Statistica Sinica **19** (2009), 707-720

BAYESIAN NONPARAMETRIC CONSTRUCTION OF THE FLEMING-VIOT PROCESS WITH FERTILITY SELECTION

Matteo Ruggiero and Stephen G. Walker

Università degli Studi di Pavia and University of Kent

Abstract: This paper provides a construction in the Bayesian framework of the Fleming-Viot measure-valued diffusion with diploid fertility selection, and highlights new connections between Bayesian nonparametrics and population genetics. Via a generalisation of the Blackwell-MacQueen Pólya-urn scheme, a Markov particle process is defined such that the associated process of empirical measures converges to the Fleming-Viot diffusion. The stationary distribution, known from Ethier and Kurtz (1994), is then derived through an application of the Dirichlet process mixture model and shown to be the de Finetti measure of the particle process. The Fleming-Viot process with haploid selection is derived as a special case.

Key words and phrases: Blackwell-MacQueen urn-scheme, Dirichlet process mixture, fertility selection, Fleming-Viot process, Gibbs sampler.

1. Introduction And Preliminaries

The Fleming-Viot process, introduced by Fleming and Viot (1979), is a diffusion on the space $\mathscr{P}(\mathcal{X})$ of Borel probability measures on $\mathcal{X}$, endowed with the topology of weak convergence, where $\mathcal{X}$ is a locally compact complete separable metric space called the *type space.* The general form of the generator which provides the Fleming-Viot process is given by Ethier and Kurtz (1993) as

$$\begin{aligned}\mathbb{A}\phi(\mu) =&\frac{1}{2}\int_{\mathcal{X}}\int_{\mathcal{X}}\mu(\mathrm{d}x)\{\delta_x(\mathrm{d}y)-\mu(\mathrm{d}y)\}\frac{\partial^2\phi(\mu)}{\partial\mu(x)\partial\mu(y)}\\&+\int_{\mathcal{X}}\mu(\mathrm{d}x)\,G\left(\frac{\partial\phi(\mu)}{\partial\mu(\cdot)}\right)(x)+\int_{\mathcal{X}}\int_{\mathcal{X}}\mu(\mathrm{d}x)\mu(\mathrm{d}y)R\left(\frac{\partial\phi(\mu)}{\partial\mu(\cdot)}\right)(x,y)\\&+\int_{\mathcal{X}}\int_{\mathcal{X}}\mu(\mathrm{d}x)\mu(\mathrm{d}y)(\sigma(x,y)-\langle\sigma,\mu^2\rangle)\frac{\partial\phi(\mu)}{\partial\mu(\cdot)},\end{aligned}$$

where $\partial\phi(\mu)/\partial\mu(x)=\lim_{\varepsilon\to0^+}\varepsilon^{-1}\{\phi(\mu+\varepsilon\delta_x)-\phi(\mu)\}$, μ^2 denotes product measure, and we take the domain $\mathscr{D}(\mathbb{A})$ to be the set of all $\phi\in B(\mathscr{P}(\mathcal{X}))$, where $B(\mathscr{P}(\mathcal{X}))$ is the set of bounded, Borel measurable functions on $\mathscr{P}(\mathcal{X})$, of the form $\phi(\mu)=F(\langle f_1,\mu\rangle,\ldots,\langle f_m,\mu\rangle)=F(\langle\mathbf{f},\mu\rangle)$, where $\langle f,\mu\rangle=\int f\mathrm{d}\mu$ and, for

$m \geq 1$, we have $f_1, \ldots, f_m \in \mathscr{D}(G)$ and $F \in C^2(\mathbb{R}^m)$. Also, G is the generator of a Feller semi-group on the space $\hat{C}(\mathcal{X})$ of continuous functions vanishing at infinity, known as the *mutation operator*, R is a bounded linear operator from $B(\mathcal{X})$ to $B(\mathcal{X}^2)$, known as the *recombination operator*, and $\sigma \in B_{\mathrm{sym}}(\mathcal{X}^2)$ is called *selection intensity function.* Here $B_{\mathrm{sym}}(\mathcal{X}^2)$ is the set of nonnegative, bounded, symmetric, Borel measurable functions on $\mathcal{X}^2$. We assume throughout the paper that $R \equiv 0$, i.e., there is no recombination in the model.

Ethier and Kurtz (1986) showed that when there is no selection nor recombination, and the mutation operator is

$$Gf(x) = \frac{1}{2}\theta \int [f(z) - f(x)]\nu_0(\mathrm{d}z), \tag{1.1}$$

where $\theta > 0$ and ν_0 is a non atomic probability measure on $\mathcal{X}$, then the stationary distribution of the Fleming-Viot process (in this case often called *neutral diffusion model*) is the Dirichlet process with parameter (θ, ν_0), denoted by Π_{θ,ν_0}. The Dirichlet process, introduced by Ferguson (1973), is defined as follows. Let α be a finite measure on $\mathcal{X}$, endowed with its Borel sigma-algebra $\mathcal{B}(\mathcal{X})$. A random probability measure μ^* on $\mathcal{X}$ is said to be a Dirichlet process with parameter α if for every measurable partition $B_1, \ldots, B_k$ of $\mathcal{X}$, the vector $(\mu^*(B_1), \ldots, \mu^*(B_k))$ has the Dirichlet distribution with parameters $(\alpha(B_1), \ldots, \alpha(B_k))$. A recent contribution by Walker, Hatjispyros and Nicoleris (2007) showed how the neutral diffusion model is strictly related to Bayesian nonparametrics.

