{(M)} = 1,$$

which uses (3.33), and of course $\int_S fg \, dP = 1$ because $\int_{\Omega^\circ} R_T \, dP_\mu = 1$. Thus, Lemma 3.8 implies that

$$(3.52) \qquad \int_{\Omega^\circ} H \, dQ^{(M)}\Phi_M^{-1} = \int_{\Omega^\circ} H\hat{R}_T^{(M)} \, dP^{(M)}\Phi_M^{-1} \to \int_{\Omega^\circ} HR_T \, dP_\mu = \int_{\Omega^\circ} H \, dQ_\mu.$$

(We assumed H to be $\mathcal{F}_{T-1}$-measurable so that it would be $\mathcal{F}_{\lceil MT \rceil/M}$-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 (3.28). Given arbitrary E, ν_0, and h (satisfying (1.3) of course), define

$$(3.53) \qquad \tilde{E} = E \times [0,1], \qquad \tilde{\nu}_0 = \nu_0 \times \lambda, \qquad \tilde{h}(x,v) \equiv h(x),$$

where λ is Lebesgue measure, and apply the theorem under (3.28), 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 $\Lambda : C_{\mathcal{P}^\circ(\tilde{E})}[0,\infty) \mapsto \Omega^\circ$ given by $\Lambda(\tilde{\omega}) = \{\tilde{\omega}_t \pi^{-1}, \ t \geq 0\}$ is continuous, and hence

$$(3.54) \qquad Q^{(M)}\Phi_M^{-1} = \tilde{Q}^{(M)}\tilde{\Phi}_M^{-1}\Lambda^{-1} \Rightarrow \tilde{Q}_{\mu\times\delta_0}\Lambda^{-1} = Q_\mu,$$

as required.

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

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

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,

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

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

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

LEMMA 4.1. Assume (1.3). Then $\Pi_0(\mathcal{P}^\circ(E)) = 1$ and $e^{2\langle h_0,\cdot\rangle} \in L^1(\Pi_0)$. In addition, Π_h, defined by (4.2), is such that $\mathcal{L}_h$ 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 $\mathcal{L}_h$. 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). Then Π_h, defined by (4.2), is a reversible stationary distribution for the Fleming–Viot process with generator $\mathcal{L}_h$, and it is the unique stationary distribution for this process.

Proof. The reversibility is known if h is bounded, so let $h_K = (-K) \vee (h \wedge K)$. Then

$$\int_{\mathcal{P}(E)} \varphi(\mu)\mathcal{T}_{h_K}(t)\psi(\mu)\,\Pi_{h_K}(d\mu) = \int_{\mathcal{P}(E)} \psi(\mu)\mathcal{T}_{h_K}(t)\varphi(\mu)\,\Pi_{h_K}(d\mu) \tag{4.3}$$

for all $\varphi, \psi \in C(\mathcal{P}(E))$ and $t \geq 0$, where $\{\mathcal{T}_{h_K}(t)\}$ is the semigroup corresponding to $\mathcal{L}_{h_K}$. Using Lemma 2.1 and the notation of Section 2, as well as (4.2), we see that (4.3) implies that

$$\int_{\mathcal{P}(E)} \varphi(\mu)\mathbf{E}^{P_\mu}[\psi(\mu_t)R_t^{0,h_K}]e^{2\langle h_K,\mu\rangle}\,\Pi_0(d\mu) \tag{4.4}$$
$$= \int_{\mathcal{P}(E)} \psi(\mu)\mathbf{E}^{P_\mu}[\varphi(\mu_t)R_t^{0,h_K}]e^{2\langle h_K,\mu\rangle}\,\Pi_0(d\mu)$$

for all $\varphi, \psi \in C(\mathcal{P}(E))$ and $t \geq 0$. Letting $K \to \infty$ and using Lemmas 2.2 and 4.1 to justify the interchanges of limits and integrals, we deduce 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 $\Pi_1, \Pi_2 \in \mathcal{P}(\mathcal{P}^\circ(E))$. We will show that this leads to a contradiction.

