ELSEVIER
Stochastic Processes and their Applications 54 (1994) 1–27
stochastic processes and their applications

Convergence to Fleming–Viot processes in the weak atomic topology

S.N. Ethier*,[a], Thomas G. Kurtz[b]

[a] *Department of Mathematics, University of Utah, Salt Lake City, UT 84112, USA*
[b] *Department of Mathematics, University of Wisconsin–Madison, Madison, WI 53706, USA*

Received 25 March 1991; revised 4 January 1994

Abstract

Stochastic models for gene frequencies can be viewed as probability-measure-valued processes. Fleming and Viot introduced a class of processes that arise as limits of genetic models as the population size and the number of possible genetic types tend to infinity. In general, the topology on the process values in which these limits exist is the topology of weak convergence; however, convergence in the weak topology is not strong enough for many genetic applications. A new topology on the space of finite measures is introduced in which convergence implies convergence of the sizes and locations of atoms, and conditions are given under which genetic models converge in this topology. As an application, Kingman's Poisson–Dirichlet limit is extended to models with selection.

Key words: Fleming–Viot process; Measure-valued diffusion; Convergence in distribution; Weak topology

1. Introduction

Consider a population of M individuals, each of which is assigned a "type" x from a set of types that we identify with a complete separable metric space (E,r). (The generality of the type space allows for a broad range of interpretations of the term "type". See Ethier and Kurtz (1993) for a survey of examples.) At each of a discrete set of times we form a new "generation" of M individuals in the following manner. For each new individual, a pair of parents is selected from the current generation of M individuals with types $x_1, \ldots, x_M \in E$. The probability that the ith and jth individuals are selected as parents is

$$\frac{w_M(x_i,x_j)}{\sum_{k,l=1}^{M} w_M(x_k,x_l)}, \tag{1.1}$$

*Corresponding author.

SSDI 0304-4149(94)00006-F

where $w_M(x_i,x_j) \geq 0$ is symmetric and represents the "fitness" of the pair (x_i,x_j). Conditioned on the types of the parents, the probability distribution of the "tentative" type of the new individual is given by a transition function $\eta_M(x_i,x_j,\Gamma)$ from $E \times E$ to E. This tentative type z then "mutates" according to a transition function $P_M(z,\Gamma)$ from E to E. Conditioned on the current generation $(x_1,\ldots,x_M)$, the M types in the new generation are independent and identically distributed.

For a genetic model of a diploid population (i.e., one in which chromosomes occur in pairs), M would be even, $(x_1,\ldots,x_M)$ would be the collection of gametic types in the population, and $w_M(x_i,x_j)$ would be the fitness of the diploid organism with genotype (x_i,x_j).

Since the order in $(x_1,\ldots,x_M)$ is not important, we now identify the population consisting of $(x_1,\ldots,x_M) \in E^M$ with the empirical measure

$$\frac{1}{M}\sum_{i=1}^{M}\delta_{x_i} \in \mathscr{P}(E), \tag{1.2}$$

where $\mathscr{P}(E)$ denotes the set of Borel probability measures on E, and we define the $\mathscr{P}(E)$-valued process Z_M by letting $Z_M(t)$ be the measure corresponding to generation $[Mt]$. Assume there exist a symmetric function $\sigma \in B(E \times E)$, $\alpha \geq 0$, and a transition function $\eta(x,y,\Gamma)$ from $E \times E$ to E such that, for M sufficiently large,

$$w_M(x,y) = 1 + M^{-1}\sigma(x,y) \tag{1.3}$$

and

$$\eta_M(x,y,\Gamma) = \left(1 - \frac{\alpha}{M}\right)\left(\frac{1}{2}\delta_x(\Gamma) + \frac{1}{2}\delta_y(\Gamma)\right) + \frac{\alpha}{M}\eta(x,y,\Gamma), \tag{1.4}$$

and define the operators Q_M and B_M on $B(E)$ by

$$Q_M f(x) = \int_E f(y)P_M(x,\mathrm{d}y), \quad B_M = M(Q_M - I). \tag{1.5}$$

Under appropriate hypotheses on the type space E, the convergence of B_M, and the continuity of η and σ, the sequence of measure-valued processes Z_M converges in distribution as $M \to \infty$ to a $\mathscr{P}(E)$-valued Markov process Z of a type introduced by Fleming and Viot (1979). We define the generator $\mathscr{A}$ for a Fleming–Viot process as follows.

For $m \geq 2$ and $1 \leq i < j \leq m$, define the *sampling operators* $\Phi_{ij} : B(E^m) \mapsto B(E^{m-1})$ by letting $\Phi_{ij}f$ be the function obtained from f by replacing x_j by x_i and renumbering the variables (e.g., for $f(x_1,x_2,x_3) \in B(E^3)$, $\Phi_{12}f(x_1,x_2) = f(x_1,x_1,x_2)$, $\Phi_{13}f(x_1,x_2) = f(x_1,x_2,x_1)$, and $\Phi_{23}f(x_1,x_2) = f(x_1,x_2,x_2)$). The sampling operators arise in the definition of $\mathscr{A}$ as the limit of the terms in the generator for Z_M corresponding to the sampling from the previous generation.

Let L_0 be a closed subspace of $\bar{C}(E)$ that is separating. Suppose that B generates a strongly continuous positive conservative contraction semigroup $\{T(t)\}$ on L_0 given by a transition function $P(t,x,\Gamma)$, i.e.,

$$T(t)f(x) = \int_E f(y)P(t,x,\mathrm{d}y). \tag{1.6}$$

B will be called the *mutation operator* for the process and is the limit, in an appropriate sense, of the operators B_M. For each $m \geq 1$, define the semigroup $\{T^{(m)}(t)\}$ on $B(E^m)$ by

$$T^{(m)}(t)f(x_1,\ldots,x_m) = \int_E \cdots \int_E f(y_1,\ldots,y_m)P(t,x_1,\mathrm{d}y_1)\cdots P(t,x_m,\mathrm{d}y_m), \quad (1.7)$$

and let $B^{(m)}$ denote its generator; note that $\mathscr{D}(B^{(m)})$ is a subspace of $B(E^m)$.

Let $\eta(x_1,x_2,\Gamma)$ be a transition function from $E \times E$ to E. For $m \geq 1$ and $i = 1, \ldots, m$, define the *recombination operators* $H_{im} : B(E^m) \mapsto B(E^{m+1})$ by

$$H_{im}f(x_1, \ldots, x_{m+1}) = \int_E f(x_1, \ldots, x_{i-1},z,x_{i+1}, \ldots, x_m)\eta(x_i,x_{m+1},\mathrm{d}z), \quad (1.8)$$

and assume that $H_{im} : \bar{C}(E^m) \mapsto \bar{C}(E^{m+1})$. For example, in a two-locus model (see Ethier and Griffiths (1990)), $E = E_1 \times E_2$ and $\eta((y_1,z_1),(y_2,z_2),\cdot) = \frac{1}{2}\delta_{(y_1,z_2)} + \frac{1}{2}\delta_{(y_2,z_1)}$. $\alpha \geq 0$ will denote the *recombination intensity*.

Given σ in $B_{\mathrm{sym}}(E \times E)$, the set of symmetric functions in $B(E \times E)$, define $\bar{\sigma} = \sup_{x,y,z} |\sigma(x,y) - \sigma(y,z)|$, and for $m \geq 1$ and $i = 1, \ldots, m$, define the *selection operators* $K_{im} : B(E^m) \mapsto B(E^{m+2})$ by

$$K_{im}f(x_1, \ldots, x_{m+2}) = \frac{\sigma(x_i,x_{m+1}) - \sigma(x_{m+1},x_{m+2})}{\bar{\sigma}} f(x_1, \ldots, x_m), \quad (1.9)$$

where $0/0 = 0$. σ is called the *selection intensity function*.

Unless otherwise specified, we will consider $\mathscr{P}(E)$ as a metric space with the Prohorov metric ρ, and $\mathscr{B}(\mathscr{P}(E)) \equiv \mathscr{B}(\mathscr{P}(E),\rho)$ will denote the collection of Borel subsets of $\mathscr{P}(E)$. For $m \geq 1$ and $f \in B(E^m)$, define $F \in B(\mathscr{P}(E))$ by $F(\mu) = \langle f, \mu^m \rangle = \int_{E^m} f \,\mathrm{d}\mu^m$, where μ^m denotes the m-fold product measure of μ, and let $\mathscr{L}^m(E) \subset B(\mathscr{P}(E))$ be the collection of functions of this form. Note that $\mathscr{L}^m(E)$ is a linear subspace of $B(\mathscr{P}(E))$ closed under bounded pointwise convergence. For $f \in \mathscr{D}(B^{(m)})$, define $\mathscr{A}F \in B(\mathscr{P}(E))$ by

$$\begin{aligned}\mathscr{A}F(\mu) = &\sum_{1\leq i<j\leq m} (\langle \Phi_{ij}f, \mu^{m-1}\rangle - \langle f,\mu^m\rangle) + \langle B^{(m)}f, \mu^m\rangle \\ &+\alpha\sum_{i=1}^{m}(\langle H_{im}f,\mu^{m+1}\rangle - \langle f,\mu^m\rangle) + \bar{\sigma}\sum_{i=1}^{m}\langle K_{im}f,\mu^{m+2}\rangle,\end{aligned} \quad (1.10)$$

or, more generally, let $\tilde{\mathscr{A}} \subset B(\mathscr{P}(E)) \times B(\mathscr{P}(E))$ consist of all pairs

$$\begin{aligned}\Big(\langle f,\mu^m\rangle, &\sum_{1\leq i<j\leq m} (\langle \Phi_{ij}f, \mu^{m-1}\rangle - \langle f,\mu^m\rangle) + \langle g, \mu^m\rangle \\ &+\alpha\sum_{i=1}^{m}(\langle H_{im}f,\mu^{m+1}\rangle - \langle f,\mu^m\rangle) + \bar{\sigma}\sum_{i=1}^{m}\langle K_{im}f,\mu^{m+2}\rangle\Big)\end{aligned} \quad (1.11)$$

with $(f,g) \in \hat{B}^{(m)}$ and $m \geq 1$, where $\hat{B}^{(m)}$ is the bounded pointwise closure of $\{(f,B^{(m)}f) : f \in \mathscr{D}(B^{(m)})\}$. Note that any solution of the martingale problem for $\mathscr{A}$ will be a solution for $\tilde{\mathscr{A}}$.

Existence of solutions of the martingale problem for $\mathscr{A}$, continuous in the weak topology on $\mathscr{P}(E)$, has been proved under a variety of assumptions (e.g., Fleming and Viot, 1979; Kurtz, 1981; Ethier and Kurtz, 1986, 1987). Uniqueness follows by a duality argument. The following theorem extends Theorem 2.3(b) of Ethier and Kurtz (1987) to the above model including recombination. The proof is essentially the same and is omitted.

Theorem 1.1. (a) *Let E be compact, let $L_0 = C(E)$, and let $\sigma \in C_{\text{sym}}(E \times E)$. Let Z_M, B_M, and B be as above, and let $\nu_M \in \mathscr{P}(\mathscr{P}(E))$ denote the distribution of $Z_M(0)$. Suppose that $\nu_M \Rightarrow \nu \in \mathscr{P}(\mathscr{P}(E))$ and that*

$$B \subset \operatorname*{ex-lim}_{M \to \infty} B_M \tag{1.12}$$

(or equivalently e^{tB_M} converges strongly to $T(t)$ on $C(E)$ for each $t \geq 0$; see Ethier and Kurtz, 1986, Chapter 1). Then, under the above hypotheses on $w_M, \sigma, \eta_M, \alpha$, and η, there exists a unique solution Z of the $C_{(\mathscr{P}(E),\rho)}[0,\infty)$ martingale problem for $\mathscr{A}$ with initial distribution ν, and $Z_M \Rightarrow Z$ in $D_{(\mathscr{P}(E),\rho)}[0,\infty)$, where ρ denotes the Prohorov metric on $\mathscr{P}(E)$ (i.e., $\mathscr{P}(E)$ is given the weak topology).

(b) *Let E be compact, let $L_0 = C(E)$, and let $\sigma \in B_{\text{sym}}(E \times E)$. Then, under the above assumptions on B, α, and η, and for each $\nu \in \mathscr{P}(\mathscr{P}(E))$, existence and uniqueness of solutions of the $C_{(\mathscr{P}(E),\rho)}[0,\infty)$ martingale problem for $\mathscr{A}$ with initial distribution ν hold.*

As will be observed in Section 3, there are many situations in which this convergence result is not adequate. In particular, in the derivation of finite sampling formulas, one needs to know that the distribution of the sizes of the atoms of Z_M converges, but convergence in the weak topology on $\mathscr{P}(E)$ does not imply convergence of the sizes of the atoms. In Section 2, we introduce a new metric ρ_{a} on the space of finite Borel measures on E under which the sizes and locations of atoms become continuous functions. We refer to the topology determined by this metric as the *weak atomic topology*. In Section 3, we give conditions that imply that the convergence in Theorem 1.1 is in $D_{(\mathscr{P}(E),\rho_{\text{a}})}[0,\infty)$. We also give similar convergence results for sequences of Fleming–Viot processes and stationary distributions. (We will see that $\mathscr{B}(\mathscr{P}(E),\rho_{\text{a}}) = \mathscr{B}(\mathscr{P}(E),\rho)$, so that the notion of a $\mathscr{P}(E)$-valued random variable will not be ambiguous.) Section 4 is devoted to an extension to models with selection of Kingman's (1975) limit theorem (as $n \to \infty$) for the descending order statistics of the allele frequencies in the stationary neutral n-allele diffusion model. This result requires the use of the weak atomic topology introduced in Section 2.