Assuming (1.1) holds, define $\phi^m(\mu) = \langle f, \mu^m \rangle$, for $f \in B(\mathcal{X}^m)$ and μ^m being a m-fold product measure, and consider a diploid selection function $\sigma \in B_{\mathrm{sym}}(\mathcal{X}^2)$ and no recombination. Then the generator of the Fleming-Viot process is (cf., e.g., Donnelly and Kurtz (1999))

$$\begin{aligned} &\sum_{i=1}^{m} \langle G_i f, \mu^m \rangle + \frac{1}{2} \sum_{1 \leq k \neq i \leq m} \langle \Phi_{ki} f - f, \mu^m \rangle \\ &+ \sum_{i=1}^{m} \left(\langle \sigma_{i\cdot}(\cdot,\cdot) f, \mu^{m+1} \rangle - \langle \sigma(\cdot,\cdot) \otimes f, \mu^{m+2} \rangle \right), \end{aligned} \tag{1.2}$$

where G_i is (1.1) operating on f as a function of x_i alone, $\Phi_{ki} f$ is the function of $m-1$ variables obtained by setting the ith and the kth variables in f equal, $\sigma_{i\cdot}(\cdot,\cdot)$ denotes $\sigma(x_i, x_{m+1})$, and $\sigma(\cdot,\cdot) \otimes f$ denotes $\sigma(x_{m+1}, x_{m+2}) f(x_1, \ldots, x_m)$. Ethier and Kurtz (1994) showed that in this case the stationary distribution is

$$\Pi(\mathrm{d}\mu) = C e^{\langle \sigma, \mu^2 \rangle} \Pi_{\theta,\nu_0}(\mathrm{d}\mu), \tag{1.3}$$

where C is a constant and $\langle \sigma, \mu^2 \rangle = \iint \sigma(x,y) \mu(\mathrm{d}x)\, \mu(\mathrm{d}y)$.

The purpose of the present work is to extend Walker, Hatjispyros and Nicoleris (2007) and further detail how Bayesian nonparametrics is connected to population genetics, and to the Fleming-Viot diffusion in particular. More specifically, this paper provides an explicit construction of the Fleming-Viot process with diploid selection, whose generator is (1.2), based on ideas borrowed from the Bayesian nonparametric literature. Our arguments present interesting applications of some typically Bayesian models, as Pólya predictive schemes, Gibbs sampling procedures, and Bayesian hierarchical models, to population genetics. The outline of the construction is the following. First a generalisation of the Blackwell-MacQueen Pólya urn scheme is introduced by means of a mixture model and the key predictive distribution is computed. Then we define an $\mathcal{X}^n$-valued Markov jump process, representing the evolution in time of the individuals, whose transitions are based on the new predictive, and it is shown that, with a particular choice for the selection function, the associated process of empirical measures converges weakly to a Fleming-Viot process with diploid fertility selection. By exploiting the properties of the Gibbs sampling algorithm, it is then shown that the stationary distribution of the measure-valued diffusion, known to be (1.3), is the de Finetti measure of the exchangeable variables introduced in the mixture model. In Section 3 some insight into the functional form of $\Pi(\mathrm{d}\mu)$ will be provided from a Bayesian viewpoint. Finally, the Fleming-Viot process with haploid selection is derived as a special case.

2. Conditional Predictive Density

Let $p_n(\mathrm{d}x_1,\ldots,\mathrm{d}x_n)$, for n even (which we assume henceforth), be the exchangeable law associated with the Blackwell-MacQueen urn-scheme (cf. (2.3) below and Blackwell and MacQueen (1973)). Let P_n denote a pairing of $\{1,\ldots,n\}$ such that, given P_n, k is paired with j_k. The distribution of the pairings is assumed to be uniform. We will also use $\tilde{\sigma}_n(x_k,x_{j_k})$, where $\tilde{\sigma}_n\in B_{\mathrm{sym}}(\mathcal{X}^2)$. Then we consider a generalisation of p_n via the function $\tilde{\sigma}_n(x,y)$, by introducing

$$q_n(\mathrm{d}x_1,\ldots,\mathrm{d}x_n,P_n)\propto\ p_n(\mathrm{d}x_1,\ldots,\mathrm{d}x_n)\prod_k\tilde{\sigma}_n(x_k,x_{j_k}).\tag{2.1}$$

In Section 3 a representation of (2.1) in terms of a mixture model is provided. It is clear that integrating out P_n, we obtain an exchangeable law. Removing one element of the vector $(x_1,\ldots,x_n)$, say x_i, then the predictive, jointly with P_n, is

$$q_n(\mathrm{d}x_i,P_n|\,\mathbf{x}_{-i})\propto\ p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\,\tilde{\sigma}_n(x_i,x_{j_i}),$$

where $\mathbf{x}_{-i}$ denotes $(x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$. Now

$$q_n(P_n|\,\mathbf{x}_{-i})\propto\int p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\,\tilde{\sigma}_n(x_i,x_{j_i});$$

and since from (2.1) we can write

$$q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i},P_n)=\frac{p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\;\tilde\sigma_n(x_i,x_{j_i})}{\int p_n(\mathrm{d}w|\,\mathbf{x}_{-i})\;\tilde\sigma_n(w,x_{j_i})},$$

we obtain

$$\begin{aligned}q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})&\propto\sum_{j\neq i}^n q_n(P_n|\,\mathbf{x}_{-i})\frac{p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\;\tilde\sigma_n(x_i,x_j)}{\int p_n(\mathrm{d}w|\,\mathbf{x}_{-i})\;\tilde\sigma_n(w,x_j)}\\&\propto p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\sum_{j\neq i}^n\tilde\sigma_n(x_i,x_j).\end{aligned}$$

Thus we have

$$q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})=\frac{p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\sum_{j\neq i}^n\tilde\sigma_n(x_i,x_j)}{\int p_n(\mathrm{d}w|\,\mathbf{x}_{-i})\sum_{j\neq i}^n\tilde\sigma_n(w,x_j)}.\tag{2.2}$$

