The Annals of Applied Probability
1992, Vol. 2, No. 1, 24–35

ON THE STATIONARY DISTRIBUTION OF THE NEUTRAL DIFFUSION MODEL IN POPULATION GENETICS[1]

BY S. N. ETHIER AND THOMAS G. KURTZ

University of Utah and University of Wisconsin-Madison

Let S be a compact metric space, let $\theta > 0$, and let $P(x, dy)$ be a one-step Feller transition function on $S \times \mathscr{B}(S)$ corresponding to a weakly ergodic Markov chain in S with unique stationary distribution ν_0. The neutral diffusion model, or Fleming–Viot process, with type space S, mutation intensity $\frac{1}{2}\theta$ and mutation transition function $P(x, dy)$, assumes values in $\mathscr{P}(S)$, the set of Borel probability measures on S with the topology of weak convergence, and is known to be weakly ergodic and have a unique stationary distribution $\Pi \in \mathscr{P}(\mathscr{P}(S))$.

Define the Markov chain $\{X(\tau),\ \tau \in \mathbf{Z}_+\}$ in $S^2 \cup S^3 \cup \cdots$ as follows. Let $X(0) = (\xi, \xi) \in S^2$, where ξ is an S-valued random variable with distribution ν_0. From state $(x_1, \ldots, x_n) \in S^n$, where $n \geq 2$, one of two types of transitions occurs. With probability $\theta/(n(n-1+\theta))$ a transition to state $(x_1, \ldots, x_{i-1}, \xi_i, x_{i+1}, \ldots, x_n) \in S^n$ occurs $(1 \leq i \leq n)$, where ξ_i is distributed according to $P(x_i, dy)$. With probability $(n-1)/((n+1)n(n-1+\theta))$ a transition to state $(x_1, \ldots, x_{j-1}, x_i, x_j, \ldots, x_n) \in S^{n+1}$ occurs $(1 \leq i \leq n,\ 1 \leq j \leq n+1)$. Letting τ_n denote the hitting time of S^n, we show that the empirical measure determined by the n coordinates of $X(\tau_{n+1} - 1)$ converges almost surely as $n \to \infty$ to a $\mathscr{P}(S)$-valued random variable with distribution Π.

1. Introduction. The neutral K-type diffusion model in population genetics is the diffusion process in the $(K-1)$-dimensional simplex

$$(1.1)\quad \Delta_K = \{p = (p_1, \ldots, p_K) \colon p_1 \geq 0, \ldots, p_K \geq 0,\ p_1 + \cdots + p_K = 1\}$$

with generator

$$(1.2)\qquad L = \frac{1}{2}\sum_{i,j=1}^{K} p_i(\delta_{ij} - p_j)\frac{\partial^2}{\partial p_i\,\partial p_j} + \sum_{i=1}^{K}\left(\sum_{j=1}^{K} q_{ji}p_j\right)\frac{\partial}{\partial p_i},$$

where q_{ij} $(i \neq j)$ is the intensity of a mutation from type i to type j and $q_{ii} = -\sum_{j:\, j \neq i} q_{ij}$. [Here the domain of L is $C^2(\Delta_K)$; for $f \in C^2(\Delta_K)$, Lf is the restriction to Δ_K of $L\tilde{f}$, where $\tilde{f}$ is an arbitrary $C^2(\mathbf{R}^K)$ extension of f.]

If the infinitesimal matrix (q_{ij}) is irreducible, it is known [Shiga (1981)] that this diffusion has a unique stationary distribution $\pi \in \mathscr{P}(\Delta_K)$, which is absolutely continuous with respect to $(K-1)$-dimensional Lebesgue measure

Received May 1990; revised February 1991.

[1]Research supported in part by NSF.

AMS 1980 *subject classifications*. Primary 60G57, 60J70, 92A10.

Key words and phrases. Population genetics, measure-valued diffusion, Fleming–Viot process, infinitely-many-sites model, genealogical tree.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Applied Probability*.
www.jstor.org

on Δ_K. In the special case of parent-independent mutation, that is,

$$q_{ij} = \frac{1}{2}\theta_j > 0, \qquad i, j \in \{1, \dots, K\}, i \neq j, \tag{1.3}$$

Wright (1949) discovered that

$$\pi(dp) = \frac{\Gamma(\theta_1 + \cdots + \theta_K)}{\Gamma(\theta_1) \cdots \Gamma(\theta_K)} p_1^{\theta_1 - 1} \cdots p_K^{\theta_K - 1}\, dp_1 \cdots dp_{K-1}; \tag{1.4}$$

see Ethier and Kurtz (1981) for a proof. In general, however, the Lebesgue density of π does not seem to be known.

Our principal aim here is to provide a relatively simple construction of a Δ_K-valued random variable with distribution π. Moreover, we can obtain a more general result by working within the framework of the Fleming–Viot (1979) process, which allows the number of types to be finite or infinite.