2. The weak atomic topology on the space of measures

Let (E,r) be a complete separable metric space. For μ in $\mathscr{M}_{\text{f}}(E)$, the space of finite, positive, Borel measures on E with the weak topology, define μ^* to be the purely atomic measure given by $\mu^* = \sum \mu(\{x\})^2 \delta_x$. Note that μ^* is also given by $\mu^*(A) = \inf \sum \mu(E_i)^2$, where the infimum is over all countable collections of Borel sets $\{E_i\}$ with $A \subset \cup E_i$, and by

$$\mu^*(A) = \lim_n \sum_i \mu(A \cap E_i^{(n)})^2 , \tag{2.1}$$

where $\{E_i^{(n)}\}$, $n = 1,2,\ldots$, is a sequence of countable Borel partitions of E with the property that $\lim_n \sup_i \operatorname{diam}(E_i^{(n)}) = 0$.

Let ρ denote the Prohorov metric on $\mathcal{M}_{\mathrm{f}}(E)$, let $\Psi : [0,\infty) \mapsto [0,1]$ be continuous and nonincreasing with $\Psi(0) = 1$ and $\Psi(1) = 0$, and define

$$\rho_{\mathrm{a}}(\mu,\nu) = \rho(\mu,\nu) + \sup_{0<\varepsilon\le 1}\left|\int_E\int_E \Psi\left(\frac{r(x,y)}{\varepsilon}\right)\mu(\mathrm{d}x)\mu(\mathrm{d}y) - \int_E\int_E \Psi\left(\frac{r(x,y)}{\varepsilon}\right)\nu(\mathrm{d}x)\nu(\mathrm{d}y)\right|. \tag{2.2}$$

ρ_{a} is easily seen to be a metric on $\mathcal{M}_{\mathrm{f}}(E)$. Note that

$$\lim_{\varepsilon\to 0}\int_E\int_E \Psi\left(\frac{r(x,y)}{\varepsilon}\right)\mu(\mathrm{d}x)\mu(\mathrm{d}y) = \mu^*(E). \tag{2.3}$$

The supremum in (2.2) could be restricted to rational ε, and hence, for fixed $\nu \in \mathcal{M}_{\mathrm{f}}(E)$, $\rho_{\mathrm{a}}(\mu,\nu)$ is a $\mathcal{B}(\mathcal{M}_{\mathrm{f}}(E),\rho)$-measurable function of μ. It follows that $\mathcal{B}(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}}) = \mathcal{B}(\mathcal{M}_{\mathrm{f}}(E),\rho)$, and consequently, notation such as $\mathcal{B}(\mathcal{M}_{\mathrm{f}}(E))$ is not ambiguous.

Lemma 2.1. *Let $\mu_n, \mu \in \mathcal{M}_{\mathrm{f}}(E)$. Suppose $\mu_n \Rightarrow \mu$. Then $\{\mu_n^*\}$ is relatively compact, and any limit point $\nu \in \mathcal{M}_{\mathrm{f}}(E)$ satisfies $\nu(A) \le \mu^*(A)$ for all $A \in \mathcal{B}(E)$ (in particular, it is purely atomic), and $\nu = \mu^*$ if and only if $\nu(E) = \mu^*(E)$. Consequently, $\mu_n^* \Rightarrow \mu^*$ if and only if $\mu_n^*(E) \to \mu^*(E)$.*

Proof. A sequence in $\mathcal{M}_{\mathrm{f}}(E)$ is relatively compact if and only if it is bounded and tight. Since $\mu_n^*(E) \le \mu_n(E)^2$ and $\mu_n^*(K^{\mathrm{c}}) \le \mu_n(K^{\mathrm{c}})^2$ for all compact $K \subset E$, the relative compactness of $\{\mu_n\}$ implies the relative compactness of $\{\mu_n^*\}$. Assume, for simplicity, that $\mu_n^* \Rightarrow \nu$. Then, for every $g \in \bar{C}(E)$ with $g \ge 0$,

$$\begin{aligned}\int_E g(x)\nu(\mathrm{d}x) &= \lim_{n\to\infty}\int_E g(x)\mu_n^*(\mathrm{d}x)\\ &\le \inf_{0<\varepsilon\le 1}\lim_{n\to\infty}\int_E\int_E g(x)\Psi\left(\frac{r(x,y)}{\varepsilon}\right)\mu_n(\mathrm{d}x)\mu_n(\mathrm{d}y)\\ &= \inf_{0<\varepsilon\le 1}\int_E\int_E g(x)\Psi\left(\frac{r(x,y)}{\varepsilon}\right)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &= \int_E g(x)\mu^*(\mathrm{d}x),\end{aligned} \tag{2.4}$$

which implies the desired inequality. Given ν, $\eta \in \mathcal{M}_{\mathrm{f}}(E)$ satisfying $\nu(A) \le \eta(A)$ for all $A \in \mathcal{B}(E)$, $\nu = \eta$ if and only if $\nu(E) = \eta(E)$. (Consider complements.) □

Lemma 2.2. *Let $\mu_n, \mu \in \mathcal{M}_{\mathrm{f}}(E)$. Then $\rho_{\mathrm{a}}(\mu_n,\mu) \to 0$ if and only if $\mu_n \Rightarrow \mu$ and $\mu_n^* \Rightarrow \mu^*$.*

Proof. By (2.3), if $\rho_{\mathrm{a}}(\mu_n,\mu) \to 0$, then $\mu_n^*(E) \to \mu^*(E)$ and it follows from Lemma 2.1 that $\mu_n^* \Rightarrow \mu^*$.

Assume now that $\mu_n \Rightarrow \mu$ and $\mu_n^* \Rightarrow \mu^*$. Let $f_n(\varepsilon)$ and $f(\varepsilon)$ denote the integrals in the definition of $\rho_{\mathrm{a}}(\mu_n,\mu)$. We must show that $f_n \to f$ uniformly on (0, 1]. We know that for $0 < \varepsilon_0 < 1, f_n \to f$ uniformly on $[\varepsilon_0, 1]$, so there exists a sequence $\varepsilon_n \to 0$ sufficiently slowly that

$$\sup_{\varepsilon_n \le \varepsilon \le 1} |f_n(\varepsilon) - f(\varepsilon)| \to 0 . \tag{2.5}$$

On the other hand,

$$\begin{aligned} &\sup_{0<\varepsilon\le\varepsilon_n} |f_n(\varepsilon) - f(\varepsilon)| \\ &\quad \le \sup_{0<\varepsilon\le\varepsilon_n} |f_n(\varepsilon) - f_n(0+)| + |f_n(0+) - f(0+)| + \sup_{0<\varepsilon\le\varepsilon_n} |f(0+) - f(\varepsilon)| \\ &\quad = |f_n(\varepsilon_n) - f_n(0+)| + |f_n(0+) - f(0+)| + |f(0+) - f(\varepsilon_n)| \\ &\quad \le |f_n(\varepsilon_n) - f(\varepsilon_n)| + 2|f(\varepsilon_n) - f(0+)| + 2|f(0+) - f_n(0+)| , \end{aligned} \tag{2.6}$$

where the equality uses the monotonicity of f_n and f. On the right-hand side of (2.6), the first term tends to 0 by (2.5), the second term tends to 0 by definition, and the third term tends to 0 by the convergence of μ_n^* to μ^* (see (2.3)). Thus, the proof is complete. □

Recall that we are assuming that (E, r) is complete and separable.

Lemma 2.3 *$(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$ is complete and separable.*

Proof. Completeness follows from the completeness of $(\mathscr{M}_{\mathrm{f}}(E), \rho)$. To check separability, let $\{E_i^{(n)}\}$, $n = 1, 2, \ldots$, be a sequence of countable Borel partitions of E with $\mathrm{diam}(E_i^{(n)}) < 1/n$. Let $x_i^n \in E_i^{(n)}$, and given $\mu \in \mathscr{M}_{\mathrm{f}}(E)$, define

$$\mu_n = \sum_{i=1}^{\infty} \mu(E_i^{(n)}) \delta_{x_i^n} . \tag{2.7}$$

It follows easily that $\mu_n \Rightarrow \mu$ and $\mu_n^* \Rightarrow \mu^*$, and hence $\rho_{\mathrm{a}}(\mu_n, \mu) \to 0$. Let $D = \{x_i^n : i, n = 1, 2, \ldots\}$, and let $\mathscr{M}_0 = \{\sum_{i=1}^k a_i \delta_{x_i} : a_i \in \boldsymbol{Q}_+,\ x_i \in D,\ k \ge 1\}$. (Note that $\mathscr{M}_0$ is countable.) Since μ_n in (2.7) can be approximated arbitrarily closely in the ρ_{a} metric by elements of $\mathscr{M}_0$, it follows that $\mathscr{M}_0$ is dense in $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$. □

An examination of the proof of Lemma 2.2 immediately gives the following.

Lemma 2.4. *A collection of measures $\{\mu_\alpha\} \subset \mathscr{M}_{\mathrm{f}}(E)$ is relatively compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$ if and only if it is relatively compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho)$ and*

$$\lim_{\varepsilon\to 0} \sup_\alpha \left(\int_E \int_E \Psi\left(\frac{r(x,y)}{\varepsilon}\right) \mu_\alpha(\mathrm{d}x)\mu_\alpha(\mathrm{d}y) - \mu_\alpha^*(E) \right) = 0 . \tag{2.8}$$

In particular, if K is compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho)$ and $\{\varepsilon_k\}$ is a positive sequence tending to 0, then

$$\hat{K} = \left\{ \mu \in K : \int_E \int_E \Psi\left(\frac{r(x,y)}{\varepsilon_k}\right) \mu(\mathrm{d}x)\mu(\mathrm{d}y) - \mu^*(E) \le \frac{1}{k} \ for\ all\ k \ge 1 \right\} \tag{2.9}$$

is compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$.

The next lemma makes precise the sense in which sizes and locations of atoms are continuous in the weak atomic topology.

Lemma 2.5. (a) *If $\rho_a(\mu_n, \mu) \to 0$, then the sizes and locations of the atoms of μ_n converge to the sizes and locations of the atoms of μ in the sense that for each atom $\alpha\delta_x$ of μ, there exists a sequence of atoms $\alpha_n\delta_{x_n}$ of μ_n such that $\lim_{n\to\infty}(\alpha_n, x_n) = (\alpha, x)$, and any sequence of atoms $\alpha_n\delta_{x_n}$ of μ_n satisfying $\inf_n \alpha_n > 0$ contains a subsequence converging to an atom of μ.*

(b) *Suppose that $\mu_n \Rightarrow \mu$. Let $\{\alpha_i^n\delta_{x_i^n}\}$ be the set of atoms of μ_n ordered so that $\alpha_1^n \geq \alpha_2^n \geq \ldots$, and let $\{\alpha_i\delta_{x_i}\}$ be the set of atoms of μ with $\alpha_1 \geq \alpha_2 \geq \ldots$. Then $\rho_a(\mu_n, \mu) \to 0$ if and only if $\alpha_i^n \to \alpha_i$ for each i. If $\rho_a(\mu_n, \mu) \to 0$ and $\alpha_k > \alpha_{k+1}$ for some $k \geq 1$, then the set of locations $\{x_1^n \ldots, x_k^n\}$ converges to $\{x_1, \ldots, x_k\}$. In particular, if $\alpha_1 > \alpha_2 > \ldots$, then $x_i^n \to x_i$ for each $i \geq 1$.*

(c) *Suppose that $\mu_n \Rightarrow \mu$ and that μ is purely atomic. Then $\rho_a(\mu_n, \mu) \to 0$ if and only if $\sum_i |\alpha_i^n - \alpha_i| \to 0$, where α_i^n and α_i are as in part* (b).

Remark 2.6. Let $\mu_n^* = \sum_i a_i^n \delta_{x_i^n}$ and $\mu^* = \sum_i a_i \delta_{x_i}$. Let $f \in \bar{C}(E \times (0, \infty))$ and suppose that for some $\varepsilon > 0$, $f(x,a) = 0$ for all $a \leq \varepsilon$ and $x \in E$. By part (a), $\rho_a(\mu_n, \mu) \to 0$ implies $\sum_i f(x_i^n, a_i^n) \to \sum_i f(x_i, a_i)$.

Proof. Let a_i^n, x_i^n, a_i, and x_i be as in the remark. If we show that for each i there exists a sequence $\{i_n\}$ such that $(a_{i_n}^n, x_{i_n}^n) \to (a_i, x_i)$, the first conclusion of part (a) will follow. Fix i. By the definition of the Prohorov metric, we have

$$\begin{aligned} a_i = \mu^*(\{x_i\}) &\leq \mu_n^*(\{x \in E : r(x, x_i) \leq \rho(\mu_n^*, \mu^*)\}) + \rho(\mu_n^*, \mu^*) \\ &\leq \mu^*(\{x \in E : r(x, x_i) \leq 2\rho(\mu_n^*, \mu^*)\}) + 2\rho(\mu_n^*, \mu^*) \to a_i\,. \end{aligned} \tag{2.10}$$

Hence

$$\sum_{j : r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)} a_j^n \to a_i\,. \tag{2.11}$$

It remains to show that $\max\{a_j^n : r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)\} \to a_i$. By standard properties of weak convergence

$$\begin{aligned} &\limsup_{n\to\infty} \sum_{j : r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)} \sqrt{a_j^n} \\ &\leq \limsup_{n\to\infty} \mu_n(\{x \in E : r(x, x_i) \leq \rho(\mu_n^*, \mu^*)\}) \leq \mu(\{x_i\}) = \sqrt{a_i}\,. \end{aligned} \tag{2.12}$$

Let j_n be the index of the largest value of a_j^n with $r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)$. Then, by (2.11) and (2.12),

$$\begin{aligned} \limsup_{n\to\infty} \sqrt{a_{j_n}^n} \sum_{j : r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)} \sqrt{a_j^n} &\leq a_i \\ &= \lim_{n\to\infty} \sum_{j : r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)} a_j^n \\ &\leq \liminf_{n\to\infty} \sqrt{a_{j_n}^n} \sum_{j : r(x_j^n, x_i) \leq \rho(\mu_n^*, \mu^*)} \sqrt{a_j^n}\,, \end{aligned} \tag{2.13}$$

and it follows that $a^n_{j_n} \to a_i$, proving the first conclusion of part (a).