When p_n is derived from the Dirichlet process prior, the predictive for x_i is given by the Blackwell-MacQueen urn scheme

$$p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})=\frac{\theta\,\nu_0(\mathrm{d}x_i)+\sum_{k\neq i}\delta_{x_k}(\mathrm{d}x_i)}{\theta+n-1}.\tag{2.3}$$

Then the predictive (2.2) can be written as

$$\begin{aligned}q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})&=\frac{\theta\sum_{j\neq i}^n\tilde\sigma_n(x_i,x_j)\;\nu_0(\mathrm{d}x_i)+\sum_{k\neq i}^n\sum_{j\neq i}^n\;\tilde\sigma_n(x_i,x_j)\;\delta_{x_k}(\mathrm{d}x_i)}{\int\Big(\theta\sum_{j\neq i}^n\tilde\sigma_n(w,x_j)\;\nu_0(\mathrm{d}w)+\sum_{k\neq i}^n\sum_{j\neq i}^n\;\tilde\sigma_n(w,x_j)\;\delta_{x_k}(\mathrm{d}w)\Big)}\\&=\frac{\theta_{n,i}\,\nu_{n,i}(\mathrm{d}x_i)+\sum_{k\neq i}^n\sum_{j\neq i}^n\tilde\sigma_n(x_i,x_j)\,\delta_{x_k}(\mathrm{d}x_i)}{\theta_{n,i}+\sum_{k\neq i}^n\sum_{j\neq i}^n\tilde\sigma_n(x_k,x_j)},\end{aligned}\tag{2.4}$$

where

$$\theta_{n,i}=\theta\int\sum_{j\neq i}^n\tilde\sigma_n(w,x_j)\,\nu_0(\mathrm{d}w),\tag{2.5}$$

$$\nu_{n,i}(\mathrm{d}w)=\frac{\sum_{j\neq i}^n\tilde\sigma_n(w,x_j)\,\nu_0(\mathrm{d}w)}{\int\sum_{j\neq i}^n\tilde\sigma_n(w,x_j)\,\nu_0(\mathrm{d}w)}.\tag{2.6}$$

Expression (2.4) is the transition density of the Markov particle process defined in Section 4, and the full conditional distribution driving the Markov chain constructed via a Gibbs sampler in the next section. Observe that in (2.4) a larger $\tilde\sigma_n$ implies a larger probability for the first coordinate of being selected to update x_i; that means the larger the $\tilde\sigma_n$ the higher the fitness of the individual who is going to have an offspring. Here fitness means the tendency to have the property or quality that makes an individual more likely to give birth. In this framework, this is mathematically modeled by the pair-dependent coefficient $\tilde{\sigma}_n$ that quantifies the degree of possession of that property or, in other words, the level of fitness. In population genetics terms, such a function describes the intensity of fertility selection. When $\tilde{\sigma}_n(x, y) \equiv 1$ for all n we recover the Dirichlet case, that is (2.3).

Note that, from (2.1), it is also possible to derive the distribution of the pairing. Indeed

$$q_n(P_n) \propto \int_{\mathcal{X}^n} p_n(\mathrm{d}x_1 \ldots \mathrm{d}x_n) \prod_k \tilde{\sigma}_n(x_k, x_{j_k}),$$

from which is also clear the key role of the selection function: since a pair with higher fitness will give a higher value of $\tilde{\sigma}_n$, those individuals which are fitter when paired will increase the probability of that specific pair occurring.

3. Gibbs Sampling

The Gibbs sampler (see Geman and Geman (1984) and Gelfand and Smith (1990)) is an iterative procedure that belongs to the more general class of Markov Chain Monte Carlo algorithms, whose use is nowadays wide-spread in the Bayesian literature. The Gibbs sampler can be synthetically outlined as follows. Suppose we want to sample from $f(z_1, \ldots, z_n|\vartheta)$, but this is not feasible. Given a vector of initial values $(z_1^{(0)}, \ldots, z_n^{(0)})$, we sample updates for each z_i from the so-called full conditional distribution $f(z_i|z_1, \ldots, z_{i-1}, z_{i+1}, \ldots, z_n, \vartheta)$, which is typically easier to deal with. This is done iteratively, so that

$$\begin{aligned}
z_1^{(1)} &\sim f(z_1|z_2^{(0)}, \ldots, z_n^{(0)}, \vartheta),\\
z_2^{(1)} &\sim f(z_2|z_1^{(1)}, z_3^{(0)}, \ldots, z_n^{(0)}, \vartheta),\\
&\vdots\\
z_n^{(1)} &\sim f(z_n|z_1^{(1)}, \ldots, z_{n-1}^{(1)}, \vartheta),\\
z_1^{(2)} &\sim f(z_1|z_2^{(1)}, \ldots, z_n^{(1)}, \vartheta),
\end{aligned}$$

and so on. This produces a Markov chain whose stationary distribution is given by the joint distribution $f(z_1, \ldots, z_n|\vartheta)$, so that after a sufficient number of iterations an arbitrarily large sample from the distribution of interest is available (see for example Gelfand and Smith (1990)). Observe that the stationary distribution of a single component z_i is $f(z_i|\vartheta)$. Note also that the result still holds if the coordinates are updated in a random order (called *random scan*), as long as all coordinates are visited infinitely often.