Let S be a compact metric space, let $\theta > 0$ and let $P(x, dy)$ be a one-step Feller transition function on $S \times \mathscr{B}(S)$. Here S is the set of types, θ is twice the mutation intensity and $P(x, dy)$ is the distribution of the type of a mutant offspring of a type x parent. The neutral diffusion model, or Fleming–Viot process, with parameters S, θ and P, is the diffusion process in $\mathscr{P}(S)$, the set of Borel probability measures on S with the topology of weak convergence, with generator $\mathscr{L}$ defined by

$$\begin{aligned}(\mathscr{L}\varphi)(\mu) &= \tfrac{1}{2} \sum_{i,j=1}^{m} \big(\langle f_i f_j, \mu\rangle - \langle f_i, \mu\rangle\langle f_j, \mu\rangle\big) F_{z_i z_j}\big(\langle f_1, \mu\rangle, \dots, \langle f_m, \mu\rangle\big) \\ &\quad + \tfrac{1}{2}\theta \sum_{i=1}^{m} \big(\langle Pf_i, \mu\rangle - \langle f_i, \mu\rangle\big) F_{z_i}\big(\langle f_1, \mu\rangle, \dots, \langle f_m, \mu\rangle\big),\end{aligned} \tag{1.5}$$

$$\mathscr{D}(\mathscr{L}) = \big\{\varphi \in C(\mathscr{P}(S)) \colon \varphi(\mu) \equiv F(\langle f_1, \mu\rangle, \dots, \langle f_m, \mu\rangle), \\ F \in C^2(\mathbf{R}^m),\ f_1, \dots, f_m \in C(S), m \in \mathbf{N}\big\},$$

where $\langle f, \mu\rangle = \int_S f\,d\mu$ and $P\colon C(S) \mapsto C(S)$ is given by

$$(Pf)(x) = \int_S f(y) P(x, dy). \tag{1.6}$$

The $C_{\mathscr{P}(S)}[0, \infty)$ martingale problem for $\mathscr{L}$ is well posed [Kurtz (1981)]. We assume only that there exists $\nu_0 \in \mathscr{P}(S)$ such that

$$\lim_{t \to \infty} \big(e^{\theta(P - I)t/2} f\big)(x) = \langle f, \nu_0\rangle, \qquad f \in C(S), x \in S, \tag{1.7}$$

an assumption slightly weaker than the one stated in the abstract.

Note that if $S = \{1, \dots, K\}$ and $\frac{1}{2}\theta(P(i, \{j\}) - \delta_{ij}) = q_{ij}$ for all $i, j \in S$, then we can identify $\mathscr{P}(S)$ with Δ_K and $\mathscr{L}$ reduces to L. However, the extra generality includes, for example, the infinitely-many-neutral-alleles diffusion model, which corresponds to the special case of the preceding paragraph in which

$$P(x, \{y\}) = 0, \qquad x, y \in S, \tag{1.8}$$

that is, with probability 1 every mutant allele is new. [Of course, (1.8) requires that S be uncountable.]

Under the weak ergodicity assumption (1.7) on the mutation semigroup, it is well known that the diffusion process in $\mathscr{P}(S)$ with generator $\mathscr{L}$ is weakly ergodic and has a unique stationary distribution $\Pi \in \mathscr{P}(\mathscr{P}(S))$. (Unfortunately, a published proof does not yet exist; in any case, the proof is straightforward, using a dual process.) In the special case of parent-independent mutation, that is,

$$(1.9) \qquad P(x, dy) = \nu_0(dy), \qquad x \in S,$$

Ethier and Kurtz (1992) showed that Π is the distribution of the $\mathscr{P}(S)$-valued random variable

$$(1.10) \qquad \sum_{i=1}^{\infty} \rho_i \delta_{V_i},$$

where $\rho_1 > \rho_2 > \cdots$ have the Poisson–Dirichlet distribution with parameter θ [Kingman (1975)] and $V_1, V_2, \ldots$ are i.i.d. ν_0 and independent of $(\rho_1, \rho_2, \ldots)$. That this result, with $S = \{1, \ldots, K\}$, implies (1.4) is a consequence of Section 3 of Donnelly and Tavaré (1987).

To construct a $\mathscr{P}(S)$-valued random variable with distribution Π without assuming (1.9), we define the Markov chain $\{X(\tau), \tau \in \mathbf{Z}_+\}$ in $S^2 \cup S^3 \cup \cdots$ as follows. Let $X(0) = (\xi, \xi) \in S^2$, where ξ is an S-valued random variable with distribution ν_0. From state $(x_1, \ldots, x_n) \in S^n$, where $n \geq 2$, one of two types of transitions occurs. With probability $\theta/(n(n-1+\theta))$ the ith coordinate mutates $(1 \leq i \leq n)$, in which case a transition to state $(x_1, \ldots, x_{i-1}, \xi_i, x_{i+1}, \ldots, x_n) \in S^n$ occurs, where ξ_i is distributed according to $P(x_i, dy)$. With probability $(n-1)/((n+1)n(n-1+\theta))$ the ith coordinate is duplicated and placed in the jth position $(1 \leq i \leq n, 1 \leq j \leq n+1)$, in which case a transition to state $(x_1, \ldots, x_{j-1}, x_i, x_j, \ldots, x_n) \in S^{n+1}$ occurs.

We further define, for each $n \geq 2$, the hitting time τ_n by

$$(1.11) \qquad \tau_n = \min\{\tau \in \mathbf{Z}_+ : X(\tau) \in S^n\}$$

and $\eta_n : S^n \mapsto \mathscr{P}(S)$ by

$$(1.12) \qquad \eta_n(x_1, \ldots, x_n) = n^{-1}(\delta_{x_1} + \cdots + \delta_{x_n});$$

note that $\eta_n(x_1, \ldots, x_n)$ is the empirical measure on S determined by the n (not necessarily distinct) points $x_1, \ldots, x_n \in S$. Our main result is the following.

THEOREM 1.1. *Assume* (1.7). *Then the limit*

$$(1.13) \qquad \lim_{n \to \infty} \eta_n(X(\tau_{n+1} - 1))$$

exists almost surely in the topology of $\mathscr{P}(S)$ *and has distribution* Π.

It should be emphasized that neither (1.8) nor (1.9) is needed here. In particular, the theorem includes the K-type case discussed earlier, assuming only the irreducibility of the infinitesimal matrix (q_{ij}).