As for the second conclusion of part (a), given a sequence of atoms $\alpha_n\delta_{x_n}$ of μ_n with $\inf_n \alpha_n > 0$, tightness implies that $\{x_n : n \geq 1\}$ is relatively compact. Therefore, there exist $\alpha > 0$, $x \in E$, and a subsequence $\{n'\}$ along which $(\alpha_{n'}, x_{n'}) \to (\alpha, x)$. Letting $S_\varepsilon(x) = \{y \in E : r(y,x) < \varepsilon\}$, we have $\alpha \leq \limsup_{n'} \mu_{n'}(\overline{S_\varepsilon(x)}) \leq \mu(\overline{S_\varepsilon(x)})$ for every $\varepsilon > 0$, hence $\mu(\{x\}) \geq \alpha$. To complete the proof of part (a), we must show that this last inequality in an equality. Suppose not. Then $\mu(\{x\}) = \beta > \alpha$, so by the first conclusion of part (a), there exists a sequence of atoms $\beta_n\delta_{y_n}$ of μ_n such that $(\beta_n, y_n) \to (\beta, x)$. But then $\alpha + \beta \leq \limsup_{n'} \mu_{n'}(\overline{S_\varepsilon(x)}) \leq \mu(\overline{S_\varepsilon(x)})$ for every $\varepsilon > 0$, hence $\alpha + \beta \leq \beta$, a contradiction.

The necessity in part (b) follows from part (a). If $\alpha^n_i \to \alpha_i$ for each i, then $\mu^*_n(E) = \sum_i (\alpha^n_i)^2 \to \sum_i (\alpha_i)^2 = \mu^*(E)$, and the result follows by Lemmas 2.1 and 2.2. The last part of (b) follows from (a).

Part (c) is left to the reader. □

Example 2.7. Let $\boldsymbol{n} = (n_1, \ldots, n_d) \in \boldsymbol{N}^d$, $n = n_1 + \cdots + n_d$, and $\Lambda_{\boldsymbol{n}} = \{(x_1, \ldots, x_n) \in E^n : \exists y_1, \ldots, y_d$ distinct in E with n_k of the x_i equal to y_k for $k = 1, \ldots, d\}$. Then the mapping $\varphi_{\boldsymbol{n}} : (\mathscr{P}(E), \rho_{\mathrm{a}}) \mapsto [0,1]$ given by $\varphi_{\boldsymbol{n}}(\mu) = \mu^n(\Lambda_{\boldsymbol{n}})$ is continuous. Note that this would not be the case, in general, if ρ_{a} were replaced by ρ.

Lemma 2.8. *Suppose $x \in D_{(\mathscr{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$ (resp., $C_{(\mathscr{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$). If $\{x(t) : t \leq T\}$ is relatively compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$ for each $T > 0$, then $x \in D_{(\mathscr{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$ (resp., $C_{(\mathscr{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$).*

Proof. Since convergence in ρ_{a} implies convergence in ρ, any ρ_{a}-limit point of $x(s)$ as $s \to t+$ must be $x(t)$. Consequently, $\lim_{s\to t+} \rho_{\mathrm{a}}(x(s), x(t)) = 0$. A similar argument applies to left limits. □

As noted earlier, $\mathscr{B}(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}}) = \mathscr{B}(\mathscr{M}_{\mathrm{f}}(E), \rho)$, so we can speak of $\mathscr{M}_{\mathrm{f}}(E)$-valued random variables without ambiguity.

Lemma 2.9. *Let $\{\Gamma_\alpha\}$ be a collection of random variables with values in $\mathscr{M}_{\mathrm{f}}(E)$. Suppose it is relatively compact as a collection of random variables in $(\mathscr{M}_{\mathrm{f}}(E), \rho)$. Then it is relatively compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$ if and only if for every $\delta > 0$ there exists $\varepsilon > 0$ such that*

$$\inf_\alpha P\left\{ \int_E \int_E \Psi\left(\frac{r(x,y)}{\varepsilon}\right) \Gamma_\alpha(dx)\Gamma_\alpha(dy) - \Gamma^*_\alpha(E) \leq \delta \right\} \geq 1 - \delta\,. \tag{2.14}$$

Proof. By Prohorov's theorem, given $\eta > 0$ we must show the existence of $\hat{K}$ compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$ such that $\inf_\alpha P\{\Gamma_\alpha \in \hat{K}\} \geq 1 - \eta$. Select K compact in $(\mathscr{M}_{\mathrm{f}}(E), \rho)$ such that $\inf_\alpha P\{\Gamma_\alpha \in K\} \geq 1 - \eta/2$ and a positive sequence $\{\varepsilon_k\}$ tending to 0 such that

$$\inf_\alpha P\left\{ \int_E \int_E \Psi\left(\frac{r(x,y)}{\varepsilon_k}\right) \Gamma_\alpha(\mathrm{d}x)\Gamma_\alpha(\mathrm{d}y) - \Gamma^*_\alpha(E) \leq \frac{1}{k} \right\} \geq 1 - \frac{\eta}{2^{k+1}}\,. \tag{2.15}$$

Then, if $\hat{K}$ is defined as in (2.9), $\inf_\alpha P\{\Gamma_\alpha \in \hat{K}\} \geq 1 - \eta$.

The converse is a consequence of Lemma 2.4. □

The following lemma gives simple conditions under which $\{\Gamma_\alpha\}$ is relatively compact as a collection of random variables in $(\mathcal{M}_{\mathrm{f}}(E), \rho)$.

Lemma 2.10. *Let $\{\Gamma_\alpha\}$ be a collection of $\mathcal{M}_{\mathrm{f}}(E)$-valued random variables. For each α, define the positive Borel measure μ_α on E by $\mu_\alpha(A) = E[\Gamma_\alpha(A)]$. If $\{\mu_\alpha\} \subset \mathcal{M}_{\mathrm{f}}(E)$ and $\{\mu_\alpha\}$ is relatively compact in $(\mathcal{M}_{\mathrm{f}}(E), \rho)$, then $\{\Gamma_\alpha\}$ is relatively compact as a collection of $(\mathcal{M}_{\mathrm{f}}(E), \rho)$-valued random variables. If the random variables Γ_α are $\mathcal{P}(E)$-valued, then the relative compactness of $\{\mu_\alpha\}$ is necessary as well as sufficient for the relative compactness of $\{\Gamma_\alpha\}$.*

Proof. Recall that $\{\mu_\alpha\}$ is relatively compact in $(\mathcal{M}_{\mathrm{f}}(E), \rho)$ if and only if $\sup_\alpha \mu_\alpha(E) < \infty$ and for every $\varepsilon > 0$ there exists a compact $K \subset E$ such that $\sup_\alpha \mu_\alpha(K^{\mathrm{c}}) \leq \varepsilon$. Noting that

$$P\{\Gamma_\alpha(E) \geq c\} \leq \frac{\mu_\alpha(E)}{c} \tag{2.16}$$

and that for each compact $K \subset E$ and $\delta > 0$

$$P\{\Gamma_\alpha(K^{\mathrm{c}}) \geq \delta\} \leq \frac{\mu_\alpha(K^{\mathrm{c}})}{\delta}, \tag{2.17}$$

we see that for each $\varepsilon > 0$ there exist $c > 0$ and compact sets $K_1 \subset K_2 \subset \cdots \subset E$ such that

$$\sup_\alpha P(\{\Gamma_\alpha(E) \geq c\} \cup \cup_i \{\Gamma_\alpha(K_i^{\mathrm{c}}) \geq 2^{-i}\}) \leq \varepsilon. \tag{2.18}$$

Since $\hat{K} \equiv \{\mu \in \mathcal{M}_{\mathrm{f}}(E) : \mu(E) \leq c,\ \mu(K_i^{\mathrm{c}}) \leq 2^{-i},\ i = 1,2,\ldots\}$ is compact in $(\mathcal{M}_{\mathrm{f}}(E), \rho)$, and since (2.18) implies $\inf_\alpha P\{\Gamma_\alpha \in \hat{K}\} \geq 1 - \varepsilon$, the relative compactness follows by Prohorov's theorem.

If $\{\Gamma_\alpha\}$ is relatively compact, then for every $\delta > 0$ there exists a compact $K \subset E$ such that $\sup_\alpha P\{\Gamma_\alpha(K^{\mathrm{c}}) \geq \delta\} \leq \delta$, and hence if the Γ_α are $\mathcal{P}(E)$-valued, then $\sup_\alpha E[\Gamma_\alpha(K^{\mathrm{c}})] \leq 2\delta$. □

Let (H, r_0) be a metric space, and suppose that r_1 is a second metric on H such that $r_0 \leq r_1$ and $\mathcal{B}(H, r_0) = \mathcal{B}(H, r_1)$. A collection $\{Z_\alpha\}$ of H-valued processes with almost all sample paths in $D_{(H, r_0)}[0, \infty)$ satisfies the *compact containment* condition in (H, r_1) if for each $T > 0$ and $\delta > 0$, there exists an r_1-compact set $K_{T,\delta} \subset H$ such that

$$\inf_\alpha P\{Z_\alpha(t) \in K_{T,\delta},\ t \leq T\} \geq 1 - \delta. \tag{2.19}$$

See Ethier and Kurtz (1986, Chapter 3) for further discussion. If a single $\mathcal{M}_{\mathrm{f}}(E)$-valued process Z with almost all sample paths in $D_{(\mathcal{M}_{\mathrm{f}}(E), \rho)}[0, \infty)$ (resp., $C_{(\mathcal{M}_{\mathrm{f}}(E), \rho)}[0, \infty)$) satisfies the compact containment condition in $(\mathcal{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$, then by Lemma 2.8 the process has almost all sample paths in $D_{(\mathcal{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})}[0, \infty)$ (resp., $C_{(\mathcal{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})}[0, \infty)$). Lemmas 2.1 and 2.2 give the following, perhaps simpler, criterion for sample path continuity.

Lemma 2.11. *Suppose Z is an $\mathcal{M}_{\mathrm{f}}(E)$-valued process with almost all sample paths in $C_{(\mathcal{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$. If $\{Z^*(t,E),\ t\geq 0\}$ has almost all sample paths in $C_{[0,\infty)}[0,\infty)$, then Z has almost all sample paths in $C_{(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$.*

Theorem 2.12. *Suppose that a collection $\{Z_\alpha\}$ of $\mathcal{M}_{\mathrm{f}}(E)$-valued processes is relatively compact in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$. Then it is relatively compact in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$ (in particular, each Z_α has almost all sample paths in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$) if and only if the compact containment condition holds in $(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})$.*

Remark 2.13. (a) By the argument used to prove Lemma 2.9, assuming relative compactness of $\{Z_\alpha\}$ in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$, the compact containment condition will hold in $(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})$ if and only if for each $T>0$ and $\delta>0$, there exists an $\varepsilon>0$ such that

$$\inf_\alpha P\left\{\sup_{t\leq T}\left(\int_E\int_E\Psi\left(\frac{r(x,y)}{\varepsilon}\right)Z_\alpha(t,\mathrm{d}x)Z_\alpha(t,\mathrm{d}y)-Z_\alpha^*(t,E)\right)\leq\delta\right\}\geq 1-\delta\,. \tag{2.20}$$

(b) As a consequence of Theorem 2.12 and Remark 2.13(a), a sequence $\{Z_n\}$ of processes with sample paths in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$ is relatively compact in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho_{\mathrm{a}})}[0,\infty)$ if and only if it is relatively compact in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$ and for each $T>0$ and $\delta>0$, there exists an $\varepsilon>0$ such that

$$\liminf_{n\to\infty} P\left\{\sup_{t\leq T}\left(\int_E\int_E\Psi\left(\frac{r(x,y)}{\varepsilon}\right)Z_n(t,\mathrm{d}x)Z_n(t,\mathrm{d}y)-Z_n^*(t,E)\right)\leq\delta\right\}\geq 1-\delta\,. \tag{2.21}$$