In this section the theoretical properties of the Gibbs sampler are exploited to derive the stationary distribution of a chain of random distribution functions, which will be of great help in Section 4 for the derivation of the stationary distribution of the Fleming-Viot process. We can represent q_n in (2.1) in the following way. We take $\mu \sim \Pi_{\theta,\nu_0}$ and $x_1,\ldots,x_n|\mu$ to be independent and identically distributed according to μ. We then assume there is a conditional density for continuous variables $(y_1,\ldots,y_n)$, with each y_i defined on the real line, whereby

$$\tilde{p}_n(y_1=1,\ldots,y_n=1|x_1,\ldots,x_n,P_n)=\prod_k n\,\tilde{\sigma}(x_k,x_{j_k}),$$

the product being over $n/2$ terms. Hence, $q_n(\mathrm{d}x_1,\ldots,\mathrm{d}x_n,P_n)$ is the conditional law of $(x_1,\ldots,x_n,P_n)$ given $(y_1=1,\ldots,y_n=1)$. Here we are considering an n factor in the product because we are going to take $\tilde{\sigma}$ of order n^{-1}. The reason for this choice will be made clear in Section 4.

Consider now a Gibbs sampler algorithm implemented on $(x_1,\ldots,x_n,\mu)$, where at each iteration $x_1,\ldots,x_n$ are sampled from the full conditionals (2.4), that is $q_n(\mathrm{d}x_i|\,x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n)$, and μ is sampled from the Dirichlet process conditional on $(x_1,\ldots,x_n)$, denoted by $\Pi_{\theta,\nu_0}(\cdot|\,x_1,\ldots,x_n)$. Note that (see Ferguson (1973))

$$\Pi_\alpha(\,\cdot\,|\,x_1,\ldots,x_n)=\Pi_{\alpha+\sum_{i=1}^n\delta_{x_i}}(\,\cdot\,).$$

From the properties of the Gibbs sampler it follows that the stationary distribution of the $\mathcal{X}^n$-valued Markov chain generated by $(x_1,\ldots,x_n)$ is given by $q_n(dx_1\cdots dx_n)$. Furthermore, since

$$\tilde{p}_n(y_1=1,\ldots,y_n=1|\,\mu,P_n)=\left\{n\iint\tilde{\sigma}_n(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\}^{\frac{n}{2}}$$

which does not depend on P_n, the stationary distribution of the chain of random distribution functions is the law of $(\mu|y_1=1,\ldots,y_n=1)$, denoted by $\Pi_n(\mathrm{d}\mu)$, given according to Bayes' theorem by

$$\Pi_n(\mathrm{d}\mu)=\frac{\Pi_{\theta,\nu_0}(\mathrm{d}\mu)\tilde{p}_n(y_1=1,\ldots,y_n=1|\mu)}{\int\Pi_{\theta,\nu_0}(\mathrm{d}\nu)\tilde{p}_n(y_1=1,\ldots,y_n=1|\nu)};$$

i.e., it is proportional to

$$\left\{n\iint\tilde{\sigma}_n(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\}^{\frac{n}{2}}\Pi_{\theta,\nu_0}(\mathrm{d}\mu). \tag{3.1}$$

Note that $\mathscr{P}(\mathcal{X})$ is compact if $\mathcal{X}$ is, in which case $\{\Pi_n,\,n\geq 1\}$ is tight. If we now set $\tilde{\sigma}_n(x,y)=n^{-1}[1+2n^{-1}\sigma(x,y)]$, where $\sigma(x,y)\in B_{\mathrm{sym}}(\mathcal{X}^2)$, we obtain

$$\Pi_n(\mathrm{d}\mu)\propto\left\{1+\iint\frac{1}{n/2}\,\sigma(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\}^{\frac{n}{2}}\Pi_{\theta,\nu_0}(\mathrm{d}\mu)$$

and, taking the limit for $n \to \infty$, yields

$$\Pi_\infty(\mathrm{d}\mu) \propto \exp\left\{\iint \sigma(x,y)\mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\}\Pi_{\theta,\nu_0}(\mathrm{d}\mu).$$

This is the stationary distribution of the chain of random distribution functions when the sample size, i.e. the dimension of the chain, grows to infinity, and is also the de Finetti measure of the sequence $(x_1, x_2, \ldots | y_1 = 1, y_2 = 1, \ldots)$. The de Finetti measure of an infinite exchangeable sequence $(z_1, z_2, \ldots)$ is defined as the unique probability measure Q on the space $\mathscr{P}(\mathcal{X})$ such that, for every $n > 0$,

$$(z_1, \ldots, z_n) \sim \int_{\mathscr{P}(\mathcal{X})} \xi^n(\mathrm{d}\cdot)Q(\mathrm{d}\xi),$$

where ξ is a probability measure sampled from Q and ξ^n denotes a n-fold product measure. In other words, conditional on ξ, the random variables $(z_1, \ldots, z_n)$ are iid ξ. See Aldous (1985) for more details.

4. The Particle Process and the Associated Measure-Valued Process

In this section we construct an $\mathcal{X}^n$-valued Markov particle process based on (2.4) and an associated $\mathscr{P}(\mathcal{X})$-valued process, and show that, in a special case for the function $\tilde{\sigma}_n$, for large n the latter converges in distribution to the Fleming-Viot process with diploid fertility selection.

Consider a vector of n particles, and define the particle process as follows. Instantaneously after each transition, a particle x_i, for $1 \leq i \leq n$, is selected with uniform probability, and a holding time is sampled from an exponential distribution of parameter $\lambda_{n,i} = \lambda_n(x_i)$. At the next transition, the ith particle is replaced with a random sample from (2.4). Since the holding time depends on x_i, which belongs to the current state only, the process is clearly Markovian. Consider now the Markov chain embedded at jump times. Since the transition laws are given by $q_n(\mathrm{d}x_i | \mathbf{x}_{-i})$, the chain is otherwise obtained by implementing a Gibbs sampler on $q_n(\mathrm{d}x_1, \ldots, \mathrm{d}x_n)$, of which (2.4) is the full conditional distribution. This ensures that q_n is the stationary distribution of the $\mathcal{X}^n$-valued chain and, given that between jumps the vector is constant, the continuous time process.