The key step in the proof is to establish the following lemma, which relates the stationary distribution Π of the Fleming–Viot process to the Markov chain $\{X(\tau),\ \tau \in \mathbf{Z}_+\}$ and generalizes Theorem 5.4 of Ethier and Griffiths (1987).

LEMMA 1.2. *Assume* (1.7). *Then*

$$X(\tau_{n+1} - 1) \quad \textit{has distribution} \quad \int_{\mathscr{P}(S)} \mu^n(\cdot)\Pi(d\mu) \tag{1.14}$$

for each $n \geq 2$, *where* μ^n *is the n-fold product measure* $\mu \times \cdots \times \mu$.

Incidentally, the lemma provides an efficient algorithm for simulating a random sample of size n from μ, where μ is distributed according to Π: simply run the Markov chain $\{X(\tau),\ \tau \in \mathbf{Z}_+\}$ until it hits S^{n+1}, and disregard the last step. The expected number of steps required is

$$E[\tau_{n+1}] = n - 1 + \theta \sum_{k=1}^{n-1} \frac{1}{k}. \tag{1.15}$$

The proof of Lemma 1.2 is somewhat nonintuitive, so it might be worthwhile examining separately the special case $n = 2$, which can be easily understood.

Since Π is stationary,

$$\int_{\mathscr{P}(S)} (\mathscr{L}\varphi)(\mu)\Pi(d\mu) = 0 \tag{1.16}$$

for all $\varphi \in \mathscr{D}(\mathscr{L})$. Taking $\varphi(\mu) \equiv \langle f, \mu\rangle$, where $f \in C(S)$, this gives $\int_{\mathscr{P}(S)}\langle f, \mu\rangle\Pi(d\mu) = \int_{\mathscr{P}(S)}\langle Pf, \mu\rangle\Pi(d\mu)$, hence

$$\begin{aligned}\int_{\mathscr{P}(S)} \langle f, \mu\rangle \Pi(d\mu) &= \int_{\mathscr{P}(S)} \left\langle e^{-\theta t/2} \sum_{k=0}^{\infty} \frac{(\theta t/2)^k}{k!} P^k f, \mu \right\rangle \Pi(d\mu) \\ &= \int_{\mathscr{P}(S)} \langle e^{\theta(P-I)t/2} f, \mu\rangle \Pi(d\mu) \qquad (1.17)\end{aligned}$$

for all $t > 0$. Letting $t \to \infty$ and recalling (1.7), we have

$$\int_{\mathscr{P}(S)} \langle f, \mu\rangle\Pi(d\mu) = \langle f, \nu_0\rangle, \qquad f \in C(S). \tag{1.18}$$

Next, apply (1.16) with $\varphi(\mu) \equiv \langle f, \mu\rangle\langle g, \mu\rangle$, where $f, g \in C(S)$, and use (1.18) to obtain

$$\begin{aligned}&\int_{\mathscr{P}(S)} \langle f, \mu\rangle\langle g, \mu\rangle\Pi(d\mu) \\ &\quad = \frac{1}{1+\theta}\langle fg, \nu_0\rangle + \frac{\theta}{1+\theta}\left(\frac{1}{2}\int_{\mathscr{P}(S)}\langle Pf, \mu\rangle\langle g, \mu\rangle\Pi(d\mu)\right. \qquad (1.19) \\ &\qquad \left. + \frac{1}{2}\int_{\mathscr{P}(S)}\langle f, \mu\rangle\langle Pg, \mu\rangle\Pi(d\mu)\right).\end{aligned}$$

Iterating this identity readily yields

$$\int_{\mathscr{P}(S)} \langle f, \mu \rangle \langle g, \mu \rangle \Pi(d\mu) = \sum_{k=0}^{\infty} \left(\frac{\theta}{1+\theta} \right)^k \frac{1}{1+\theta} \sum_{i=0}^{k} \binom{k}{i} 2^{-k} \langle P^i f \cdot P^{k-i} g, \nu_0 \rangle, \quad f, g \in C(S). \tag{1.20}$$

On the other hand, the Markov chain $\{X(\tau), \tau \in \mathbf{Z}_+\}$, where for the moment we allow $X(0)$ to be an arbitrary S^2-valued random variable, clearly satisfies

$$E[(f \times g)(X(1)) | X(1) \in S^2, X(0)] = \tfrac{1}{2}(Pf \times g + f \times Pg)(X(0)) \tag{1.21}$$

for all $f, g \in C(S)$, where $(f \times g)(x_1, x_2) \equiv f(x_1) g(x_2)$. Iterating, we have

$$\begin{aligned} E[(f \times g)(X(\tau_3 - 1)) | \tau_3 - 1 = k, X(0)] &= E[(f \times g)(X(k)) | X(k) \in S^2, X(0)] \\ &= \sum_{i=0}^{k} \binom{k}{i} 2^{-k} (P^i f \times P^{k-i} g)(X(0)), \end{aligned} \tag{1.22}$$

where the first equality uses the conditional independence of $\{X(k+1) \in S^3\}$ and $X(k)$, given $X(k) \in S^2$. Now, since $X(0) = (\xi, \xi)$ by definition, where ξ has distribution ν_0, and since $\tau_3 - 1$ has a geometric distribution on $\mathbf{Z}_+$ with parameter $1/(1+\theta)$, we conclude that

$$E[(f \times g)(X(\tau_3 - 1))] = \sum_{k=0}^{\infty} \left(\frac{\theta}{1+\theta} \right)^k \frac{1}{1+\theta} \sum_{i=0}^{k} \binom{k}{i} 2^{-k} \langle P^i f \cdot P^{k-i} g, \nu_0 \rangle, \quad f, g \in C(S). \tag{1.23}$$

Comparing (1.20) and (1.23), we find that $X(\tau_3 - 1)$ has distribution $\int_{\mathscr{P}(S)} \mu^2(\cdot) \Pi(d\mu)$, which is the special case of Lemma 1.2 in which $n = 2$.