Proof. The necessity of the compact containment condition follows from Prohorov's theorem. To see that it is sufficient in this setting, it is enough to consider a sequence $\{Z_n\}$ that converges in distribution to a process Z in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho)}[0,\infty)$. Let $\varepsilon>0$. By the continuous mapping theorem,

$$\begin{aligned}U_n^\varepsilon&\equiv\int_E\int_E\Psi\left(\frac{r(x,y)}{\varepsilon}\right)Z_n(\cdot,\mathrm{d}x)Z_n(\cdot,\mathrm{d}y)\\&\Rightarrow\int_E\int_E\Psi\left(\frac{r(x,y)}{\varepsilon}\right)Z(\cdot,\mathrm{d}x)Z(\cdot,\mathrm{d}y)\equiv U^\varepsilon\,,\end{aligned} \tag{2.22}$$

and in fact $(Z_n,U_n^\varepsilon)\Rightarrow(Z,U^\varepsilon)$ in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho)\times\boldsymbol{R}}[0,\infty)$. The relative compactness of $\{U_n^\varepsilon\}$ and (2.21) imply the relative compactness of $\{U_n^0\}$ ($U_n^0\equiv Z_n^*(\cdot,E)$), and in fact the relative compactness of $\{(Z_n,U_n^\varepsilon)\}$ in $D_{(\mathcal{M}_{\mathrm{f}}(E),\rho)\times\boldsymbol{R}}[0,\infty)$ implies the relative compactness of $\{(Z_n,U_n^0)\}$. This assertion follows from Ethier and Kurtz (1986, Theorem 3.2.2; see also Problem 3.18). Since

$$\begin{aligned}\limsup_{n\to\infty}\rho_0(P_{U_n^0},P_{U^0})&\leq\limsup_{n\to\infty}\{\rho_0(P_{U_n^0},P_{U_n^\varepsilon})+\rho_0(P_{U_n^\varepsilon},P_{U^\varepsilon})+\rho_0(P_{U^\varepsilon},P_{U^0})\}\\&\leq\delta+e^{-T}+\rho_0(P_{U^\varepsilon},P_{U^0})\,,\end{aligned} \tag{2.23}$$

where ρ_0 denotes the Prohorov metric on $D_{\boldsymbol{R}}[0,\infty)$, P_U denotes the distribution of U, and ε, δ, and T are as in (2.21), it follows that $(Z_n, U_n^0) \Rightarrow (Z, U^0)$. By Lemma 2.1, the mapping $(\mu, \mu^*(E)) \mapsto \mu$ from $(\mathscr{M}_{\mathrm{f}}(E), \rho) \times \boldsymbol{R}$ to $(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})$ is continuous. Therefore, $Z_n \Rightarrow Z$ in $D_{(\mathscr{M}_{\mathrm{f}}(E), \rho_{\mathrm{a}})}[0,\infty)$ by the continuous mapping theorem. □

3. Convergence to Fleming–Viot processes

In Ethier and Kurtz (1987, Theorem 2.4), it was shown that Fleming–Viot processes with bounded mutation operators take values in $\mathscr{P}_{\mathrm{a}}(E)$, the set of purely atomic Borel probability measures on E. That paper did not consider models with recombination; however, the extension of the result to include recombination operators of the form described above follows by the same proof. In particular, if B is of the form

$$Bf(x) = \frac{1}{2}\theta \int_E (f(z) - f(x))\lambda(x, \mathrm{d}z), \tag{3.1}$$

where θ is a positive constant and λ is nonatomic, and if η in the recombination operator is also nonatomic, then the $\boldsymbol{R}^\infty$-valued process obtained from the Fleming–Viot process Z by taking the sequence of descending order statistics of the sizes of the atoms of Z is an infinitely-many-alleles diffusion of the type considered in Ethier and Kurtz (1981) with "mutation" intensity $\frac{1}{2}\theta + \alpha$. (Of course, the recombination operators of primary interest involve purely atomic η; however, the original process with type space E can be replaced by a new process with type space $E \times [0,1]$ and nonatomic "recombination" in such a way that the projection of the new process onto E is a version of the original process.)

Lemma 2.11 allows one to conclude that these processes have continuous sample paths in the weak atomic topology. (Shiga (1990) has obtained stronger results.)

Theorem 3.1. *Let Z be a Fleming–Viot process with bounded mutation operator B. If Z has almost all sample paths in $C_{(\mathscr{P}(E),\rho)}[0,\infty)$, then Z has almost all sample paths in $C_{(\mathscr{P}(E),\rho_{\mathrm{a}})}[0,\infty)$.*

Proof. For $m \geq 1$, $f \in \bar{C}(E^m)$, and $F(\mu) = \langle f, \mu^m \rangle$,

$$F(Z(t)) - \int_0^t \mathscr{A}F(Z(s))\,\mathrm{d}s \tag{3.2}$$

is an a.s. continuous martingale. It follows from the boundedness of B that the collection of $f \in B(E^m)$ for which (3.2) is an a.s. continuous martingale is closed under bounded pointwise convergence, and hence is all of $B(E^m)$. Taking $m = 2$ and $f = I_D$ with $D = \{(x,y) \in E^2 : x = y\}$, we see that $Z^*(t, E) = F(Z(t))$ is a.s. continuous in t, and the theorem follows by Lemma 2.11. □

The fact that Z is purely atomic for bounded mutation operators is fundamental to the derivation of a variety of sampling distributions. (See for example Ethier and Kurtz, 1986, Theorem 10.4.7). Since, however, Fleming–Viot processes are of interest as approximations to finite-population models, it is necessary to justify the validity of the sampling distributions derived from Fleming–Viot processes as approximations to the sampling distributions based on the finite-population models. This justification was done in Ethier and Kurtz (1986, Theorem 10.4.6), for the stationary distribution of a particular neutral model. It should be emphasized, however, that previously proved weak-approximation theorems (e.g., Fleming and Viot, 1979; Kurtz, 1981; Ethier and Kurtz, 1986, 1987) that take the weak topology on the state space $\mathscr{P}(E)$ do not imply convergence of the corresponding sampling distributions since the latter depend on the distributions of the sizes of the atoms (cf. Example 2.7). We have introduced the weak atomic topology precisely because it does imply convergence of the sizes of the atoms, and we can verify convergence of the sampling distributions by verifying convergence in distribution of the processes in $D_{(\mathscr{P}(E),\rho_a)}[0,\infty)$.

Theorem 3.2. *Let $Q_M, \eta_M, \alpha, \eta, w_M, \sigma$, and Z_M be as in the Introduction, and let E be compact. Suppose that*

$$Q_M f(x) = \left(1 - \frac{\theta_M(x)}{M}\right) f(x) + \frac{\theta_M(x)}{M} \int_E f(y)\lambda_M(x,dy) \tag{3.3}$$

for M sufficiently large, where $\theta_M \in B(E)$ is nonnegative and λ_M is a transition function on E, and that

$$\sup_M \sup_x \theta_M(x) < \infty\,, \tag{3.4}$$

$$\lim_{\varepsilon\to 0} \limsup_{M\to\infty} \sup_{x,y} \theta_M(x)\lambda_M(x,S_\varepsilon(y)) = 0\,, \tag{3.5}$$

where $S_\varepsilon(y) = \{z \in E : r(z,y) < \varepsilon\}$,

$$\lim_{\varepsilon\to 0} \sup_{x,y,z} \eta(x,y,S_\varepsilon(z)) = 0\,, \tag{3.6}$$

and (as we have already assumed)

$$\sup_{x,y} |\sigma(x,y)| < \infty\,. \tag{3.7}$$

If in addition $\{Z_M(0)\}$ is relatively compact in $(\mathscr{P}(E),\rho_a)$, then $\{Z_M\}$ is relatively compact in $D_{(\mathscr{P}(E),\rho_a)}[0,\infty)$.

Remark 3.3. The assumption that E is compact can be replaced by the assumption that $\{Z_M\}$ is relatively compact in $D_{(\mathscr{P}(E),\rho)}[0,\infty)$, the compact containment condition then taking the place of the compactness of E. See, for example, Remark 3.7.3 in Ethier and Kurtz (1986).

Proof of Theorem 3.2. Let $F(\mu) = \langle f, \mu^m\rangle$, where $f \in B(E^m)$ and $m \geq 1$. For each M let $\mathscr{P}_M(E)$ be the collection of all Borel probability measures on E of the form (1.2). For $\mu \in \mathscr{P}_M(E)$, define

$$\mathscr{A}_M F(\mu) = ME[\langle f, Z_M(1/M)^m\rangle - \langle f, \mu^m\rangle | Z_M(0) = \mu]\,. \tag{3.8}$$

The measure $Z_M(1/M)$ can be written

$$\frac{1}{M}\sum_{k=1}^{M}\delta_{Y_k}\,, \tag{3.9}$$

where, conditioned on $Z_M(0)=\mu$, the random variables $Y_1,\ldots,Y_M$ are independent and identically distributed with

$$\begin{aligned}
&E[g(Y_k)|Z_M(0)=\mu]\\
&\quad=\frac{\int_E\int_E\int_E(1+M^{-1}\sigma(x,y))Q_Mg(z)\eta_M(x,y,\mathrm{d}z)\mu(\mathrm{d}x)\mu(\mathrm{d}y)}{\int_E\int_E(1+M^{-1}\sigma(x,y))\mu(\mathrm{d}x)\mu(\mathrm{d}y)}\\
&\quad=\langle g,\mu\rangle+M^{-1}\{\langle B_Mg,\mu\rangle+\alpha(\langle H_{11}g,\mu^2\rangle-\langle g,\mu\rangle)+\bar\sigma\langle K_{11}g,\mu^3\rangle\}+\mathrm{O}(M^{-2})
\end{aligned} \tag{3.10}$$

for each $g\in B(E)$, and it follows as in the proof of Theorem 10.4.1 of Ethier and Kurtz (1986) that

$$\begin{aligned}
&E[\langle f,Z_M(1/M)^m\rangle|Z_M(0)=\mu]\\
&\quad=\sum_{1\le i<j\le m}M^{-m}\frac{M!}{(M-m+1)!}E[\Phi_{ij}f(Y_1,\ldots,Y_{m-1})|Z_M(0)=\mu]\\
&\qquad+\mathrm{O}(M^{-2})+M^{-m}\frac{M!}{(M-m)!}E[f(Y_1,\ldots,Y_m)|Z_M(0)=\mu]\,,
\end{aligned} \tag{3.11}$$

and hence that

$$\begin{aligned}
\mathscr{A}_MF(\mu)=&\sum_{1\le i<j\le m}(\langle\Phi_{ij}f,\mu^{m-1}\rangle-\langle f,\mu^m\rangle)+\langle B_Mf,\mu^m\rangle\\
&+\alpha\sum_{i=1}^{m}(\langle H_{im}f,\mu^{m+1}\rangle-\langle f,\mu^m\rangle)\\
&+\bar\sigma\sum_{i=1}^{m}\langle K_{im}f,\mu^{m+2}\rangle+\mathrm{O}(M^{-1})\,,
\end{aligned} \tag{3.12}$$

where B_M, like B below (1.7), is extended from $B(E)$ to $B(E^m)$. Noting that $\sup_M\sup_{\mu\in\mathscr{P}_M(E)}|\mathscr{A}_MF(\mu)|<\infty$ for each $f\in B(E^m)$, we have relative compactness of $\{Z_M\}$ in $D_{(\mathscr{P}(E),\rho)}[0,\infty)$ by Theorems 3.9.1 and 3.9.4 of Ethier and Kurtz (1986).

To complete the proof of the theorem we need only verify (2.21). Fix $\varepsilon>0$, and let

$$f_\varepsilon(x,y)=\Psi\left(\frac{r(x,y)}{\varepsilon}\right)-I_{\{0\}}(r(x,y))\,. \tag{3.13}$$

Then, setting $\underline{\theta}=\inf_M\inf_x\theta_M(x)$, we have for $F_\varepsilon(\mu)=\langle f_\varepsilon,\mu^2\rangle$,

$$\begin{aligned}
\mathscr{A}_MF_\varepsilon(\mu)\le\,&2\int_E\int_E\theta_M(x)\lambda_M(x,S_\varepsilon(y))\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\
&+2\alpha\int_E\int_E\int_E\eta(x,y,S_\varepsilon(z))\mu(\mathrm{d}x)\mu(\mathrm{d}y)\mu(\mathrm{d}z)\\
&+(2\bar\sigma-1-2\underline{\theta}-2\alpha)F_\varepsilon(\mu)+\mathrm{O}(M^{-1})\,.
\end{aligned} \tag{3.14}$$

Let $c > 2\bar{\sigma} - 1 - 2\underline{\theta} - 2\alpha$ and note that for $M > |c|$

$$U_M(t) \equiv \left(\frac{M}{M+c}\right)^{[Mt]} \langle f_\varepsilon, Z_M(t)^2\rangle - \int_0^{[Mt]} \left(\frac{M}{M+c}\right)^{[Ms]+1} (\mathscr{A}_M F_\varepsilon(Z_M(s)) - c\langle f_\varepsilon, Z_M(s)^2\rangle)\,\mathrm{d}s \tag{3.15}$$

is a martingale. Given $\delta > 0$, by (3.5) and (3.6), ε can be selected sufficiently small and M, depending on ε, sufficiently large so that $\mathscr{A}_M F_\varepsilon(Z_M(s)) - c\langle f_\varepsilon, Z_M(s)^2\rangle \le \delta^2$. It follows that for fixed ε and M sufficiently large,

$$V_M(t) \equiv U_M(t) + \delta^2 c^{-1}\left(1 - \left(\frac{M}{M+c}\right)^{[Mt]}\right) \tag{3.16}$$

is a nonnegative submartingale that bounds $\mathrm{e}^{-(c\vee 0)t}\langle f_\varepsilon, Z_M(t)^2\rangle$, and hence

$$P\left\{\sup_{t\le T} \langle f_\varepsilon, Z_M(t)^2\rangle \ge \delta e^{(c\vee 0)T}\right\} \le P\left\{\sup_{t\le T} V_M(t) \ge \delta\right\}$$

$$\le \delta^{-1}(E[\langle f_\varepsilon, Z_M(0)^2\rangle] + \delta^2 T e^{-(c\wedge 0)T}). \tag{3.17}$$

Letting $M \to \infty$ and then $\varepsilon \to 0$, we see that (3.17) implies (2.21). □

The proof of the following is essentially the same as the proof of Theorem 3.2.