For $f \in B(\mathcal{X}^n)$, the generator of the $\mathcal{X}^n$-valued process is

$$\begin{aligned} A^n f(\mathbf{x}) = &\sum_{i=1}^n \frac{\lambda_{n,i}}{n} \int \left[f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x})\right] \\ &\times \left(\frac{\theta_{n,i}\,\nu_{n,i}(\mathrm{d}y) + \sum_{k\neq i}^n \sum_{j\neq i}^n \tilde{\sigma}_n(y, x_j)\,\delta_{x_k}(\mathrm{d}y)}{\theta_{n,i} + \sum_{k\neq i}^n \sum_{j\neq i}^n \tilde{\sigma}_n(x_k, x_j)}\right), \end{aligned}$$

where $\eta_i(\mathbf{x}|z) = (x_1, \ldots, x_{i-1}, z, x_{i+1}, \ldots, x_n)$, which equals

$$\sum_{i=1}^{n} \frac{\lambda_{n,i}\,\theta_{n,i}}{n\Big(\theta_{n,i} + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n} \tilde{\sigma}_n(x_k, x_j)\Big)} \int [f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x})]\nu_{n,i}(\mathrm{d}y)$$
$$+ \sum_{i=1}^{n}\sum_{k\neq i}^{n}\sum_{j\neq i}^{n} \frac{\lambda_{n,i}\,\tilde{\sigma}_n(x_k, x_j)}{n\Big(\theta_{n,i} + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n} \tilde{\sigma}_n(x_k, x_j)\Big)} [f(\eta_i(\mathbf{x}|x_k)) - f(\mathbf{x})].$$

If we let the Poisson intensity rate be

$$\lambda_{n,i} = \frac{n\Big(\theta_{n,i} + \sum_{k\neq i}^{n}\sum_{j\neq i}^{n} \tilde{\sigma}_n(x_k, x_j)\Big)}{2}, \tag{4.1}$$

we obtain

$$\sum_{i=1}^{n} \frac{1}{2}\theta_{n,i} \int [f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x})]\nu_{n,i}(\mathrm{d}y) + \sum_{1\leq k\neq i\neq j\leq n} \frac{1}{2}\tilde{\sigma}_n(x_k, x_j)[f(\eta_i(\mathbf{x}|x_k)) - f(\mathbf{x})]. \tag{4.2}$$

Consider now the same choice for $\tilde{\sigma}_n$ as in Section 3, i.e.,

$$\tilde{\sigma}_n(x, y) = \frac{1}{n} + \frac{2}{n^2}\,\sigma(x, y), \tag{4.3}$$

where $\sigma \in B_{\mathrm{sym}}(\mathcal{X}^2)$. Substituting (2.5), (2.6), and (4.3) in (4.2) yields

$$\begin{aligned} A^n f(\mathbf{x}) =& \frac{1}{n} \sum_{1\leq j\neq i\leq n} G_i^{n,\sigma_j} f(\mathbf{x}) \\ &+ \frac{1}{2} \sum_{1\leq k\neq i\leq n} [f(\eta_i(\mathbf{x}|x_k)) - f(\mathbf{x})] \\ &+ \frac{1}{n^2} \sum_{1\leq k\neq i\neq j\leq n} \sigma(x_k, x_j)[f(\eta_i(\mathbf{x}|x_k)) - f(\mathbf{x})], \end{aligned}$$

where

$$G^{n,\sigma_j} f(x) = \frac{1}{2}\theta \int [f(y) - f(x)] \left(1 + \frac{2\sigma(y, x_j)}{n}\right) \nu_0(\mathrm{d}y) \tag{4.4}$$

and G_i^{n,σ_j} is the operator G^{n,σ_j} applied to the ith argument of f.

As in Donnelly and Kurtz (1999), define the probability measure $\mu^{(m)}$ on $\mathcal{X}^m$, $m \leq n$, by

$$\mu^{(m)} = \frac{1}{n(n-1)\ldots(n-m+1)} \sum_{1\leq i_1\neq\cdots\neq i_m\leq n} \delta_{(x_{i_1},\ldots,x_{i_m})}. \tag{4.5}$$

Also, for $f \in B(\mathcal{X}^m)$, $m \le n$, define $\phi^{(m)}(\mu) = \langle f, \mu^{(m)} \rangle$ and $\mathbb{A}^n \phi^{(m)}(\mu) = \langle A^n f, \mu^{(m)} \rangle$, where $\langle f, \mu \rangle = \int f \mathrm{d}\mu$. Then the generator for the process of empirical measures in the n-dimensional case is

$$\begin{aligned}\mathbb{A}^n \phi^{(n)}(\mu) = &\frac{1}{n} \sum_{1 \le j \ne i \le n} \langle G_i^{n,\sigma_j} f, \mu^{(n)} \rangle \\ &+ \frac{1}{2} \sum_{1 \le k \ne i \le n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle \\ &+ \frac{1}{n^2} \sum_{1 \le k \ne i \ne j \le n} \langle \sigma_{kj}(\cdot,\cdot)(\Phi_{ki} f - f), \mu^{(n)} \rangle,\end{aligned}$$

where $\sigma_{kj}(\cdot,\cdot)$ denotes $\sigma(x_k, x_j)$, and $\Phi_{ki} f$ is the function of f where the coordinate at level k has replaced the coordinate at level i.