In Section 2 we generalize the preceding argument to prove Lemma 1.2. In Section 3 we apply the martingale convergence theorem to show that the limit (1.13) exists almost surely; then, invoking a lemma of Dawson and Hochberg (1982), we use Lemma 1.2 to show that (1.13) has distribution Π, as required. Finally, in an Appendix, we state a result that expresses the moment measures of Π in terms of the genealogical-tree probabilities in the stationary infinitely-many-sites model of Ethier and Griffiths (1987), thereby generalizing (1.20). But because this result is not needed here, its proof is left to the interested reader.

Without further mention, we assume (1.7) throughout.

2. The moment measures of Π and the stopped Markov chain. Applying (1.16) to the monomial $\varphi(\mu) \equiv \langle f_1, \mu \rangle \cdots \langle f_n, \mu \rangle$, where $f_1, \dots, f_n \in C(S)$ and $n \in \mathbf{N}$, we have

$$n(n-1+\theta)\int_{\mathscr{P}(S)} \langle f_1, \mu\rangle \cdots \langle f_n, \mu\rangle \Pi(d\mu)$$

$$(2.1) \qquad = 2 \sum_{1 \le i < j \le n} \int_{\mathscr{P}(S)} \langle f_i f_j, \mu\rangle \prod_{l:\, l \ne i, j} \langle f_l, \mu\rangle \Pi(d\mu)$$

$$+ \theta \sum_{i=1}^{n} \int_{\mathscr{P}(S)} \langle Pf_i, \mu\rangle \prod_{l:\, l \ne i} \langle f_l, \mu\rangle \Pi(d\mu).$$

We begin by showing that this system uniquely determines the moments $\int_{\mathscr{P}(S)} \langle f_1, \mu\rangle \cdots \langle f_n, \mu\rangle \Pi(d\mu)$ for all $f_1, \dots, f_n \in C(S)$ and $n \in \mathbf{N}$. For each $n \ge 2$, $i \in \{1, \dots, n-1\}$, and $j \in \{1, \dots, n\}$, define $\Phi_{ij}^{(n)}\colon C(S)^n \mapsto C(S)^{n-1}$ by

$$\Phi_{ij}^{(n)}(f_1, \dots, f_n)$$

$$(2.2) \qquad = \begin{cases} \left(f_1, \dots, f_{i-1}, f_i f_j, f_{i+1}, \dots, \check{f_j}, \dots, f_n\right), & \text{if } i < j, \\ \left(f_1, \dots, \check{f_j}, \dots, f_i, f_j f_{i+1}, f_{i+2}, \dots, f_n\right), & \text{if } i \ge j, \end{cases}$$

where the $\check{}$ notation signifies deletion of the component in question.

Lemma 2.1. *For each $n \in \mathbf{N}$, let $\Lambda_n\colon C(S)^n \mapsto \mathbf{R}$ satisfy $|\Lambda_n(f_1, \dots, f_n)| \le \|f_1\| \cdots \|f_n\|$ for all $f_1, \dots, f_n \in C(S)$. Suppose that*

$$(2.3) \qquad n(n-1+\theta)\Lambda_n(f_1, \dots, f_n) = \sum_{i=1}^{n-1} \sum_{j=1}^{n} \Lambda_{n-1}\left(\Phi_{ij}^{(n)}(f_1, \dots, f_n)\right) + \theta \sum_{i=1}^{n} \Lambda_n(f_1, \dots, f_{i-1}, Pf_i, f_{i+1}, \dots, f_n)$$

for all $f_1, \dots, f_n \in C(S)$ and $n \in \mathbf{N}$. Then

$$(2.4) \qquad \Lambda_n(f_1, \dots, f_n) = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=1}^{n} \sum_{k=0}^{\infty} (1-p_n)^k p_n \times \sum_{\alpha \in (\mathbf{Z}_+)^n:\, |\alpha| = k} \binom{k}{\alpha} n^{-k} \Lambda_{n-1}\left(\Phi_{ij}^{(n)}(P^{\alpha_1} f_1, \dots, P^{\alpha_n} f_n)\right)$$

for all $f_1, \dots, f_n \in C(S)$ and $n \ge 2$, where $p_n = (n-1)/(n-1+\theta)$.

Suppose further that Λ_1 is linear and $\Lambda_1(1) = 1$. Then

$$(2.5) \qquad \Lambda_n(f_1, \dots, f_n) = \int_{\mathscr{P}(S)} \langle f_1, \mu\rangle \cdots \langle f_n, \mu\rangle \Pi(d\mu)$$

for all $f_1, \dots, f_n \in C(S)$ and $n \in \mathbf{N}$.

