INTERNATIONAL CENTRE FOR ECONOMIC RESEARCH

ICER

WORKING PAPER SERIES

Stephen G. Walker and Matteo Ruggiero

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

Working Paper no. 13/2007
April 2007

APPLIED MATHEMATICS AND QUANTITATIVE METHODS
WORKING PAPER SERIES

Bayesian Nonparametric Construction of the Fleming-Viot Process with Fertility Selection

Stephen G. Walker
Institute of Mathematics, Statistics and Actuarial Science
University of Kent
CT2 7NZ Canterbury, UK
e-mail: S.G.Walker@kent.ac.uk

Matteo Ruggiero[1]
Department of Statistics and Applied Mathematics
University of Turin
Piazza Arbarello 8, 10122 Torino, Italy
e-mail: matteo.ruggiero@unito.it

Abstract

This paper provides the construction in a Bayesian setting of the Fleming-Viot measure-valued process 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.

Keywords: Fleming-Viot process; Measure-valued process; Fertility selection; Gibbs sampler; Dirichlet process mixture model; Blackwell-MacQueen urn-scheme.

AMS Subject Classification: Primary 60G57, 60J35; secondary 60J60, 92D15.

[1]Matteo Ruggiero is supported by CRT Foundation, Turin, and by MIUR, grant 2006/133449

1 Introduction and preliminaries

The Fleming-Viot process, introduced by Fleming and Viot (1979), is a diffusion on the space $\mathscr{P}(E)$ of Borel probability measures on E, endowed with the topology of weak convergence, where E 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 (cf. Ethier and Kurtz, 1993)

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

where

$$\frac{\partial\phi(\mu)}{\partial\mu(x)}=\lim_{\varepsilon\to 0^+}\varepsilon^{-1}\{\phi(\mu+\varepsilon\delta_x)-\phi(\mu)\}$$

and we take the domain $\mathscr{D}(\mathbb{A})$ to be the set of all $\phi\in B(\mathscr{P}(E))$, where $B(\mathscr{P}(E))$ is the set of bounded functions on $\mathscr{P}(E)$, of the form

$$\phi(\mu)=F(\langle f_1,\mu\rangle,\dots,\langle f_m,\mu\rangle)=F(\langle\mathbf{f},\mu\rangle),$$

where $\langle f,\mu\rangle=\int f\mathrm{d}\mu$ and, for $m\geq 1$, $f_1,\dots,f_m\in\mathscr{D}(A)$ and $F\in C^2(\mathbb{R}^m)$. Also, G is the generator of a Feller semi-group on the space $\hat{C}(E)$ of continuous functions vanishing at infinity, known as the *mutation operator*, R is a bounded linear operator from $B(E)$ to $B(E^2)$, known as the *recombination operator*, and $\sigma\in B_{\mathrm{sym}}(E^2)$ is called *selection intensity function*. We assume throughout the paper that $R\equiv 0$, i.e. there is no recombination in the model.

Ethier and Kurtz (1994) showed that when there is no selection nor recombination, and the mutation operator is

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

where $\theta>0$ and ν_0 is a non atomic probability measure, 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}. A recent contribution by Walker *et al.* (2007) showed how the neutral diffusion model is strictly related to Bayesian nonparametrics.

Assuming (1) holds, let $\phi(\mu) = \langle f, \mu^m \rangle$, for $f \in B(E^m)$, where μ^m denotes a m-fold product measure, and consider a diploid selection function $\sigma \in B_{\mathrm{sym}}(E^2)$ and no recombination. Then the generator of the Fleming-Viot process is

$$\frac{1}{2}\sum_{1\le i\neq j\le m}\big(\langle \Phi_{ij}f, \mu^{m-1}\rangle - \langle f, \mu^m\rangle\big) + \sum_{j=1}^{m}\langle G_j f, \mu^m\rangle \qquad (2)$$
$$+\sum_{j=1}^{m}\big(\langle \sigma_{j\cdot}(\cdot,\cdot)f, \mu^{m+1}\rangle - \langle \sigma(\cdot,\cdot)\otimes f, \mu^{m+2}\rangle\big)$$

where G_j is G operating on f as a function of x_j alone, $\Phi_{ij}f$ is the function of $m-1$ variables obtained by setting the ith and the jth variables in f equal, $\sigma_{j\cdot}(\cdot,\cdot)$ denotes $\sigma(x_j, x_{m+1})$ 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

$$\Pi(\mathrm{d}\mu) = Ce^{\langle \sigma, \mu^2\rangle}\Pi_{\theta,\nu_0}(\mathrm{d}\mu) \qquad (3)$$

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

The purpose of the present work is to extend Walker *et al.* (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 the construction in a Bayesian setting of the Fleming-Viot process with diploid selection, whose generator is given by (2) and whose stationary distribution is known to be (3). The outline of the construction is the following. First, a hierarchical model in the Bayesian nonparametric framework is introduced, and the key predictive density, which generalises the Blackwell-MacQueen Pólya urn scheme (Blackwell and MacQueen, 1973), is computed. Then an E^n-valued Markov jump process is constructed, whose transitions are based on the new predictive density, and it is shown that, with a particular choice for the selection function, the associated process of empirical measures converges to a Fleming-Viot process with diploid fertility selection. Finally, by means of Gibbs sampling techniques, it is shown that the stationary distribution of the measure-valued diffusion is the de Finetti measure of the exchangeable variables introduced in the hierarchical model. The Fleming-Viot process with haploid selection is then derived as a special case.