Theorem 3.4. *For each $n \ge 1$, let Z_n be a Fleming–Viot process with type space E, mutation operator B_n, recombination determined by α_n and η_n, and selection intensity function σ_n, and assume that $\{Z_n\}$ is relatively compact in $D_{(\mathscr{P}(E),\rho)}[0,\infty)$. Suppose that for each $n \ge 1$,*

$$B_n f(x) = \theta_n(x) \int_E (f(y) - f(x))\lambda_n(x, \mathrm{d}y), \tag{3.18}$$

where $\theta_n \in B(E)$ is nonnegative and λ_n is a transition function on E, and that

$$\sup_n \sup_x \theta_n(x) < \infty, \tag{3.19}$$

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_{x,y} \theta_n(x)\lambda_n(x, S_\varepsilon(y)) = 0, \tag{3.20}$$

$$\lim_{\varepsilon\to 0} \limsup_{n\to\infty} \sup_{x,y,z} \alpha_n\eta_n(x, y, S_\varepsilon(z)) = 0, \tag{3.21}$$

and

$$\sup_n \sup_{x,y} |\sigma_n(x, y)| < \infty. \tag{3.22}$$

If in addition $\{Z_n(0)\}$ is relatively compact in $(\mathscr{P}(E), \rho_a)$, then $\{Z_n\}$ is relatively compact in $D_{(\mathscr{P}(E),\rho_a)}[0,\infty)$.

The estimates on $\mathscr{A}_M$ given above also yield the corresponding results for stationary distributions.

Theorem 3.5. *Under the assumptions of Theorem 3.2 (namely, (3.3)–(3.7)), let $\{\Pi_M\}$ be a corresponding sequence of stationary distributions and assumes that it is relatively compact in the weak topology on $\mathscr{P}(\mathscr{P}(E),\rho)$. If $2\bar{\sigma} < 1+2\underline{\theta}+2\alpha$, then $\{\Pi_M\}$ is relatively compact in the weak topology on $\mathscr{P}(\mathscr{P}(E),\rho_{\mathrm{a}})$.*

Remark 3.6. Of course if E is compact, then $(\mathscr{P}(E),\rho)$ is compact and any collection of distributions on $(\mathscr{P}(E),\rho)$ is relatively compact in the weak topology. More generally, define $\pi_M \in \mathscr{P}(E)$ by $\pi_M(\Gamma) = \int_E \mu(\Gamma)\Pi_M(\mathrm{d}\mu)$. If $\{\pi_M\}$ is relatively compact in $(\mathscr{P}(E),\rho)$, then $\{\Pi_M\}$ is relatively compact in the weak topology on $\mathscr{P}(\mathscr{P}(E),\rho)$. If $\sigma = 0$ and $\alpha = 0$, then π_M is a stationary distribution for the mutation process, i.e.,

$$\int_E B_M f(x)\pi_M(\mathrm{d}x) = 0, \quad f \in B(E). \tag{3.23}$$

Consequently, in this case, relative compactness for the stationary distributions for the mutation processes implies relative compactness (using ρ) for the stationary distributions for the Fleming–Viot processes.

Proof of Theorem 3.5. Let F_ε be as in the proof of Theorem 3.2. Then

$$\begin{aligned}0 &= \int_{\mathscr{P}(E)} \mathscr{A}_M F_\varepsilon(\mu)\Pi_M(\mathrm{d}\mu)\\ &\le \int_{\mathscr{P}(E)} \Big(2\int_E\int_E \theta_M(x)\lambda_M(x,S_\varepsilon(y))\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &\quad +2\alpha\int_E\int_E\int_E \eta(x,y,S_\varepsilon(z))\mu(\mathrm{d}x)\mu(\mathrm{d}y)\mu(\mathrm{d}z)\\ &\quad -(1+2\underline{\theta}+2\alpha-2\bar{\sigma})F_\varepsilon(\mu)\Big)\,\Pi_M(\mathrm{d}\mu) + \mathrm{O}(M^{-1}),\end{aligned} \tag{3.24}$$

and hence for $\delta > 0$, there exist ε_0 and M_ε such that $\varepsilon < \varepsilon_0$ and $M > M_\varepsilon$ imply

$$\begin{aligned}&(1+2\underline{\theta}+2\alpha-2\bar{\sigma})\int_{\mathscr{P}(E)} F_\varepsilon(\mu)\Pi_M(\mathrm{d}\mu)\\ &\le \int_{\mathscr{P}(E)} \Big(2\int_E\int_E \theta_M(x)\lambda_M(x,S_\varepsilon(y))\mu(\mathrm{d}x)\mu(\mathrm{d}y)\\ &\quad +2\alpha\int_E\int_E\int_E \eta(x,y,S_\varepsilon(z))\mu(\mathrm{d}x)\mu(\mathrm{d}y)\mu(\mathrm{d}z)\Big)\,\Pi_M(\mathrm{d}\mu) + \mathrm{O}(M^{-1})\\ &\le \delta^2.\end{aligned} \tag{3.25}$$

Letting Γ_M have distribution Π_M, then

$$\begin{aligned}&\limsup_{M\to\infty} P\left\{\int_E\int_E \Psi\left(\frac{r(x,y)}{\varepsilon}\right)\Gamma_M(\mathrm{d}x)\Gamma_M(\mathrm{d}y) - \Gamma_M^*(E) \ge \delta\right\}\\ &\le \frac{\delta}{1+2\underline{\theta}+2\alpha-2\bar{\sigma}},\end{aligned} \tag{3.26}$$

and the relative compactness of $\{\Pi_M\}$ follows by Lemma 2.9. □

The proof of the following theorem is essentially the same as that of Theorem 3.5.

Theorem 3.7. *Let $\{\theta_n\}$, $\{\lambda_n\}$, $\{\alpha_n\}$, $\{\eta_n\}$, and $\{\sigma_n\}$ be as in Theorem 3.4, and suppose that $\sup_n \alpha_n < \infty$. Let $\{\Pi_n\}$ be a corresponding sequence of stationary distributions and assume that it is relatively compact in the weak topology on $\mathscr{P}(\mathscr{P}(E),\rho)$. If $2\bar{\sigma} < 1 + 2\underline{\theta} + 2\underline{\alpha}$, where $\underline{\alpha} = \inf_n \alpha_n$, then $\{\Pi_n\}$ is relatively compact in the weak topology on $\mathscr{P}(\mathscr{P}(E),\rho_a)$.*

Remark 3.8. In Theorems 1.1, 3.2–3.5, and 3.7, the fixed type space E can be replaced by a sequence of type spaces $\{E_n\}$ (or $\{E_M\}$) with $E_n \subset E$ asymptotically dense in E.

Finally, we note that the conditions of the previous theorem imply a strengthening of the usual ergodic theorem for Fleming–Viot processes.

Theorem 3.9. *Let Z be a Fleming–Viot process with mutation operator*

$$Bf(x) = \theta(x)\int_E (f(y) - f(x))\lambda(x,\mathrm{d}y)\,, \tag{3.27}$$

where $\theta \in B(E)$ is nonnegative and λ is a transition function on E, recombination is determined by α and η, and the selection intensity function is $\sigma \in B_{\mathrm{sym}}(E\times E)$. Suppose

$$\lim_{\varepsilon\to 0}\sup_{x,y}\theta(x)\lambda(x,S_\varepsilon(y)) = 0 \tag{3.28}$$

and

$$\lim_{\varepsilon\to 0}\sup_{x,y,z}\eta(x,y,S_\varepsilon(z)) = 0\,. \tag{3.29}$$

Suppose further that $Z(t)$ converges in distribution in $(\mathscr{P}(E),\rho)$ as $t\to\infty$. If $2\bar{\sigma} < 1 + 2\underline{\theta} + 2\alpha$, then $Z(t)$ converges in distribution in $(\mathscr{P}(E),\rho_a)$.

Remark 3.10. Note that if E is compact and $\lambda(x,\cdot)$ is nonatomic for all x and weakly continuous in x, then (3.28) holds.

Proof of Theorem 3.9. Let $0 > c > 2\bar{\sigma} - 1 - 2\underline{\theta} - 2\alpha$. Then, arguing as in the proof of Theorem 3.2, we have for ε sufficiently small,

$$E[\langle f_\varepsilon, Z(t)^2\rangle e^{-ct}] \le E[\langle f_\varepsilon, Z(0)^2\rangle] + \delta^2\frac{1-e^{-ct}}{c}\,, \quad t\ge 0, \tag{3.30}$$

and hence

$$\sup_t E[\langle f_\varepsilon, Z(t)^2\rangle] \le E[\langle f_\varepsilon, Z(0)^2\rangle] + \frac{\delta^2}{|c|}\,, \tag{3.31}$$

which implies (2.14). □

Example 3.11. Let $E = [0,1]\times[0,1]$, $\alpha = 0$, $\sigma = 0$, and

$$Bf(x,y) = \int_0^1 (f(z,y) - f(x,y))\,\mathrm{d}z + \frac{1}{y}\int_0^y (f(x,z) - f(x,y))\,\mathrm{d}z\,, \tag{3.32}$$

where the second term is taken to be zero for $y = 0$. Then B is a bounded operator and is the generator of a Feller process on a compact state space. Ergodicity in $(\mathscr{P}(E), \rho)$ follows by duality (see, for example, Ethier and Kurtz, 1994), since the unique stationary distribution for B is given by linear Lebesgue measure on $[0,1] \times \{0\}$. Ergodicity does not hold in $(\mathscr{P}(E), \rho_a)$. In fact, if the initial distribution of Z is supported on $[0,1] \times (0,1]$, then the sequence of descending order statistics of the sizes of the atoms of $Z(t)$ converges in distribution to the Poisson–Dirichlet distribution with parameter 4, while the stationary distribution has atoms giving a Poisson–Dirichlet distribution with parameter 2.

Note that the jump distribution giving B is nonatomic, but that it does not satisfy (3.28). The jump intensity is, however, discontinuous.

4. Kingman's Poisson–Dirichlet limit with selection

We begin with two lemmas that rely on the symmetry-preserving transformation of Fukushima and Stroock (1986); however, in the case of the second lemma, the hypotheses in the latter paper are not satisfied, so we first isolate the identity of Fukushima and Stroock that we need here.

Let L be a linear operator on $B(E)$ whose domain $\mathscr{D}(L)$ is an algebra that is closed under exponentiation ($h \in \mathscr{D}(L)$ implies $e^h \in \mathscr{D}(L)$). Define

$$[f,h]_L = L(fh) - fLh - hLf\,, \quad f,h \in \mathscr{D}(L)\,. \tag{4.1}$$

Fix $h \in \mathscr{D}(L)$. Suppose

$$[f,e^h]_L = [f,h]_L e^h\,, \quad f \in \mathscr{D}(L)\,, \tag{4.2}$$

and define the linear operator L_h on $B(E)$ by

$$L_h f = Lf + [f,h]_L, \quad \mathscr{D}(L_h) = \mathscr{D}(L)\,. \tag{4.3}$$

Then it is immediate that

$$(fL_h g - gL_h f)e^{2h} = fe^h L(ge^h) - ge^h L(fe^h), \quad f,g \in \mathscr{D}(L_h)\,. \tag{4.4}$$

In particular, if $\mu \in \mathscr{P}(E)$ satisfies

$$\int_E fLg \,\mathrm{d}\mu = \int_E gLf \,\mathrm{d}\mu, \quad f,g \in \mathscr{D}(L)\,, \tag{4.5}$$

then

$$\int_E fL_h g \,\mathrm{d}\mu_h = \int_E gL_h f \,\mathrm{d}\mu_h, \quad f,g \in \mathscr{D}(L_h)\,, \tag{4.6}$$

where

$$\mathrm{d}\mu_h = e^{2h}\,\mathrm{d}\mu \Big/ \int_E e^{2h}\,\mathrm{d}\mu\,. \tag{4.7}$$

In other words, assuming (4.2), if L is symmetric with respect to μ, then L_h is symmetric with respect to μ_h.

For $n \geq 2$ define

$$\Delta_n = \left\{ x = (x_1, \ldots, x_n) : x_1 \geq 0, \ldots, x_n \geq 0, \sum_{i=1}^{n} x_i = 1 \right\} . \tag{4.8}$$

The following lemma gives Wright's formula (Wright, 1949; Watterson, 1977) (cf. Problem 10.2 of Ethier and Kurtz, 1986).

Lemma 4.1. *Fix $n \geq 2$. Let $\gamma_1 > 0, \ldots, \gamma_n > 0$ and let (σ_{ij}) be a real, symmetric, $n \times n$ matrix. Then the distribution $\mu \in \mathscr{P}(\Delta_n)$, defined for the appropriate constant C by*

$$\mu(\mathrm{d}x) = C x_1^{\gamma_1 - 1} \cdots x_n^{\gamma_n - 1} \exp\left\{ \sum_{i,j=1}^{n} \sigma_{ij} x_i x_j \right\} \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} , \tag{4.9}$$

is the unique stationary distribution for the diffusion in Δ_n with generator

$$A = \frac{1}{2} \sum_{i,j=1}^{n} x_i(\delta_{ij} - x_j) \frac{\partial^2}{\partial x_i \partial x_j} + \sum_{i=1}^{n} b_i(x) \frac{\partial}{\partial x_i}, \quad \mathscr{D}(A) = C^2(\Delta_n) , \tag{4.10}$$

where

$$b_i(x) = \frac{1}{2} \{ \gamma_i - (\gamma_1 + \cdots + \gamma_n) x_i \} + x_i \left(\sum_{j=1}^{n} \sigma_{ij} x_j - \sum_{k,l=1}^{n} \sigma_{kl} x_k x_l \right) . \tag{4.11}$$

Moreover, μ is reversible.