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

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

$$(\tilde{A}f)(x_1,x_2) = \tfrac{1}{2}\theta \int_E (f(y,y) - f(x_1,x_2)\,\nu_0(dy) \tag{4.5}$$

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

$$\tilde{h}_i(x_1,x_2) = h(x_i). \tag{4.6}$$

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

$$\int_{\mathcal{P}^\circ(E)} \int_{\mathcal{P}^\circ(E)} (\mu_1 \times \mu_2)(\cdot)\,\Pi_1(d\mu_1)\,\Pi_2(d\mu_2). \tag{4.7}$$

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 $(\tilde{\Omega}, \tilde{\mathcal{F}}, P)$, $\{\tilde{\mu}_t \pi_1^{-1},\ t \geq 0\}$ and $\{\tilde{\mu}_t \pi_2^{-1},\ t \geq 0\}$ are Fleming–Viot processes with generator $\mathcal{L}_0$ 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

$$(4.8) \qquad \mathcal{P}^\circ(E \times E) = \{\mu \in \mathcal{P}(E \times E) : \mu \pi_i^{-1} \in \mathcal{P}^\circ(E) \text{ for } i = 1, 2\}$$

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

$$(4.9) \qquad \tilde{d}^\circ(\mu, \nu) = \tilde{d}(\mu, \nu) + \sum_{i=1}^{2} \int_0^\infty \Big(1 \wedge \sup_{0 \leq \rho \leq r} |\langle e^{\rho h_0}, \mu \pi_i^{-1}\rangle - \langle e^{\rho h_0}, \nu \pi_i^{-1}\rangle|\Big) e^{-r}\, dr$$

where $\tilde{d}$ is a metric on $\mathcal{P}(E \times E)$ that induces the topology of weak convergence. Then $(\mathcal{P}^\circ(E \times E), \tilde{d}^\circ)$ is a complete separable metric space and $\tilde{d}^\circ(\mu_n, \mu) \to 0$ if and only if $\mu_n \Rightarrow \mu$ and $e^{\rho h_0}$ is $\{\mu_n \pi_1^{-1}\} \cup \{\mu_n \pi_2^{-1}\}$-uniformly integrable for each $\rho > 0$. We now define

$$(4.10) \qquad \tilde{\Omega}^\circ = C_{(\mathcal{P}^\circ(E\times E), \tilde{d}^\circ)}[0, \infty) \subset \tilde{\Omega} = C_{(\mathcal{P}(E\times E), \tilde{d})}[0, \infty).$$

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

Then, exactly as in Lemma 1.3,

$$(4.11) \qquad \tilde{R}_t^{(i)} = \exp\Big\{\langle \tilde{h}_i, \tilde{\mu}_t\rangle - \langle \tilde{h}_i, \tilde{\mu}_0\rangle - \int_0^t \Big[\tfrac{1}{2}(\langle \tilde{h}_i^2, \tilde{\mu}_s\rangle - \langle \tilde{h}_i, \tilde{\mu}_s\rangle^2) + \tfrac{1}{2}\theta(\langle h, \nu_0\rangle - \langle \tilde{h}_i, \tilde{\mu}_s\rangle)\Big]\, ds\Big\}$$

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

$$(4.12) \qquad dQ_i = \tilde{R}_t^{(i)}\, dP \quad \text{on} \quad \tilde{\mathcal{F}}_t^\circ, \qquad t \geq 0,\ i = 1, 2,$$

and exactly as in Lemma 1.4 we conclude that, for $i = 1, 2$, Q_i is a solution of the $\tilde{\Omega}^\circ$ martingale problem for $\mathcal{L}_{\tilde{h}_i}$ with initial distribution Π_i. Letting

$$(4.13) \qquad \tau_N = \inf\{t \geq 0 : \langle \tilde{h}_1^2, \tilde{\mu}_t\rangle + \langle \tilde{h}_2^2, \tilde{\mu}_t\rangle \geq N\}$$

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

$$(4.14) \qquad \tilde{R}_t^{(i)} \geq c_N(T), \qquad 0 \leq t \leq T \wedge \tau_N,\ i = 1, 2.$$

Consequently, for $i = 1, 2$,

$$(4.15) \qquad \Pi_i(G) = Q_i\{\tilde{\mu}_T \pi_i^{-1} \in G\} \geq c_N(T) P\{\tilde{\mu}_T \pi_i^{-1} \in G,\ \tau_N > T\} \geq c_N(T) P\{\tilde{\mu}_T \pi_i^{-1} \in G,\ \tau_N > T,\ \tau \leq T\}$$

for all Borel sets G. But the right side of (4.15) 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.

References

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

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

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. (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.*, to appear.

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

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.