2 The underlying model

For each even n, let P_n denote a pairing of $\{1,\dots,n\}$, such that given P_n, k is paired with j_k.

Consider the Dirichlet process mixture model, introduced by Lo (1984), in the following setting:

$$\tilde{p}_n(y_1=1,\dots,y_n=1|\, x_1,\dots,x_n,P_n) = \prod_k n\,\tilde{\sigma}_n(x_k,x_{j_k}) \tag{4}$$

$$x_1,\dots,x_n|\,\mu \overset{i.i.d.}{\sim} \mu$$

$$\mu \sim \Pi_{\theta,\nu_0}$$

$$P_n \sim \pi(P_n) \propto 1.$$

The product in (4), which is taken over $n/2$ terms that cover all pairs, is to be regarded as the density of the vector $(y_1,\dots,y_n)$ computed at $(1,\dots,1)$, conditionally on $(x_1,\dots,x_n)$ and on the pairing. The bounded symmetric function $\tilde{\sigma}_n(x,y)\in B_{\rm sym}(E^2)$ is assumed to be chosen for each n. The vector $(x_1,\dots,x_n)$ is exchangeable, that is $x_1,\dots,x_n$ are i.i.d. μ conditionally on μ, and μ is a random distribution function distributed as a Dirichlet process with parameters (θ,ν_0), denoted by Π_{θ,ν_0}. Last, $\pi(P_n)$ is the distribution of the pairing, assumed to be uniform.

It is well known that given a sample of size $n-1$ from a random distribution function which is a Dirichlet process, the predictive density for the next observation is

$$p_n(\mathrm{d}x_n|\, x_1,\dots,x_{n-1}) = \frac{\theta\,\nu_0(\mathrm{d}x_n)+\sum_{k=1}^{n-1}\delta_{x_k}(\mathrm{d}x_n)}{\theta+n-1}$$

which is known as the Blackwell-MacQueen urn-scheme (see Blackwell and MacQueen, 1973). For this reason the notation $p_n(x_1,\dots,x_n)$ will be used to denote the unconditional joint density of $x_1,\dots,x_n$.

We are interested in constructing a Markov process which converges to the Fleming-Viot process with diploid selection, and derive its stationary distribution. Given the hierarchical model, in the next section a predictive distribution for the x's conditionally on the y's will be derived, which will be the transition density of the E^n-valued particle process we will define in Section 5. The hierarchical model will also play a central role in the derivation of the de Finetti measure of the infinite exchangeable sequence $x_1, x_2, \dots$ conditionally on $y_1, y_2, \dots$, in that the key step will be Gibbs sampling the joint law of the x's and μ, conditionally on the y's, which will be done in Section 4. Finally it will remain to show that the so found de Finetti measure is the distribution, when n tends to infinity and for fixed $t \geq 0$, of the limiting empirical measure of the particle forming the E^n-valued Markov process, and is also the stationary distribution of the measure-valued process associated with the constructed particle process; that is of the Fleming-Viot process with diploid selection.

3 Conditional predictive density

Let $p_n(x_1, \dots, x_n)$ be the exchangeable density associated with the Blackwell-MacQueen urn-scheme as above. The hierarchical model induces a generalisation of p_n via the functions $\tilde{\sigma}_n(x, y)$, by writing

$$q_n(x_1, \dots, x_n, P_n | \, \mathbf{y} = \mathbf{1}) \propto \; p_n(x_1, \dots, x_n) \prod_k \tilde{\sigma}_n(x_k, x_{j_k}) \tag{5}$$

where $\mathbf{y} = \mathbf{1}$ denotes $(y_1 = 1, \dots, y_n = 1)$, the conditioning on which will be from now on implicit. Removing one element of the vector $x_1, \dots, x_n$, 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 $(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$. 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 (5) we can write

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

we obtain

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

Thus we have

$$q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i}) = \frac{p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)}{\int p_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)}. \tag{6}$$