Observe that for $f \in B(\mathcal{X}^m)$, $m \le n$,

$$\sum_{i=m+1}^{n} \sum_{j \ne i}^{n} \langle G_i^{n,\sigma_j} f, \mu^{(n)} \rangle = 0,$$

$$\sum_{i=m+1}^{n} \sum_{k \ne i}^{n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle = 0,$$

$$\sum_{i=m+1}^{n} \sum_{k \ne i}^{n} \sum_{j \ne i}^{n} \langle \sigma_{kj}(\cdot,\cdot)(\Phi_{ki} f - f), \mu^{(n)} \rangle = 0,$$

given that in all cases x_i is not an argument of f and thus f does not change. Further, that

$$\sum_{i=1}^{m} \sum_{k=m+1}^{n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle = 0$$

given that $\langle \Phi_{ki} f, \mu^{(n)} \rangle = \langle f, \mu^{(n)} \rangle$ when x_k is not an argument of f. Hence, when $f \in B(\mathcal{X}^m)$, $m \le n$, we have

$$\begin{aligned}\mathbb{A}^n \phi^{(m)}(\mu) = &\frac{1}{n} \sum_{1 \le j \ne i \le m} \langle G_i^{n,\sigma_j} f, \mu^{(m)} \rangle \\ &+ \frac{n-m}{n} \sum_{i=1}^{m} \langle G_i^{n,\sigma_{m+1}} f, \mu^{(m+1)} \rangle \\ &+ \frac{1}{2} \sum_{1 \le k \ne i \le m} \langle \Phi_{ki} f - f, \mu^{(m)} \rangle \\ &+ \frac{1}{n^2} \sum_{1 \le k \ne i \ne j \le m} \Big(\langle \sigma_{kj}(\cdot,\cdot) \Phi_{ki} f, \mu^{(m)} \rangle - \langle \sigma_{kj}(\cdot,\cdot) f, \mu^{(m)} \rangle \Big)\end{aligned}$$

$$
\begin{aligned}
&+\frac{n-m}{n^2}\sum_{i=1}^{m}\sum_{j\neq i}^{m}\Big(\langle\sigma_{ij}(\cdot,\cdot)f,\mu^{(m)}\rangle-\langle\sigma_{\cdot j}(\cdot,\cdot)f,\mu^{(m+1)}\rangle\Big)\\
&+\frac{n-m}{n^2}\sum_{1\leq k\neq i\leq m}\Big(\langle\sigma_{k\cdot}(\cdot,\cdot)\Phi_{ki}f,\mu^{(m+1)}\rangle-\langle\sigma_{k\cdot}(\cdot,\cdot)f,\mu^{(m+1)}\rangle\Big)\\
&+\frac{(n-m)^2}{n^2}\sum_{i=1}^{m}\Big(\langle\sigma_{i\cdot}(\cdot,\cdot)f,\mu^{(m+1)}\rangle-\langle\sigma(\cdot,\cdot)\otimes f,\mu^{(m+2)}\rangle\Big), \quad (4.6)
\end{aligned}
$$

where we used the notation $\sigma_{h\cdot}(\cdot,\cdot)f=\sigma(x_h,x_{m+1})f(x_1,\ldots,x_m)$ and $\sigma(\cdot,\cdot)\otimes f=\sigma(x_{m+1},x_{m+2})f(x_1,\ldots,x_m)$.

Note that in the fifth term we have σ_{ij}, since the operator Φ_{ki} has replaced the particle at level i in f with x_k, which is the reason for the different dimension of integration. The same applies to the last term.

Given now that (4.4) converges to

$$Gf(x)=\frac{1}{2}\theta\int[f(y)-f(x)]\nu_0(\mathrm{d}y),$$

we have that $\langle G_i^{n,\sigma_{m+1}}f,\mu^{(m+1)}\rangle$ converges to $\langle G_if,\mu^{(m+1)}\rangle=\langle G_if,\mu^{(m)}\rangle$. The last identity is due to the fact that the $(m+1)$-th dimension is referred to an argument of σ in (4.4), which vanishes in the limit.

Since in addition, for large n, $\mu^{(m)}$ is essentially product measure, the limiting operator is

$$
\begin{aligned}
\mathbb{A}\phi^m(\mu)=&\sum_{i=1}^{m}\langle G_if,\mu^m\rangle+\frac{1}{2}\sum_{1\leq k\neq i\leq m}\langle\Phi_{ki}f-f,\mu^m\rangle\\
&+\sum_{i=1}^{m}\left(\langle\sigma_{i\cdot}(\cdot,\cdot)f,\mu^{m+1}\rangle-\langle\sigma(\cdot,\cdot)\otimes f,\mu^{m+2}\rangle\right), \quad (4.7)
\end{aligned}
$$

where $\phi^m(\mu)=\langle f,\mu^m\rangle$ for $f\in B(\mathcal{X}^m)$. $\mathbb{A}\phi^m(\mu)$ is the generator of the Fleming-Viot process with diploid fertility selection (cf. (1.2)). The above computation implies that the measure-valued process, constructed by means of the generalised Pólya urn scheme, converges in distribution to the Fleming-Viot process with fertility selection.

Theorem 4.1. *Let $\mathcal{X}$ be a compact Polish space, $f\in B(\mathcal{X}^m)$ for $m\leq n$, $\mu^{(m)}$ be as in* (4.5), *and $\sigma\in B_{\mathrm{sym}}(\mathcal{X}^2)$. Let the mutation operator G^{n,σ_j} be defined by* (4.4) *and $\Phi_{ki}:\mathcal{X}^n\to\mathcal{X}^{n-1}$ be defined, for $k,i\leq n$, by*

$$\Phi_{ki}f(x_1,\ldots,x_n)=f(x_1,\ldots,x_{i-1},x_k,x_{i+1},\ldots,x_n).$$

Let $\{\mu_n(t)\}_{t>0}$ be a Markov process with sample paths in $D_{\mathscr{P}(\mathcal{X})}([0,\infty))$ and with generator (4.6). *Let $\{\mu(t)\}_{t>0}$ be the Fleming-Viot process with generator* (4.7). *Then $\mu_n\Rightarrow\mu$ in the Skorohod topology.*

Proof. Given the convergence of (4.6) to (4.7), the result follows by an application of Theorem 1.6.1 and 4.2.5 of Ethier and Kurtz (1986), together with the observation that the martingale problem for (4.7) is well-posed (cf., Ethier and Kurtz (1993)).