PROOF. The proof relies on a discrete-time function-valued dual process. Fix $n \geq 2$ and $f_1, \dots, f_n \in C(S)$, and define the Markov chain $\{Y(\tau), \tau \in \mathbf{Z}_+\}$ in $C(S) \cup C(S)^2 \cup \cdots \cup C(S)^n$ as follows. Let $Y(0) = (f_1, \dots, f_n) \in C(S)^n$. From state $(g_1, \dots, g_m) \in C(S)^m$, where $2 \leq m \leq n$, one of two types of transitions occurs. With probability $1/(m(m-1+\theta))$ a transition to state $\Phi_{ij}^{(m)}(g_1, \dots, g_m) \in C(S)^{m-1}$ occurs $(1 \leq i \leq m-1, 1 \leq j \leq m)$. With probability $\theta/(m(m-1+\theta))$ a transition to state $(g_1, \dots, g_{i-1}, Pg_i, g_{i+1}, \dots, g_m) \in C(S)^m$ occurs $(1 \leq i \leq m)$. From state $g \in C(S)$ a transition to state $Pg \in C(S)$ occurs with probability 1. Letting $M(\tau) = m$ if $Y(\tau) \in C(S)^m$, it follows from (2.3) that $\{\Lambda_{M(\tau)}(Y(\tau)), \tau \in \mathbf{Z}_+\}$ is a martingale. Hence

$$\Lambda_n(f_1, \dots, f_n) = E\left[\Lambda_{M(0)}(Y(0))\right] = E\left[\Lambda_{M(\sigma)}(Y(\sigma))\right], \tag{2.6}$$

where $\sigma = \min\{\tau \in \mathbf{Z}_+ : Y(\tau) \in C(S)^{n-1}\}$. This implies (2.4). Under the additional assumptions, $\Lambda_1(f) = \langle f, \nu_0 \rangle$ for all $f \in C(S)$ by the same argument as the one used to prove (1.18). This, together with (2.4), implies that (2.3) has a unique solution, which by (2.1) yields (2.5). □

PROOF OF LEMMA 1.2. For each $n \in \mathbf{N}$ and $f_1, \dots, f_n \in C(S)$, define

$$\Lambda_n(f_1, \dots, f_n) = \int_{\mathscr{P}(S)} \langle f_1, \mu \rangle \cdots \langle f_n, \mu \rangle \Pi(d\mu) \tag{2.7}$$

and

$$\tilde{\Lambda}_n(f_1, \dots, f_n) = \begin{cases} \langle f_1, \nu_0 \rangle, & \text{if } n = 1, \\ E[(f_1 \times \cdots \times f_n)(X(\tau_{n+1} - 1))], & \text{if } n \geq 2, \end{cases} \tag{2.8}$$

where $f_1 \times \cdots \times f_n \in C(S^n)$ is defined by $(f_1 \times \cdots \times f_n)(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$. Note that $\Lambda_1 = \tilde{\Lambda}_1$ by (1.18). By (2.1) and the first conclusion of Lemma 2.1, $\{\Lambda_n\}$ satisfies (2.4). To show that $\Lambda_n = \tilde{\Lambda}_n$ for all $n \in \mathbf{N}$ and thereby complete the proof, it will suffice to show that $\{\tilde{\Lambda}_n\}$ satisfies (2.4) (with tildes inserted on both sides). The case $n = 2$ follows by (1.23), so suppose $n \geq 3$. Using the strong Markov property, we have

$$\begin{aligned} &E[(f_1 \times \cdots \times f_n)(X(\tau_{n+1} - 1)) | X(\tau_n - 1) = (x_1, \dots, x_{n-1})] \\ &\quad = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=1}^{n} E[(f_1 \times \cdots \times f_n)(X(\tau_{n+1} - 1)) | \\ &\qquad\qquad X(\tau_n) = (x_1, \dots, x_{j-1}, x_i, x_j, \dots, x_{n-1})] \\ &\quad = \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=1}^{n} \sum_{k=0}^{\infty} (1 - p_n)^k p_n \sum_{\alpha \in (\mathbf{Z}_+)^n : |\alpha| = k} \binom{k}{\alpha} n^{-k} \\ &\qquad \times (P^{\alpha_1} f_1)(x_1) \cdots (P^{\alpha_{j-1}} f_{j-1})(x_{j-1})(P^{\alpha_j} f_j)(x_i)(P^{\alpha_{j+1}} f_{j+1})(x_j) \\ &\qquad \cdots (P^{\alpha_n} f_n)(x_{n-1}) \end{aligned} \tag{2.9}$$

for all $f_1, \ldots, f_n \in C(S)$ and $(x_1, \ldots, x_{n-1}) \in S^{n-1}$. Taking expectations now gives the desired conclusion. □

3. Almost-sure convergence of the empirical measures. First, we show that the limit (1.13) exists almost surely, and then we identify its distribution. Define $\{N(\tau), \tau \in \mathbf{Z}_+\}$ by $N(\tau) = n$ if $X(\tau) \in S^n$.

LEMMA 3.1. $\lim_{\tau\to\infty} \eta_{N(\tau)}(X(\tau))$ *exists almost surely in the topology of* $\mathscr{P}(S)$.

PROOF. Define $\mu_\tau = \eta_{N(\tau)}(X(\tau))$ for $\tau = 0, 1, \ldots$. Then $\{(\mu_\tau, N(\tau)), \tau \in \mathbf{Z}_+\}$ is a Markov chain in $\bigcup_{n=2}^\infty \eta_n(S^n) \times \{n\} \subset \mathscr{P}(S) \times \mathbf{N}$ with transitions

$$(\mu, n) = \big(n^{-1}(\delta_{x_1} + \cdots + \delta_{x_n}), n\big)$$
$$\mapsto \begin{cases} \big(\mu + n^{-1}(\delta_{\xi_i} - \delta_{x_i}), n\big), \\ \qquad \text{with prob. } \theta/(n(n-1+\theta)),\ i = 1, \ldots, n, \\ \big(\mu + (n+1)^{-1}(\delta_{x_i} - \mu), n+1\big), \\ \qquad \text{with prob. } (n-1)/(n(n-1+\theta)),\ i = 1, \ldots, n, \end{cases} \tag{3.1}$$