Proof. The neutral case ($\sigma_{ij} = 0$ for $i, j = 1, \ldots, n$) is Lemma 4.1 and Remark 4.2(b) of Ethier and Kurtz (1981). Denote the generator A in that case by A^0. Then

$$[f, h]_{A^0}(x) = \sum_{i,j=1}^{n} f_{x_i}(x) x_i (\delta_{ij} - x_j) h_{x_j}(x), \quad f, h \in \mathscr{D}(A^0) , \tag{4.12}$$

so (4.2) holds. If $h(x) = \frac{1}{2} \sum_{i,j=1}^{n} \sigma_{ij} x_i x_j$, then using the symmetry of (σ_{ij}), we have $[f, h]_{A^0} = Af - A^0 f$. Letting $\mu^0 \in \mathscr{P}(\Delta_n)$ denote the Dirichlet distribution

$$\mu^0(\mathrm{d}x) = \frac{\Gamma(\gamma_1 + \cdots + \gamma_n)}{\Gamma(\gamma_1) \cdots \Gamma(\gamma_n)} x_1^{\gamma_1 - 1} \cdots x_n^{\gamma_n - 1} \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} , \tag{4.13}$$

we know that

$$\int_{\Delta_n} f A^0 g \, \mathrm{d}\mu^0 = \int_{\Delta_n} g A^0 f \, \mathrm{d}\mu^0, \quad f, g \in \mathscr{D}(A^0) . \tag{4.14}$$

Therefore, by (4.6),

$$\int_{\Delta_n} f A g \, \mathrm{d}\mu = \int_{\Delta_n} g A f \, \mathrm{d}\mu, \quad f, g \in \mathscr{D}(A) . \tag{4.15}$$

Since the closure of A generates a Feller semigroup on $C(\Delta_n)$ (Ethier, 1976), (4.15) implies the reversibility (hence stationarity) of μ. The uniqueness of stationary distributions is a result of Shiga (1981). □

Lemma 4.2. *Let E be compact, let $\theta > 0$ and $\nu \in \mathscr{P}(E)$, and define the linear operator B on $B(E)$ by $Bf = \frac{1}{2}\theta(\langle f,\nu\rangle - f)$.*

(a) *Let $(\xi_1, \xi_2, \ldots)$ have the Poisson–Dirichlet distribution with parameter θ, and let $V_1, V_2, \ldots$ be a sequence of E-valued random variables that are i.i.d. ν and independent of $(\xi_1, \xi_2, \ldots)$. Define $\Pi^0 \in \mathscr{P}(\mathscr{P}(E))$ to be the distribution of the random measure $\mu = \sum_{i=1}^{\infty} \xi_i \delta_{V_i}$. Then Π^0 is the unique stationary distribution for the neutral Fleming–Viot process with type space E and mutation operator B, and Π^0 is reversible.*

(b) *Let $\sigma \in B_{\mathrm{sym}}(E^2)$. Then $\Pi \in \mathscr{P}(\mathscr{P}(E))$, defined for the appropriate constant C by*

$$\Pi(\mathrm{d}\mu) = Ce^{\langle \sigma, \mu^2 \rangle} \Pi^0(\mathrm{d}\mu), \tag{4.16}$$

is the unique stationary distribution for the Fleming–Viot process with type space E, mutation operator B, and selection intensity function σ. Moreover, Π is reversible; in fact, if $\mathscr{A}$ is the corresponding generator, then Π is the unique $\Gamma \in \mathscr{P}(\mathscr{P}(E))$ such that

$$\int_{\mathscr{P}(E)} \varphi \mathscr{A} \psi \, \mathrm{d}\Gamma = \int_{\mathscr{P}(E)} \psi \mathscr{A} \varphi \, \mathrm{d}\Gamma, \quad \varphi, \psi \in \mathscr{D}(\mathscr{A}). \tag{4.17}$$

Remark 4.3. We cannot prove that Π is the unique $\Gamma \in \mathscr{P}(\mathscr{P}(E))$ such that

$$\int_{\mathscr{P}(E)} \mathscr{A} \varphi \, \mathrm{d}\Gamma = 0, \quad \varphi \in \mathscr{D}(\mathscr{A}). \tag{4.18}$$

But if σ were continuous, then this would follow from Echevarria's theorem (see, for example, Ethier and Kurtz, 1986, Theorem 4.9.17).

Proof of Lemma 4.2. (a) For $1 \le d \le n$, let $\pi(n,d)$ be the collection of partitions β of $\{1, \ldots, n\}$ into d nonempty subsets $\beta_1, \ldots, \beta_d$ labeled so that $\min \beta_1 < \cdots < \min \beta_d$, and for $\beta \in \pi(n,d)$ define

$$p(\beta) = (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \frac{\theta^{d-1}}{(1+\theta)\cdots(n-1+\theta)}. \tag{4.19}$$

Then, for each $n \ge 1$ and $f_1, \ldots, f_n \in C(E)$,

$$\begin{aligned} &\int_{\mathscr{P}(E)} \langle f_1, \mu \rangle \cdots \langle f_n, \mu \rangle \Pi^0(\mathrm{d}\mu) \\ &= E\left[\prod_{i=1}^{n} \left\langle f_i, \sum_{j=1}^{\infty} \xi_j \delta_{V_j} \right\rangle\right] \end{aligned}$$

$$
\begin{aligned}
&= E\left[\prod_{i=1}^{n}\sum_{j=1}^{\infty}\xi_j f_i(V_j)\right] \\
&= E\left[\sum_{(j_1,\ldots,j_n)\in N^n}\xi_{j_1}\cdots\xi_{j_n} f_1(V_{j_1})\cdots f_n(V_{j_n})\right] \\
&= \sum_{d=1}^{n}\sum_{\beta\in\pi(n,d)}\sum_{(j_1,\ldots,j_n)\sim\beta} E[\xi_{j_1}\cdots\xi_{j_n}]\prod_{k=1}^{d}\left\langle\prod_{i\in\beta_k} f_i, v\right\rangle \\
&= \sum_{d=1}^{n}\sum_{\beta\in\pi(n,d)} p(\beta)\prod_{k=1}^{d}\left\langle\prod_{i\in\beta_k} f_i, v\right\rangle ,
\end{aligned}
\tag{4.20}
$$

where $(j_1, \ldots, j_n) \sim \beta$ means that $j_k = j_l$ if and only if k and l belong to the same β_i; the last equality follows from the relationship between the Poisson–Dirichlet distribution and the Ewens sampling formula (Kingman, 1977). It is known (Ethier, 1990) that the stationary distribution satisfies the same identity (i.e., the left-hand side of (4.20) equals the right-hand side) and does so uniquely. Hence Π^0 is the stationary distribution. (See Theorem 10.4.6 of Ethier and Kurtz (1986) for a somewhat more complicated proof in the special case $E = [0,1], v =$ Lebesgue measure.) The reversibility of the stationary distribution was proved by Ethier (1990).

(b) Let $\mathscr{A}^0$ be the generator for the neutral Fleming–Viot process. From the latter reference,

$$\int_{\mathscr{P}(E)} \varphi\mathscr{A}^0\psi \,\mathrm{d}\Pi^0 = \int_{\mathscr{P}(E)} \psi\mathscr{A}^0\varphi \,\mathrm{d}\Pi^0, \quad \varphi,\psi\in\mathscr{D}(\mathscr{A}^0). \tag{4.21}$$

As in the proof of Lemma 4.1, we now want to apply (4.6) with $h(\mu) = \frac{1}{2}\langle\sigma,\mu^2\rangle$. First, however, we need to extend $\mathscr{A}^0$ to an algebra closed under exponentiation and containing h; note that h is not necessarily continuous. Thus, for

$$\varphi(\mu) = F(\langle f_1,\mu^{n_1}\rangle, \ldots, \langle f_d,\mu^{n_d}\rangle), \tag{4.22}$$

where $d \geq 1, F \in C^2(\boldsymbol{R}^d), n_1, \ldots, n_d \geq 1$, and $f_k \in B(E^{n_k})$ for $k = 1, \ldots, d$, we define

$$
\begin{aligned}
\hat{\mathscr{A}}^0\varphi(\mu) &= \frac{1}{2}\sum_{i,j=1}^{d}\sum_{k=1}^{n_i}\sum_{l=1}^{n_j}(\langle\Psi_{kl}(f_i,f_j),\mu^{n_i+n_j-1}\rangle - \langle f_i,\mu^{n_i}\rangle\langle f_j,\mu^{n_j}\rangle)F_{x_ix_j} \\
&\quad + \sum_{i=1}^{d}\sum_{1\leq k<l\leq n_i}(\langle\Phi_{kl}f_i,\mu^{n_i-1}\rangle - \langle f_i,\mu^{n_i}\rangle)F_{x_i} \\
&\quad + \sum_{i=1}^{d}\sum_{k=1}^{n_i}\langle B_k f_i,\mu^{n_i}\rangle F_{x_i} ,
\end{aligned}
\tag{4.23}
$$

where $\Psi_{kl}(f_i,f_j)$ is the function in $B(E^{n_i+n_j-1})$ obtained from $f_i(x)f_j(y)$ by replacing y_l by x_k and renumbering the variables, $\Phi_{kl}f_i$ is as in Section 1, and $B_k f$ is B acting on f as a function of its kth variable. The partial derivatives of F have the same arguments as F itself in (4.22). We define $\mathscr{D}(\hat{\mathscr{A}}^0)$ to be the space of all such φ. Approximating F by polynomials, we can easily check that

$$\{(\varphi,\hat{\mathscr{A}}^0\varphi) : \varphi\in\mathscr{D}(\hat{\mathscr{A}}^0)\} \subset \text{bp-closure}\{(\varphi,\mathscr{A}^0\varphi) : \varphi\in\mathscr{D}(\mathscr{A}^0)\} . \tag{4.24}$$

It follows from (4.21) that

$$\int_{\mathscr{P}(E)} \varphi \hat{\mathscr{A}}^0 \psi \, \mathrm{d}\Pi^0 = \int_{\mathscr{P}(E)} \psi \hat{\mathscr{A}}^0 \varphi \, \mathrm{d}\Pi^0 , \quad \varphi, \psi \in \mathscr{D}(\hat{\mathscr{A}}^0) . \tag{4.25}$$

By (4.23) (using the notation (4.1)),

$$\begin{aligned} &[F(\langle f_1, \mu^{n_1}\rangle, \ldots, \langle f_d, \mu^{n_d}\rangle), G(\langle g_1, \mu^{m_1}\rangle, \ldots, \langle g_c, \mu^{m_c}\rangle)]_{\hat{\mathscr{A}}^0} \\ &\quad = \sum_{i=1}^{d} \sum_{j=1}^{c} \sum_{k=1}^{n_i} \sum_{l=1}^{m_j} (\langle \Psi_{kl}(f_i, g_j), \mu^{n_i+m_j-1}\rangle - \langle f_i, \mu^{n_i}\rangle \langle g_j, \mu^{m_j}\rangle) F_{x_i} G_{y_j} , \end{aligned} \tag{4.26}$$

hence (4.2) holds. Let $\varphi(\mu) = \langle f, \mu^n\rangle$, where $n \geq 1$ and $f \in B(E^n)$, and let $h(\mu) = \frac{1}{2}\langle \sigma, \mu^2\rangle$. Then, by (4.26) and the symmetry of $\sigma, [\varphi, h]_{\hat{\mathscr{A}}^0} = \mathscr{A}\varphi - \mathscr{A}^0\varphi$, so by (4.6) we have

$$\int_{\mathscr{P}(E)} \varphi \mathscr{A} \psi \, \mathrm{d}\Pi = \int_{\mathscr{P}(E)} \psi \mathscr{A} \varphi \, \mathrm{d}\Pi , \quad \varphi, \psi \in \mathscr{D}(\mathscr{A}) . \tag{4.27}$$

We claim that Π is reversible. For continuous σ, this follows from Fukushima and Stroock (1986). To see that it holds in general, let $\{T_\sigma(t)\}$ be the Markov semigroup on $B(\mathscr{P}(E))$ corresponding to the Fleming–Viot process with type space E, mutation operator B, and selection intensity function σ, and temporarily denote Π by Π_σ. Let

$$\Sigma = \{\sigma \in B_{\text{sym}}(E^2) : \Pi_\sigma \text{ is reversible}\} . \tag{4.28}$$

If $\{\sigma_n\} \subset \Sigma$ and bp-$\lim_{n\to\infty} \sigma_n = \sigma_\infty$ exists, then

$$\begin{aligned} \int_{\mathscr{P}(E)} \varphi T_{\sigma_\infty}(t) \psi \, \mathrm{d}\Pi_{\sigma_\infty} &= \lim_{n\to\infty} \int_{\mathscr{P}(E)} \varphi T_{\sigma_n}(t) \psi \, \mathrm{d}\Pi_{\sigma_n} \\ &= \lim_{n\to\infty} \int_{\mathscr{P}(E)} \psi T_{\sigma_n}(t) \varphi \, \mathrm{d}\Pi_{\sigma_n} \\ &= \int_{\mathscr{P}(E)} \psi T_{\sigma_\infty}(t) \varphi \, \mathrm{d}\Pi_{\sigma_\infty} \end{aligned} \tag{4.29}$$

for all $\varphi, \psi \in C(\mathscr{P}(E))$ and $t \geq 0$, where the first and third equalities follow by coupling the dual processes as in Theorem 3.1(a) (especially, Eq. (3.3)) of Ethier and Kurtz (1987) and by (4.16). We conclude that $\sigma_\infty \in \Sigma$, so Σ is bp-closed. As noted above, Σ contains the continuous symmetric functions, hence $\Sigma = B_{\text{sym}}(E^2)$. We conclude that Π (given by (4.16)) is reversible, hence stationary. Uniqueness of stationary distributions is a consequence of results in Ethier and Kurtz (1994).