Observe that for $\sigma \equiv 0$, i.e. when there is no selection, (4.7) reduces to the generator of the neutral diffusion model whose stationary distribution is the Dirichlet process. This is consistent with the generalisation of the predictive distribution described in Section 3, which for $\sigma \equiv 0$ reduces to the the Blackwell-MacQueen case.

5. Stationary Distribution

As stated in the Introduction, it was shown by Ethier and Kurtz (1994) that the measure-valued process with generator (4.7) has stationary distribution given by (1.3). In this section we provide a different proof of this result, based on the construction of the previous sections. In particular, in Section 3 the use of the Gibbs sampler enabled us to elicit the stationary distribution of the chain of random distribution functions. What remains to do is to connect the de Finetti measure of the sequence with the empirical measure of the particles when the population size goes to infinity.

Theorem 5.2. *Let $\mathscr{X}$ be a compact Polish space, and let $\{\mu(t)\}_{t>0}$ be the Fleming-Viot process with generator* (4.7)*. Then*

$$\Pi_\infty(\mathrm{d}\mu) = C \,\exp\left\{\iint \sigma(x,y)\,\mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\} \Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{5.1}$$

is the stationary distribution of $\{\mu(t)\}_{t>0}$, where C is a constant and Π_{θ,ν_0} denotes the Dirichlet process with parameters (θ,ν_0).

Proof. In Section 4 we showed that $q_n(\mathrm{d}x_1,\ldots,\mathrm{d}x_n)$ is the stationary distribution of the $\mathscr{X}^n$-valued particle process. Then, for fixed $t \geq 0$ in steady state of the particle process, from Section 3 it follows that $(x_1,\ldots,x_n|\,\mu, y_1 = 1,\ldots,y_n = 1)$ are i.i.d. μ with $\mu \sim \Pi_n$, where Π_n is proportional to (3.1). Hence the limit as n tends to infinity of $n^{-1}\sum_{i=1}^n \delta_{x_i}$ has distribution Π_∞ (see, for example, Aldous (1985)). That is, the limiting distribution of the empirical measure of the particles $(x_1,\ldots,x_n)$ is given by the de Finetti measure of the infinite exchangeable sequence $(x_1,x_2,\ldots)$ conditional on $(y_1 = 1, y_2 = 1,\ldots)$. Since the particles are in steady state, this holds for every $s \geq t$.

Furthermore, the $C_{\mathscr{P}(\mathscr{X})}[0,\infty)$ martingale problem for $\mathbb{A}$ is well posed (cf., Ethier and Kurtz (1993)), that is, $\mathbb{A}$ characterises a unique solution. In this case Lemma 4.9.1 of Ethier and Kurtz (1986) states that if the limiting process has the same distribution at all instants, this is a stationary distribution. Uniqueness follows from Ethier and Kurtz (1994).

Clearly, when $\sigma(x,y) \equiv 0$ we recover the Dirichlet process.

6. Haploid Case

In this section we show how the parameters of the model simplify in the special case of haploid selection. The Fleming-Viot process with haploid fertility selection has generator given by

$$\sum_{i=1}^{m} \langle G_i f, \mu^m \rangle + \frac{1}{2} \sum_{1 \le k \ne i \le m} \left(\langle \Phi_{ki} f, \mu^{m-1} \rangle - \langle f, \mu^m \rangle \right)$$
$$+ \sum_{i=1}^{m} \left(\langle \sigma_i(\cdot) f, \mu^m \rangle - \langle \sigma(\cdot) \otimes f, \mu^{m+1} \rangle \right), \tag{6.1}$$

where $\sigma_i(\cdot)$ denotes $\sigma(x_i)$ and $\sigma(\cdot) \otimes f$ denotes $\sigma(x_{m+1}) f(x_1, \ldots, x_m)$ (cf., Donnelly and Kurtz (1999)). Its stationary distribution is

$$\Pi(d\mu) = C e^{2\langle \sigma, \mu \rangle} \Pi_{\theta, \nu_0}(d\mu), \tag{6.2}$$

where $\langle \sigma, \mu \rangle = \int \sigma(x) \mu(\mathrm{d}x)$. Note that (6.2) is a special case of (1.3), when $\sigma(x,y) = \sigma(x) + \sigma(y)$. See also Ethier and Shiga (2000).

A construction analogous to that exposed so far can be done starting from

$$p_n(y_1 = 1, \ldots, y_n = 1 | \, x_1, \ldots, x_n) = \prod_{i=1}^{n} \tilde{\sigma}_n(x_i)$$

together with $x_1, \ldots, x_n | \, \mu \sim^{iid} \mu$ and $\mu \sim \Pi_{\theta, \nu_0}$. Proceeding as in Section 2 we obtain

$$q_n(\mathrm{d}x | x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_n) = \frac{\theta_n \, \nu_n(\mathrm{d}x) + \sum_{i \ne j}^{n} \tilde{\sigma}_n(x_i) \, \delta_{x_i}(\mathrm{d}x)}{\theta_n + \sum_{i \ne j}^{n} \tilde{\sigma}_n(x_i)}, \tag{6.3}$$

where $\theta_n = \theta \int \tilde{\sigma}_n(x) \nu_0(\mathrm{d}x)$ and $\nu_n(\mathrm{d}x) = \tilde{\sigma}_n(x) \nu_0(\mathrm{d}x) [\int \tilde{\sigma}_n(x) \nu_0(\mathrm{d}x)]^{-1}$. Defining a Markov particle process as in Section 4, where now the transition law for the new particle is given by (6.3), and letting $\lambda_{n,i} = 2^{-1} n (\theta_n + \sum_{k \ne i} \tilde{\sigma}_n(x_k))$, we obtain

$$\sum_{i=1}^{n} \frac{1}{2} \theta_n \int [f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x})] \nu_n(\mathrm{d}y) + \sum_{1 \le k \ne i \le n} \frac{1}{2} \tilde{\sigma}_n(x_i) [f(\eta_i(\mathbf{x}|x_k)) - f(\mathbf{x})].$$