where ξ_i is distributed according to $P(x_i, dy)$. Let $f \in C(S)$. Then

$$\begin{aligned} E\big[\langle f, \mu_{k+1}\rangle - \langle f, \mu_k\rangle \,|\, (\mu_k, N(k)) = (\mu, n) = \big(n^{-1}(\delta_{x_1} + \cdots + \delta_{x_n}), n\big)\big] \\ = \frac{\theta}{n(n-1+\theta)} \sum_{i=1}^n E\big[\big\langle f, n^{-1}(\delta_{\xi_i} - \delta_{x_i})\big\rangle\big] \\ + \frac{n-1}{n(n-1+\theta)} \sum_{i=1}^n \big\langle f, (n+1)^{-1}(\delta_{x_i} - \mu)\big\rangle \\ = \frac{\theta}{n(n-1+\theta)} n^{-1} \sum_{i=1}^n (Pf - f)(x_i) \\ + \frac{n-1}{n(n-1+\theta)} (n+1)^{-1} \sum_{i=1}^n (f(x_i) - \langle f, \mu\rangle) \\ = \frac{\theta}{n(n-1+\theta)} \langle Pf - f, \mu\rangle \end{aligned} \tag{3.2}$$

for $k = 0, 1, \ldots$, so

$$\begin{aligned} M_\tau &\equiv \langle f, \mu_\tau\rangle - \sum_{k=0}^{\tau-1} E\big[\langle f, \mu_{k+1}\rangle - \langle f, \mu_k\rangle \,|\, (\mu_k, N(k))\big] \\ &= \langle f, \mu_\tau\rangle - \theta \sum_{k=0}^{\tau-1} \frac{1}{N(k)(N(k) - 1 + \theta)} \langle Pf - f, \mu_k\rangle \end{aligned} \tag{3.3}$$

is a martingale.

Now the time spent by $\{X(\tau), \tau \in \mathbf{Z}_+\}$ in S^n is geometrically distributed on $\mathbf{N}$ with parameter $(n-1)/(n-1+\theta)$, so

$$E\left[\sum_{k=0}^{\infty} \frac{1}{N(k)(N(k)-1+\theta)}\right] = E\left[\sum_{n=2}^{\infty} \frac{\tau_{n+1}-\tau_n}{n(n-1+\theta)}\right] \tag{3.4}$$
$$= \sum_{n=2}^{\infty} \frac{1}{n(n-1+\theta)} \frac{n-1+\theta}{n-1} = 1.$$

Consequently, the martingale convergence theorem implies that M_τ, and hence $\langle f, \mu_\tau \rangle$, converges almost surely as $\tau \to \infty$. But $f \in C(S)$ was arbitrary, so since $C(S)$ with the supremum norm is separable, we conclude that μ_τ converges almost surely in the topology of $\mathscr{P}(S)$ as $\tau \to \infty$. □

The next lemma is a slight modification of part of Lemma 6.1 of Dawson and Hochberg (1982).

LEMMA 3.2. *Let Π_0 in $\mathscr{P}(\mathscr{P}(S))$ be arbitrary, and let $Z_1, Z_2, \ldots$ be a sequence of S-valued random variables such that*

$$(Z_1, \ldots, Z_n) \quad \textit{has distribution} \quad \int_{\mathscr{P}(S)} \mu^n(\cdot)\Pi_0(d\mu) \tag{3.5}$$

for each $n \in \mathbf{N}$. (Such a sequence exists by Kolmogorov's extension theorem and is exchangeable.) Then $\lim_{n\to\infty} \eta_n(Z_1, \ldots, Z_n)$ exists almost surely in the topology of $\mathscr{P}(S)$ and has distribution Π_0.

REMARK. Dawson and Hochberg (1982) refer to this construction as the *canonical representation* of a random probability measure.

PROOF. It suffices to replace the countable collection of indicator functions of Dawson and Hochberg (1982) by a countable dense subset of $C(S)$. □

We now have all the ingredients necessary to prove our main result.

PROOF OF THEOREM 1.1. Let $Z_1, Z_2, \ldots$ be as in Lemma 3.2 with $\Pi_0 = \Pi$. By Lemma 1.2,

$$\eta_n(X(\tau_{n+1}-1)) =_{\mathscr{D}} \eta_n(Z_1, \ldots, Z_n), \qquad n \geq 2, \tag{3.6}$$

where $=_{\mathscr{D}}$ denotes equality in distribution. Lemmas 3.1 and 3.2 imply that both sides of (3.6) converge almost surely. By Lemma 3.2, the limit of the right side has distribution Π. Therefore the same must be true of the limit of the left side and the proof is complete. □

REMARKS. (a) In view of Lemma 3.1, the subsequential limit $\lim_{n\to\infty} \eta_n(X(\tau_{n+1}-1))$ in Theorem 1.1 can be replaced by the limit $\lim_{\tau\to\infty} \eta_{N(\tau)}(X(\tau))$ of the full sequence, slightly strengthening the conclusion of the theorem.

(b) Recall that $\{X(\tau),\ \tau \in \mathbf{Z}_+\}$ has two types of transitions, mutations and duplications. Define the slightly simpler Markov chain $\{X^{\circ}(\tau),\ \tau \in \mathbf{Z}_+\}$ in $S^2 \cup S^3 \cup \cdots$ similarly, except that when a duplication occurs, the duplicated coordinate is placed in the $(n+1)$th position. More precisely, from state $(x_1,\dots,x_n) \in S^n$, where $n \geq 2$, a transition to state $(x_1,\dots,x_n,x_i) \in S^{n+1}$ occurs $(1 \leq i \leq n)$ with probability $(n-1)/(n(n-1+\theta))$.