Finally, suppose $\Gamma \in \mathscr{P}(\mathscr{P}(E))$ satisfies (4.17). We begin by extending $\mathscr{A}$ as in (4.23). For φ as in (4.22) define

$$\hat{\mathscr{A}}\varphi(\mu) = \hat{\mathscr{A}}^0\varphi(\mu) + \bar{\sigma} \sum_{i=1}^{d} \sum_{k=1}^{n_i} \langle K_{kn_i} f_i, \mu^{n_i+2}\rangle F_{x_i} , \tag{4.30}$$

and let $\mathscr{D}(\hat{\mathscr{A}}) = \mathscr{D}(\hat{\mathscr{A}}^0)$. Then, by (4.17),

$$\int_{\mathscr{P}(E)} \varphi \hat{\mathscr{A}} \psi \, \mathrm{d}\Gamma = \int_{\mathscr{P}(E)} \psi \hat{\mathscr{A}} \varphi \, \mathrm{d}\Gamma , \quad \varphi, \psi \in \mathscr{D}(\hat{\mathscr{A}}) . \tag{4.31}$$

Noting that (4.26) holds with $\hat{\mathscr{A}}$ in place of $\hat{\mathscr{A}}^0$, we again apply (4.6), now with $h(\mu) = -\frac{1}{2}\langle\sigma, \mu^2\rangle$, to conclude that

$$\int_{\mathscr{P}(E)} \varphi(\mu)\mathscr{A}^0\psi(\mu)e^{-\langle\sigma,\mu^2\rangle}\Gamma(\mathrm{d}\mu)$$
$$= \int_{\mathscr{P}(E)} \psi(\mu)\mathscr{A}^0\varphi(\mu)e^{-\langle\sigma,\mu^2\rangle}\Gamma(\mathrm{d}\mu)\,, \quad \varphi, \psi \in \mathscr{D}(\mathscr{A}^0)\,. \tag{4.32}$$

Take $\psi \equiv 1$ and apply Echeverria's theorem to conclude that $C^{-1}e^{-\langle\sigma,\mu^2\rangle}\Gamma(\mathrm{d}\mu) = \Pi^0(\mathrm{d}\mu)$ for some $C > 0$, hence $\Gamma = \Pi$. □

Let

$$\nabla_\infty = \left\{(x_1, x_2, \ldots) : x_1 \geq x_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} x_i = 1\right\} \tag{4.33}$$

be the infinite-dimensional ordered simplex (topologized as a subset of $[0,1]^N$), and define $\zeta_n : \Delta_n \mapsto \nabla_\infty$ for each $n \geq 2$ in terms of the descending order statistics $x_{(1)} \geq x_{(2)} \geq \cdots \geq x_{(n)}$ of the coordinates $x_1, \ldots, x_n$ of vectors $x \in \Delta_n$:

$$\zeta_n(x) = (x_{(1)}, \ldots, x_{(n)}, 0, 0, \ldots)\,. \tag{4.34}$$

We can now state the main result of this section.

Theorem 4.4. *For each $n \geq 2$, let $\gamma_1^{(n)} > 0, \ldots, \gamma_n^{(n)} > 0$, and let $(\sigma_{ij}^{(n)})$ be a real, symmetric, $n \times n$ matrix. Let $\theta > 0$ and assume*

$$\lim_{n\to\infty} \sum_{j=1}^{n} \gamma_j^{(n)} = \theta, \quad \lim_{n\to\infty} \max_{1\leq j\leq n} \gamma_j^{(n)} = 0\,. \tag{4.35}$$

Define

$$\beta_i^{(n)} = \frac{\gamma_1^{(n)} + \cdots + \gamma_i^{(n)}}{\gamma_1^{(n)} + \cdots + \gamma_n^{(n)}}, \quad i = 1, \ldots, n, \quad n \geq 2\,, \tag{4.36}$$

and suppose there exists $\sigma \in B_{\mathrm{sym}}([0,1]^2)$ such that σ is continuous λ^2-a.e. (λ = Lebesgue measure on $[0,1]$), the function $\sigma_0 \in B([0,1])$ given by $\sigma_0(x) = \sigma(x,x)$ is continuous λ-a.e., and

$$\lim_{n\to\infty} \max_{1\leq i,j\leq n} |\sigma_{ij}^{(n)} - \sigma(\beta_i^{(n)}, \beta_j^{(n)})| = 0\,. \tag{4.37}$$

For each $n \geq 2$ define $\mu_n \in \mathscr{P}(\Delta_n)$ for the appropriate constant C_n by

$$\mu_n(\mathrm{d}x) = C_n x_1^{\gamma_1^{(n)}-1} \cdots x_n^{\gamma_n^{(n)}-1} \exp\left\{\sum_{i,j=1}^{n} \sigma_{ij}^{(n)} x_i x_j\right\} \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1}\,. \tag{4.38}$$

Then there exists $\mu_\infty \in \mathscr{P}(\nabla_\infty)$ such that $\mu_n\zeta_n^{-1} \Rightarrow \mu_\infty$ on ∇_∞. In fact,

$$\mu_\infty(\Lambda) = \frac{E[I_\Lambda(\xi_1, \xi_2, \ldots)\exp\{\sum_{i,j=1}^{\infty}\sigma(U_i, U_j)\xi_i\xi_j\}]}{E[\exp\{\sum_{i,j=1}^{\infty}\sigma(U_i, U_j)\xi_i\xi_j\}]} \tag{4.39}$$

for all $\Lambda \in \mathscr{B}(\nabla_\infty)$, *where the* ∇_∞*-valued random variable* $(\xi_1, \xi_2, \ldots)$ *is Poisson–Dirichlet with parameter* θ, *and* $U_1, U_2, \ldots$ *are independent uniform* $[0,1]$ *random variables, independent of* $(\xi_1, \xi_2, \ldots)$.

Proof. Theorem 3.1(b) of Ethier and Kurtz (1987) is a limit theorem for a sequence of Fleming–Viot processes with different type spaces (cf. Remark 3.8 above). Here we need a uniform estimate on the rate of convergence of their generators.

For each n, let E_n and E be compact metric spaces; let $\eta_n : E_n \mapsto E$ be Borel measurable and define $\hat{\eta}_n : \mathscr{P}(E_n) \mapsto \mathscr{P}(E)$ by $\hat{\eta}_n(\mu) = \mu\eta_n^{-1}$; let B_n and B generate Feller semigroups on $C(E_n)$ and $C(E)$; let $\sigma_n \in B_{\mathrm{sym}}(E_n^2)$ and $\sigma \in B_{\mathrm{sym}}(E^2)$; and let $\mathscr{A}_n$ and $\mathscr{A}$ be the generators of the Fleming–Viot processes associated with E_n, B_n, σ_n and E, B, σ, respectively. Assume that if $f \in \mathscr{D}(B)$, then $f \circ \eta_n \in \mathscr{D}(B_n)$. Fix $m \geq 1$ and $f_1, \ldots, f_m \in \mathscr{D}(B)$, and define $\varphi \in \mathscr{D}(\mathscr{A})$ by $\varphi(\mu) = \langle f_1, \mu\rangle \cdots \langle f_m, \mu\rangle$. Then it is easily verified that $\varphi \circ \hat{\eta}_n \in \mathscr{D}(\mathscr{A}_n)$ and

$$\begin{aligned} \|\mathscr{A}_n(\varphi \circ \hat{\eta}_n) - (\mathscr{A}\varphi) \circ \hat{\eta}_n\| &\leq \sum_{i=1}^m \|B_n(f_i \circ \eta_n) - (Bf_i) \circ \eta_n\| \prod_{l:l\neq i} \|f_l\| \\ &+2m \sup_{(x,y)\in E_n^2} |\sigma_n(x,y) - \sigma(\eta_n(x), \eta_n(y))| \prod_{l=1}^m \|f_l\| \end{aligned} \tag{4.40}$$

for each n.

In what follows we take

$$E_n = \{\beta_1^{(n)}, \ldots, \beta_n^{(n)}\}, \quad E = [0,1], \tag{4.41}$$

$$B_n f = \tfrac{1}{2}\theta_n(\langle f, \lambda_n\rangle - f), \quad Bf = \tfrac{1}{2}\theta(\langle f, \lambda\rangle - f), \tag{4.42}$$

where

$$\theta_n = \sum_{j=1}^n \gamma_j^{(n)}, \qquad \lambda_n = \sum_{i=1}^n (\gamma_i^{(n)}/\theta_n)\delta_{\beta_i^{(n)}}, \tag{4.43}$$

and $\sigma_n(\beta_i^{(n)}, \beta_j^{(n)}) = \sigma_{ij}^{(n)}$ $(i,j = 1, \ldots, n)$ and σ is as given; also, η_n is the inclusion map. Noting that

$$\begin{aligned} \|B_n(f \circ \eta_n) - (Bf) \circ \eta_n\| &\leq \tfrac{1}{2}|\theta_n\langle f \circ \eta_n, \lambda_n\rangle - \theta\langle f, \lambda\rangle| + \tfrac{1}{2}|\theta_n - \theta|\,\|f\| \\ &\leq |\theta_n - \theta|\,\|f\| + \frac{1}{2}\theta w_f\left(\max_{1\leq i\leq n} \gamma_i^{(n)}/\theta_n\right) \end{aligned} \tag{4.44}$$

for all $f \in C([0,1])$, where w_f denotes the modulus of continuity of f, we see from (4.40) that (4.35) and (4.37) imply that

$$\lim_{n\to\infty} \|\mathscr{A}_n(\varphi \circ \hat{\eta}_n) - (\mathscr{A}\varphi) \circ \hat{\eta}_n\| = 0\,. \tag{4.45}$$

For each $n \geq 2$, define $\pi_n : \Delta_n \mapsto \mathscr{P}(E_n)$ by $\pi_n(x) = \sum_{i=1}^n x_i\delta_{\beta_i^{(n)}}$, and let $\Pi_n = \mu_n\pi_n^{-1}$, where μ_n is given by (4.38). It follows from Lemma 4.1 that Π_n is the unique stationary distribution of the Fleming–Viot process with type space E_n, mutation operator B_n, and selection intensity function σ_n. Consequently, given $\varphi, \psi \in \mathscr{D}(\mathscr{A})$ of the form

$$\varphi(\mu) = \prod_{i=1}^{m} \langle f_i, \mu\rangle, \quad \psi(\mu) = \prod_{i=1}^{m} \langle g_i, \mu\rangle, \tag{4.46}$$

where $m \geq 1$ and $f_1, \ldots, f_m, g_1, \ldots, g_m \in C([0,1])$, we have

$$\int_{\mathscr{P}(E_n)} (\varphi \circ \hat{\eta}_n)\mathscr{A}_n(\psi \circ \hat{\eta}_n)\,\mathrm{d}\Pi_n - \int_{\mathscr{P}(E_n)} (\psi \circ \hat{\eta}_n)\mathscr{A}_n(\varphi \circ \hat{\eta}_n)\,\mathrm{d}\Pi_n = 0 \tag{4.47}$$

for all $n \geq 2$. Hence, by (4.45),

$$\int_{\mathscr{P}([0,1])} \varphi\mathscr{A}\psi \,\mathrm{d}\Pi_n\hat{\eta}_n^{-1} - \int_{\mathscr{P}([0,1])} \psi\mathscr{A}\varphi \,\mathrm{d}\Pi_n\hat{\eta}_n^{-1} \to 0 \tag{4.48}$$

as $n \to \infty$.

Let $\Pi^0 \in \mathscr{P}(\mathscr{P}([0,1]))$ be as in Lemma 4.2(a) (with $E = [0,1]$ and $\nu = \lambda$). We defer for the moment the proofs of the following three assertions:

(4.49) $\{\Pi_n\hat{\eta}_n^{-1}\}$ is relatively compact in $\mathscr{P}(\mathscr{P}([0,1]), \rho_a)$.

(4.50) Every subsequential limit $\Gamma \in \mathscr{P}(\mathscr{P}([0,1]))$ of $\{\Pi_n\hat{\eta}_n^{-1}\}$ satisfies $\Gamma \ll \Pi^0$ (and hence is concentrated on $\mathscr{P}_a([0,1])$, the space of purely atomic Borel probability measures on [0, 1]).

(4.51) $\mathscr{A}\varphi$ is Π^0-a.s. continuous on $(\mathscr{P}_a([0,1]), \rho_a)$.

Granting these results, the proof is easily completed. Suppose $\Pi_{n'}\hat{\eta}_{n'}^{-1} \Rightarrow \Gamma$ on $(\mathscr{P}([0,1]), \rho_a)$; then the convergence holds on $(\mathscr{P}_a([0,1]), \rho_a)$ since by (4.50) all measures are concentrated on $\mathscr{P}_a([0,1])$. By (4.51), $\varphi\mathscr{A}\psi - \psi\mathscr{A}\varphi$ is Π^0-a.s., hence by (4.50) Γ-a.s., continuous on $(\mathscr{P}_a([0,1]), \rho_a)$. It follows from (4.48) that

$$\int_{\mathscr{P}([0,1])} \varphi\mathscr{A}\psi \,\mathrm{d}\Gamma - \int_{\mathscr{P}([0,1])} \psi\mathscr{A}\varphi \,\mathrm{d}\Gamma = 0, \tag{4.52}$$

and this holds for all φ and ψ of the form (4.46). We conclude from Lemma 4.2(b) that $\Gamma = \Pi$. Hence $\Pi_n\hat{\eta}_n^{-1} \Rightarrow \Pi$ on $(\mathscr{P}_a([0,1]), \rho_a)$. For each $i \geq 1$ define s_i : $\mathscr{P}_a([0,1]) \mapsto [0,1]$ by letting $s_i(\mu)$ be the size of the ith largest atom of μ if μ has at least i atoms, and 0 otherwise. By Lemma 2.5, $(s_1, s_2, \ldots)$ is a continuous function from $(\mathscr{P}_a([0,1]), \rho_a)$ into ∇_∞. Hence $\Pi_n\hat{\eta}_n^{-1}(s_1, s_2, \ldots)^{-1} \Rightarrow \Pi(s_1, s_2, \ldots)^{-1}$ on ∇_∞. But this amounts to $\mu_n\zeta_n^{-1} \Rightarrow \mu_\infty$ on ∇_∞, where

$$\mu_\infty(A) = \Pi\{(s_1, s_2, \ldots) \in A\} = \frac{\int_{\mathscr{P}([0,1])} I_A(s_1(\mu), s_2(\mu), \ldots) e^{\langle\sigma,\mu^2\rangle}\Pi^0(\mathrm{d}\mu)}{\int_{\mathscr{P}([0,1])} e^{\langle\sigma,\mu^2\rangle}\Pi^0(\mathrm{d}\mu)} \tag{4.53}$$

for all $A \in \mathscr{B}(\nabla_\infty)$ by Lemma 4.2(b), and the right-hand side of (4.53) equals the right-hand side of (4.39) by Lemma 4.2(a).