Consider the particular choice $\tilde{\sigma}_n(x) = 1 + n^{-1} 2\sigma(x)$, where σ is a bounded nonnegative measurable function on $\mathcal{X}$; this yields

$$A^n f(\mathbf{x}) = \sum_{i=1}^{n} \frac{1}{2} \theta \int [f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x})] \{1 + \frac{2\sigma(y)}{n}\} \nu_0(\mathrm{d}y)$$

$$+\frac{1}{2}\sum_{1\le k\ne i\le n}[f(\eta_i(\mathbf{x}|x_k))-f(\mathbf{x})]+\frac{1}{n}\sum_{1\le k\ne i\le n}\sigma(x_k)[f(\eta_i(\mathbf{x}|x_k))-f(\mathbf{x})].$$

Proceeding as in Section 4 we can derive the generator for the process of the empirical measures in the n-dimensional case through

$$\begin{aligned}\mathbb{A}^n\phi^{(n)}(\mu)=&\sum_{i=1}^{n}\langle G_i^n f,\mu^{(n)}\rangle+\frac{1}{2}\sum_{1\le k\ne i\le n}\langle\Phi_{ki}f-f,\mu^{(n)}\rangle\\&+\sum_{1\le k\ne i\le n}\frac{1}{n}\sigma_k(\cdot)\langle\Phi_{ki}f-f,\mu^{(n)}\rangle,\end{aligned}$$

where $G^n f(x)=(1/2)\theta\int[f(z)-f(x)]\{1+2\sigma(z)/n\}\nu_0(\mathrm{d}z)$. When $f\in B(S^m)$, $m<n$,

$$\begin{aligned}\mathbb{A}^n\phi^{(m)}(\mu)=&\sum_{i=1}^{m}\langle G_i^n f,\mu^{(m)}\rangle+\frac{1}{2}\sum_{1\le k\ne i\le m}\langle\Phi_{ki}f-f,\mu^{(m)}\rangle\\&+\frac{1}{n}\sum_{1\le k\ne i\le m}\Big(\langle\sigma_k(\cdot)\Phi_{ki}f,\mu^{(m)}\rangle-\langle\sigma_k(\cdot)f,\mu^{(m)}\rangle\Big)\\&+\frac{n-m}{n}\sum_{i=1}^{m}\Big(\langle\sigma_i(\cdot)f,\mu^{(m)}\rangle-\langle\sigma(\cdot)\otimes f,\mu^{(m+1)}\rangle\Big).\end{aligned}$$

The limiting operator is then

$$\begin{aligned}\mathbb{A}\phi^{m}(\mu)=&\sum_{i=1}^{m}\langle G_i f,\mu^{m}\rangle+\frac{1}{2}\sum_{1\le k\ne i\le m}\langle\Phi_{ki}f-f,\mu^{m}\rangle\\&+\sum_{i=1}^{m}\big(\langle\sigma_i(\cdot)f,\mu^{m}\rangle-\langle\sigma(\cdot)\otimes f,\mu^{m+1}\rangle\big),\end{aligned}\tag{6.4}$$

where G^n has been replaced by G, defined in (1.1).

Following an analogous procedure to that used in the proof of Theorem 5.2, one can show that the stationary distribution of the Fleming-Viot process with generator (6.4) is (6.2), as we know from Ethier and Kurtz (1994) and Ethier and Shiga (2000).

Acknowledgement

M. Ruggiero is indebted to the Department of Statistics and Applied Mathematics, University of Turin, and to the Institute of Mathematics, Statistics and Actuarial Science, University of Kent, where the research was carried out. M. Ruggiero was supported by the CRT Foundation, Turin, and by MIUR, grant 2006/133449.

References

Aldous, D. (1985). Exchangeability and related topics. *Ecole d'Eté de Probabilités de Saint Flour XIII. Lecture notes in Math.* **1117**. Springer, Berlin.

Blackwell, D. and MacQueen, J. B. (1973). Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353-355.

Donnelly, P. and Kurtz, T. G. (1999). Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.* **9**, 1091-1148.

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

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

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

Ethier, S. N. and Shiga, T. (2000). A Fleming-Viot process with unbounded selection. *J. Math. Kyoto Univ.* **40**, 337-361.

Ferguson, T.S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209-230.

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

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398-409.

Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Patt. Anal. Mach. Intelligence* **6**, 721–740.

Walker, S. G., Hatjispyros, S. J. and Nicoleris, T. (2007). A Fleming-Viot process and Bayesian nonparametrics. *Ann. Appl. Probab.* **17**, 67-80.

Dipartimento di Economia Politica e Metodi Quantitativi, Università degli Studi di Pavia, Via San Felice 5, 27100 Pavia, Italy.

E-mail: matteo.ruggiero@unipv.it

Institute of Mathematics, Statistics and Actuarial Science, University of Kent, CT2 7NZ Canterbury, UK.

E-mail: S.G.Walker@kent.ac.uk

(Received September 2007; accepted January 2008)