Then the conclusion of Theorem 1.1 holds without change for the modified Markov chain. The conclusion of Lemma 1.2, however, must be weakened to

$$\eta_n(X^{\circ}(\tau_{n+1}-1)) \quad \text{has distribution} \quad \int_{\mathscr{P}(S)} \mu^n \eta_n^{-1}(\cdot)\Pi(d\mu) \tag{3.7}$$

for each $n \geq 2$. For purposes of simulation, however, this will suffice if the statistics of interest are symmetric with respect to the sample, as they usually are. The proofs of these assertions are almost immediate by noting that the Markov chain (3.1) is insensitive to the ordering of the coordinates.

APPENDIX

The moment measures of Π and genealogical trees. Here we state a result that generalizes (1.20) from $n = 2$ to arbitrary n. It requires a considerable amount of notation, much of it from Ethier and Griffiths (1987), referred to hereafter as E–G (1987). For the convenience of the reader, we repeat here the essential definitions.

Let $E = [0,1]^{\mathbf{Z}_+}$. For $n \in \mathbf{N}$ define $(\mathbf{x}_1,\dots,\mathbf{x}_n) \in E^n$ to be an *n-tree* if: (i) the coordinates x_{ij}, $j = 0,1,\dots,$ of $\mathbf{x}_i$ are distinct for fixed $i \in \{1,\dots,n\}$; (ii) whenever $i, i' \in \{1,\dots,n\}$, $j, j' \in \mathbf{Z}_+$, and $x_{ij} = x_{i'j'}$, we have $x_{i,j+l} = x_{i',j'+l}$ for $l = 0,1,\dots$; and (iii) there exist $j_1,\dots,j_n \in \mathbf{Z}_+$ such that $x_{1j_1} = \cdots = x_{nj_n}$. Let $\mathscr{T}_n \subset E^n$ be the set of all n-trees and define the equivalence relation $\sim$ on $\mathscr{T}_n$ as follows. Say that $(\mathbf{x}_1,\dots,\mathbf{x}_n) \sim (\mathbf{y}_1,\dots,\mathbf{y}_n)$ if there exists a bijection ζ: $[0,1] \mapsto [0,1]$ with $y_{ij} = \zeta(x_{ij})$ for $i = 1,\dots,n$ and $j = 0,1,\dots$. Let $\mathscr{T}_n/\sim$ denote the quotient set of equivalence classes. For $d \in \mathbf{N}$, define $(\mathscr{T}_d/\sim)_0$ to be the set of all $T \in \mathscr{T}_d/\sim$ such that $\mathbf{x}_1,\dots,\mathbf{x}_d$ are distinct whenever $(\mathbf{x}_1,\dots,\mathbf{x}_d) \in T$. Define the shift operator $\mathscr{S}$: $E \mapsto E$ by $\mathscr{S}\mathbf{x} = \mathscr{S}(x_0,x_1,\dots) = (x_1,x_2,\dots)$. For $k = 1,\dots,d$, define $\mathscr{S}_k$: $E^d \mapsto E^d$ by $\mathscr{S}_k(\mathbf{x}_1,\dots,\mathbf{x}_d) = (\mathbf{x}_1,\dots,\mathbf{x}_{k-1},\mathscr{S}\mathbf{x}_k,\mathbf{x}_{k+1},\dots,\mathbf{x}_d)$, and note that $\mathscr{S}_k$ induces a map (also denoted by $\mathscr{S}_k$) from $\mathscr{T}_d/\sim$ into $\mathscr{T}_d/\sim$. For $k = 1,\dots,d$, we say that $T \in \mathscr{T}_d/\sim$ has the property that x_{k0} *is distinct* if $x_{k0} \neq x_{ij}$ for all $(\mathbf{x}_1,\dots,\mathbf{x}_d) \in T$ and $(i,j) \neq (k,0)$. Finally, $z \in [0,1]$ is said to be a *segregating site* of $(\mathbf{x}_1,\dots,\mathbf{x}_n) \in \mathscr{T}_n$ if z appears in at least one but not all of the sequences $\mathbf{x}_1,\dots,\mathbf{x}_n$. Note that one can refer to the number of segregating sites of an equivalence class $T \in \mathscr{T}_n/\sim$. See E–G (1987) for discussion and motivation.

In addition, for $1 \le d \le n$, let $\pi(n,d)$ be the set of partitions β of $\{1,\dots,n\}$ into d unordered, nonempty subsets (e.g., $|\pi(n,n)| = 1$). However, we want to be able to refer to the individual subsets that make up a partition $\beta \in \pi(n,d)$, so we denote them by $\beta_1,\dots,\beta_d$, where $\min\beta_1 < \cdots < \min\beta_d$. For $\beta \in \pi(n,d)$, define Φ_β: $E^d \mapsto E^n$ by $\Phi_\beta(\mathbf{x}_1,\dots,\mathbf{x}_d) = (\mathbf{y}_1,\dots,\mathbf{y}_n)$, where $\mathbf{y}_i = \mathbf{x}_k$ whenever $i \in \beta_k$, and define Ψ_β: $B(S^n) \mapsto B(S^d)$ by $(\Psi_\beta f)(x_1,\dots,x_d) = f(y_1,\dots,y_n)$, where $y_i = x_k$ whenever $i \in \beta_k$. Note that Φ_β induces a map (also denoted by Φ_β) from $\mathscr{T}_d/\sim$ into $\mathscr{T}_n/\sim$. For $k = 1,\dots,d$, define P_k: $B(S^d) \mapsto B(S^d)$ by

$$(P_k f)(x_1,\dots,x_d) = \int_S f(x_1,\dots,x_{k-1},y,x_{k+1},\dots,x_d)P(x_i,dy).$$