We turn to (4.49)–(4.51). Relative compactness of $\{\Pi_n\hat{\eta}_n^{-1}\}$ in $\mathscr{P}(\mathscr{P}([0,1]), \rho)$ is automatic, so by Lemma 2.9, (4.49) will hold if for every $\delta > 0$ there exists $\varepsilon > 0$ such that

$$\inf_n \mu_n\left\{x \in \Delta_n : \sum_{i,j=1}^{n} \Psi\left(\frac{|\beta_i^{(n)} - \beta_j^{(n)}|}{\varepsilon}\right) x_i x_j - \sum_{i=1}^{n} x_i^2 \leq \delta\right\} \geq 1 - \delta, \tag{4.54}$$

or if

$$\lim_{\varepsilon\to 0}\sup_n \int_{\Delta_n}\sum_{i\neq j}\Psi\left(\frac{|\beta_i^{(n)}-\beta_j^{(n)}|}{\varepsilon}\right)x_ix_j\mu_n(\mathrm{d}x)=0\,. \tag{4.55}$$

We can replace μ_n by the Dirichlet distribution μ_n^0 (cf. (4.13)) in (4.55) since $\mu_n(\Lambda)\leq \mu_n^0(\Lambda)\exp\{\max\,\sigma_{ij}^{(n)}-\min\,\sigma_{ij}^{(n)}\}$. But then (4.55) becomes

$$\lim_{\varepsilon\to 0}\sup_n\sum_{i\neq j}\Psi\left(\frac{|\beta_i^{(n)}-\beta_j^{(n)}|}{\varepsilon}\right)\frac{\gamma_i^{(n)}\gamma_j^{(n)}}{\theta_n(\theta_n+1)}=0\,, \tag{4.56}$$

which by Lemma 2.4 holds if and only if $\{\lambda_n\}$ is relatively compact in $(\mathscr{P}([0,1]),\rho_{\mathrm{a}})$. But $\lambda_n\Rightarrow\lambda$ and

$$\lambda_n^*([0,1])=\sum_{i=1}^n(\gamma_i^{(n)}/\theta_n)^2\leq\max_{1\leq i\leq n}\gamma_i^{(n)}/\theta_n\to 0=\lambda^*([0,1])\,. \tag{4.57}$$

Hence $\lambda_n\to\lambda$ in $(\mathscr{P}([0,1]),\rho_{\mathrm{a}})$ by Lemmas 2.1 and 2.2. Thus, we have (4.49).

Next, suppose $\Pi_{n'}\hat{\eta}_{n'}^{-1}\Rightarrow\Gamma$ on $(\mathscr{P}([0,1]),\rho_{\mathrm{a}})$. Then, for every nonnegative function $\varphi\in C(\mathscr{P}([0,1]),\rho)$,

$$\begin{aligned}\int_{\mathscr{P}([0,1])}\varphi\,\mathrm{d}\Gamma&=\lim_{n'}\int_{\mathscr{P}([0,1])}\varphi\,\mathrm{d}\Pi_{n'}\hat{\eta}_{n'}^{-1}\\&\leq e^{\sup\sigma-\inf\sigma}\lim_{n'}\int_{\mathscr{P}([0,1])}\varphi\,\mathrm{d}\Pi_{n'}^0\hat{\eta}_{n'}^{-1}\\&=e^{\sup\,\sigma-\inf\,\sigma}\int_{\mathscr{P}([0,1])}\varphi\,\mathrm{d}\Pi^0\,,\end{aligned} \tag{4.58}$$

where $\Pi_n^0=\mu_n^0\pi_n^{-1}$, and hence we have (4.50).

To verify (4.51), it is sufficient to consider $\langle\sigma f,\mu^2\rangle$, where $f\in C([0,1])$. If $\mu=\sum_i\alpha_i\delta_{x_i}$, where $\alpha_1\geq\alpha_2\geq\cdots$, this is

$$\sum_{i,j=1}^{\infty}\sigma(x_i,x_j)f(x_i)\alpha_i\alpha_j=\sum_i\sigma_0(x_i)f(x_i)\alpha_i^2+\sum_{i\neq j}\sigma(x_i,x_j)f(x_i)\alpha_i\alpha_j\,. \tag{4.59}$$

By Lemma 2.5, the first term is continuous on $(\mathscr{P}_{\mathrm{a}}([0,1]),\rho_{\mathrm{a}})$ at μ if $\alpha_1>\alpha_2>\cdots$ and each x_i belongs to the continuity set of σ_0. On the other hand,

$$\begin{aligned}&\left|\sum_{i\neq j}\sigma(x_i^n,x_j^n)f(x_i^n)\alpha_i^n\alpha_j^n-\sum_{i\neq j}\sigma(x_i,x_j)f(x_i)\alpha_i\alpha_j\right|\\&\quad\leq\sum_{i\neq j}|\sigma(x_i^n,x_j^n)f(x_i^n)|\,|\alpha_i^n\alpha_j^n-\alpha_i\alpha_j|+\sum_{i\neq j}|\sigma(x_i^n,x_j^n)f(x_i^n)-\sigma(x_i,x_j)f(x_i)|\alpha_i\alpha_j\\&\quad\leq 2\|\sigma\|\,\|f\|\sum_i|\alpha_i^n-\alpha_i|+\sum_{i\neq j}|\sigma(x_i^n,x_j^n)f(x_i^n)-\sigma(x_i,x_j)f(x_i)|\alpha_i\alpha_j\,,\end{aligned} \tag{4.60}$$

so Lemma 2.5 implies that the second term on the right-hand side of (4.59) is continuous on $(\mathscr{P}_{\mathrm{a}}([0,1]),\rho_{\mathrm{a}})$ at μ if $\alpha_1>\alpha_2>\cdots$ and each pair (x_i,x_j) with $i\neq j$

belongs to the continuity set of σ. But by Lemma 4.2(a), these requirements are satisfied $\Pi^0(\mathrm{d}\mu)$-a.s. under the assumptions on σ and by properties of the Poisson–Dirichlet distribution. This establishes (4.51) and completes the proof. □

We conclude by showing that μ_∞, given by (4.39), has a simpler representation for an important class of selection intensity functions σ.

Let $\theta_1 > 0, \ldots, \theta_d > 0$, and $\theta = \theta_1 + \cdots + \theta_d$. Write

$$[0,1] = J_1 \cup \cdots \cup J_d = [0, \theta_1/\theta) \cup \cdots \cup [(\theta_1 + \cdots + \theta_{d-1})/\theta, 1], \tag{4.61}$$

and suppose that $\sigma \in B_{\mathrm{sym}}([0,1]^2)$ is given by

$$\sigma(x,y) = \sum_{k,l=1}^{d} \sigma_{kl} I_{J_k \times J_l}(x,y), \tag{4.62}$$

where (σ_{kl}) is a real, symmetric $d \times d$ matrix. Let $(\xi_1^k, \xi_2^k, \ldots)$, $k = 1, \ldots, d$, be d independent Poisson–Dirichlet random variables with parameters $\theta_1, \ldots, \theta_d$, respectively, independent of $(\gamma_1, \ldots, \gamma_d)$, a Δ_d-valued random variable whose distribution μ is defined for the appropriate constant C by

$$\mu(\mathrm{d}x) = C x_1^{\theta_1 - 1} \cdots x_d^{\theta_d - 1} \exp\left\{ \sum_{k,l=1}^{d} \sigma_{kl} x_k x_l \right\} \mathrm{d}x_1 \cdots \mathrm{d}x_{d-1}. \tag{4.63}$$

Proposition 4.5. *Under the above assumptions, the ∇_∞-valued random variable obtained by giving a common ordering to the d sequences $(\gamma_1 \xi_i^1)_{i\geq 1}, \ldots, (\gamma_d \xi_i^d)_{i \geq 1}$ has distribution μ_∞.*

Remark 4.6. μ_∞ is the distribution of allele frequencies in the stationary infinitely-many-alleles diffusion model in the situation where there are d classes of alleles; mutation to an allele in class k occurs with intensity $\frac{1}{2}\theta_k$; and the selection intensity of a genotype consisting of alleles in class k and class l is σ_{kl}. Two models of Li (1978), one for genic selection and one for recessive selection, fit into this framework. The result says that the class frequencies $\gamma_1, \ldots, \gamma_d$ are distributed according to (4.63) (cf. Lemma 4.1) and the within-class relative allele frequencies are Poisson–Dirichlet and independent of the class frequencies. In the special case of genic selection (i.e., $\sigma_{kl} = \sigma_k + \sigma_l$), this result is due to Griffiths (1983).

Proof of Proposition 4.5. Let $z_1 > z_2 > \cdots$ be the points of an inhomogeneous Poisson point process on $(0, \infty)$ with intensity function $\theta u^{-1} e^{-u}$. If $s = z_1 + z_2 + \cdots$, then Kingman (1975) has shown that $(z_1/s, z_2/s, \ldots)$ is Poisson–Dirichlet with parameter θ. Let $U_1, U_2, \ldots$ be i.i.d. uniform [0, 1], independent of $z_1 > z_2 > \cdots$. For $k = 1, \ldots, d$, let z_i^k be the ith largest z_j for which $U_j \in J_k$ and put $s_k = \Sigma_i z_i^k$; let $\eta_i^k = z_i^k / s_k$ and $\beta_k = s_k / s$. Then, by (4.39),

$$\mu_\infty(A) = \frac{E[I_A(\zeta((\beta_1 \eta_i^1)_{i\geq 1}, \ldots, (\beta_d \eta_i^d)_{i \geq 1})) \exp\{\sum_{k,l=1}^d \sigma_{kl} \beta_k \beta_l\}]}{E[\exp\{\sum_{k,l=1}^d \sigma_{kl} \beta_k \beta_l\}]}, \tag{4.64}$$

where ζ is the function that gives a common ordering to d sequences, and the result follows from the easily verified facts that $(\eta_1^k, \eta_2^k, \ldots)$, $k = 1, \ldots, d$, are independent

Poisson–Dirichlets with parameters $\theta_1, \ldots, \theta_d$, respectively, and $(\beta_1, \ldots, \beta_d)$ is Dirichlet with parameters $\theta_1, \ldots, \theta_d$, independent of the η_i^k's. (Cf. Donnelly and Tavaré, 1987, Section 3.) $\square$

Acknowledgements

The research leading to this paper was carried out while the second author was visiting the Department of Mathematics at the University of Utah. The hospitality of that department is acknowledged with appreciation. The research of both authors is supported in part by the National Science Foundation.

References

P. Donnelly and S. Tavaré, The population genealogy of the infinitely-many neutral alleles model, J. Math. Biol. 25 (1987), 381–391.

S.N. Ethier, A class of degenerate diffusion processes occurring in population genetics, Comm. Pure Appl. Math. 29 (1976), 483–493.

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

S.N. Ethier and R.C. Griffiths, The neutral two-locus model as a measure-valued diffusion, Adv. Appl. Probab. 22 (1990), 773–786.

S.N. Ethier and T.G. Kurtz, The infinitely-many-neutral-alleles diffusion model, Adv. Appl. Probab. 13 (1981), 429–452.

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

S.N. Ethier and T.G. Kurtz, The infinitely-many-alleles model with selection as a measure-valued diffusion, Lecture Notes in Biomathematics 70 (Springer, Berlin, 1987), 72–86.

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

S.N. Ethier and T.G. Kurtz, Coupling and ergodic theorems for Fleming–Viot processes, preprint (1994).

W.H. Fleming and M. Viot, Some measure-valued Markov processes in population genetics theory, Indiana Univ. Math. J. 28 (1979), 817–843.

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

R.C. Griffiths, Allele frequencies with genic selection, J. Math. Biol. 17 (1983), 1–10.

J.F.C. Kingman, Random discrete distributions, J. Roy. Statist. Soc. B 37 (1975), 1–22.

J.F.C. Kingman, The population structure associated with the Ewens sampling formula. Theoret. Popn. Biol. 11 (1977), 274–283.

T.G. Kurtz, Approximation of Population Processes, CBMS-NSF Regional Conf. Series in Appl. Math. 36 (SIAM, Philadelpha, 1981).

W.-H. Li, Maintenance of genetic variability under the joint effect of mutation, selection and random drift, Genetics 90 (1978), 349–382.

T. Shiga, Diffusion processes in population genetics, J. Math. Kyoto Univ. 21 (1981), 133–151.

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

G.A. Watterson, Heterosis or neutrality? Genetics 85 (1977), 789–814.

S. Wright, Adaptation and selection, in: Genetics, Paleontology, and Evolution, eds. G.L. Jepson et al. (Princeton Univ. Press, Princeton, 1949), pp. 365–389.