For $T \in \mathscr{T}_n/\sim$, we define P_T: $B(S^n) \mapsto B(S)$ recursively as follows. Choose $d \in \{1,\dots,n\}$, $\beta \in \pi(n,d)$ and $T_0 \in (\mathscr{T}_d/\sim)_0$ such that $T = \Phi_\beta(T_0)$. Then

$$P_T = P_{T_0}\Psi_\beta, \quad \text{where } P_{T_0} = P_{\mathscr{S}_k T_0}P_k \text{ if } x_{k0} \text{ is distinct;} \tag{A.1}$$

also, $P_T = I$ for $T \in (\mathscr{T}_1/\sim)_0$. If we define the degree of an equivalence class $T \in \mathscr{T}_n/\sim$ with s segregating sites to be $s + n$, then $\mathscr{S}_k T_0$ in (A.1) has degree $s - 1 + d$, which is less than $s + n$, thus explaining the recursion. (Note that, if $d \ge 2$, every $T \in (\mathscr{T}_d/\sim)_0$ has at least one distinct segregating site.) It is left to the reader to check that the order in which distinct segregating sites are removed does not matter.

EXAMPLE A.1. Let $T_0 \in (\mathscr{T}_3/\sim)_0$ be as in Figure 1 of E–G (1987), that is,

$$\begin{aligned} T_0 = \{&((y_1,x_0,x_1,\dots),(y_4,y_3,y_2,x_0,x_1,\dots),(y_5,y_2,x_0,x_1,\dots)):\\ &\qquad y_1,y_2,y_3,y_4,y_5,x_0,x_1,\dots \in [0,1] \text{ distinct}\}. \end{aligned} \tag{A.2}$$

Let $\beta \in \pi(6,3)$ be the partition $\beta_1 = \{1,5\}$, $\beta_2 = \{2,3,6\}$, $\beta_3 = \{4\}$. Then, for $f_1,\dots,f_6 \in C(S)$,

$$\begin{aligned} P_{T_0}\Psi_\beta(f_1 \times \cdots \times f_6) &= P_{T_0}(f_1 f_5 \times f_2 f_3 f_6 \times f_4)\\ &= P(f_1 f_5) \cdot P\big(P^2(f_2 f_3 f_6) \cdot P f_4\big) \in C(S). \end{aligned} \tag{A.3}$$

There is one more definition needed at this point. Let $\tilde{\mu} \in \mathscr{P}(\mathscr{P}(E))$ be the unique stationary distribution of the infinitely-many-sites model of E–G (1987) and for $n \in \mathbf{N}$, $d \in \{1,\dots,n\}$, $\beta \in \pi(n,d)$ and $T \in (\mathscr{T}_d/\sim)_0$, define

$$p(T,\beta) = \int_{\mathscr{P}(E)} \mu^n\big(\Phi_\beta(T)\big)\tilde{\mu}(d\mu). \tag{A.4}$$

Letting $\mathbf{n} = (|\beta_1|,\dots,|\beta_d|)$, we note that $p(T,\beta) = p(T,\mathbf{n})$, where the latter is as in (3.5) of E–G (1987).

PROPOSITION A.2. *For each $n \in \mathbf{N}$ and $f \in B(S^n)$*,

$$\text{(A.5)} \quad \int_{\mathscr{P}(S)} \langle f, \mu^n \rangle \,\Pi(d\mu) = \sum_{d=1}^{n} \sum_{\beta \in \pi(n,d)} \sum_{T \in (\mathscr{T}_d/\sim)_0} p(T,\beta) \langle P_T \Psi_\beta f, \nu_0 \rangle.$$

REMARKS. (a) Note that (A.5) is the integral of f with respect to the nth moment measure of Π.

(b) By (1.15) of E–G (1987), the special case of (A.5) in which $n = 2$ implies (1.20).

(c) If (1.9) holds, then (A.5) implies [see E–G (1987), page 537] that

$$\text{(A.6)} \quad \int_{\mathscr{P}(S)} \langle f_1, \mu \rangle \cdots \langle f_n, \mu \rangle \Pi(d\mu) = \sum_{d=1}^{n} \sum_{\beta \in \pi(n,d)} p(\beta) \prod_{k=1}^{d} \left\langle \prod_{i \in \beta_k} f_i, \nu_0 \right\rangle,$$

where $f_1, \dots, f_n \in C(S)$ and $p(\beta) = (|\beta_1| - 1)! \cdots (|\beta_d| - 1)! \theta^{d-1} / ((1 + \theta) \cdots (n - 1 + \theta))$, a result of Ethier (1990).

(d) A proof can be based on Lemma 2.1 above and Corollary 4.2 of E–G (1987).

Acknowledgments. It is a pleasure to thank Peter Donnelly for helpful discussions and R. C. Griffiths for the observation that (1.8) should not be needed.

REFERENCES

DAWSON, D. A. and HOCHBERG, K. J. (1982). Wandering random measures in the Fleming–Viot model. *Ann. Probab.* **10** 554–580.

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

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

ETHIER, S. N. and GRIFFITHS, R. C. (1987). The infinitely-many-sites model as a measure-valued diffusion. *Ann. Probab.* **15** 515–545.

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

ETHIER, S. N. and KURTZ, T. G. (1992). Convergence to Fleming–Viot processes in the weak atomic topology. Unpublished manuscript.

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

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

KURTZ, T. G. (1981). *Approximation of Population Processes*. SIAM, Philadelphia.

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

WRIGHT, S. (1949). Adaptation and selection. In *Genetics, Paleontology, and Evolution*. (G. L. Jepson, E. Mayr and G. G. Simpson, eds.) 365–389. Princeton Univ. Press.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
SALT LAKE CITY, UTAH 84112

DEPARTMENTS OF MATHEMATICS AND STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
MADISON, WISCONSIN 53706