When p_n is derived from the Dirichlet process prior, the predictive for x_i is

$$p_n(\mathrm{d}x|\,\mathbf{x}_{-i}) = \frac{\theta\,\nu_0(\mathrm{d}x) + \sum_{k\neq i}^n \delta_{x_k}(\mathrm{d}x)}{\theta + n - 1} \tag{7}$$

and the predictive (6) can be written,

$$\begin{aligned} q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i}) &= \frac{\theta\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i) + \sum_{k\neq i}^n\sum_{j\neq i}^n\,\tilde{\sigma}_n(x_i,x_j)\,\delta_{x_k}(\mathrm{d}x_i)}{\int\Big(\theta\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i) + \sum_{k\neq i}^n\sum_{j\neq i}^n\,\tilde{\sigma}_n(x_i,x_j)\,\delta_{x_k}(\mathrm{d}x_i)\Big)} \\ &= \frac{\theta_n\,\nu_n(\mathrm{d}x_i) + \sum_{k\neq i}^n\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\delta_{x_k}(\mathrm{d}x_i)}{\theta_n + \sum_{k\neq i}^n\sum_{j\neq i}^n \tilde{\sigma}_n(x_k,x_j)} \end{aligned} \tag{8}$$

where θ_n and ν_n denote

$$\theta_n = \theta\int\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i) \tag{9}$$

and

$$\nu_n(\mathrm{d}x_i) = \frac{\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i)}{\int\sum_{j\neq i}^n \tilde{\sigma}_n(x_i,x_j)\,\nu_0(\mathrm{d}x_i)}. \tag{10}$$

Expression (8) will be the transition density of the Markov particle process, of Section 5, and the full conditional distribution driving the Markov chain constructed via a Gibbs sampler in the next section.

Observe in (8) that a larger $\tilde{\sigma}_n$ implies a larger probability for the 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. 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 (7).

Note that from (5) it is also possible to derive the distribution of the pairing. Indeed

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

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

4 Gibbs sampling the model

Consider now a Gibbs sampler algorithm (see Gelfand and Smith, 1990) implemented on

$$(x_1, \dots, x_n, \mu)$$

where at each iteration $x_1, \dots, x_n$ are sampled from the full conditionals (8), that is $q_n(\mathrm{d}x_i | \mathbf{x}_{-i})$, and μ is sampled from the Dirichlet process conditional on $(x_1, \dots, x_n)$, denoted by $\Pi_{\theta,\nu_0}(\cdot | x_1, \dots, x_n)$.

The stationary distribution of the E^n-valued Markov chain generated by $(x_1, \dots, x_n)$ is given by $q_n(x_1, \dots, x_n)$. Further, since

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

which does not depend on P_n, the stationary distribution of the chain of random distribution functions is

$$\Pi_n(\mathrm{d}\mu) \propto \left\{ n \iint \tilde{\sigma}_n(x, y)\mu(\mathrm{d}x)\mu(\mathrm{d}y) \right\}^{n/2} \Pi_{\theta,\nu_0}(\mathrm{d}\mu). \tag{11}$$

Note that $\mathscr{P}(E)$ with Prohorov's metric is separable and complete, so that $\{\Pi_n,\, n \geq 1\}$ is tight. If we now put

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

and take the limit as $n \to \infty$, we obtain

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

which is the stationary distribution of the chain of random distribution functions when the sample size grows to infinity, and is also the de Finetti measure of the infinite exchangeable sequence $(x_1, x_2, \dots)$.

5 The particle process and the associated measure-valued process

In this section we construct an E^n-valued Markov particle process based on (8) and an associated $\mathscr{P}(E)$-valued process, derive their respective generators in a special case for the function $\tilde{\sigma}_n$, and show that in the limit for large n the latter converges to the generator of the Fleming-Viot process with diploid fertility selection.

Consider a vector of n particles. Instantaneously after each transition, a particle x_i, for $1 \le i \le n$, 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 (8). Since the holding time depends on x_i which belongs to the current state only, the process is clearly Markov. Note that there is a Markov chain embedded at jump times, and since the transition densities are given by $q_n(\mathrm{d}x_i|\,\mathbf{x}_{-i})$, the chain is otherwise obtained by implementing a Gibbs sampler on $q_n(x_1,\ldots,x_n)$, of which (8) is the full conditional distribution. This ensures that q_n is the stationary distribution of either the E^n-valued chain and, given that between jumps the vector is constant, the process.

The generator of the E^n-valued process is

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

where $\eta_i(\mathbf{x}|z) = (x_1,\ldots,x_{i-1},z,x_{i+1},\ldots,x_n)$. If we let the Poisson rate be

$$\lambda_{n,i} = \frac{n\left(\theta_n + \sum_{k\neq i}^n \sum_{j\neq i}^n \tilde{\sigma}_n(x_k,x_j)\right)}{2} \tag{12}$$

we obtain

$$\sum_{i=1}^n \frac{1}{2}\,\theta_n \int [f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x})]\nu_n(\mathrm{d}y) + \sum_{1\le k\neq i\neq j\le n} \frac{1}{2}\,\tilde{\sigma}_n(x_k,x_j)[f(\eta_i(\mathbf{x}|x_k)) - f(\mathbf{x})]. \tag{13}$$

Consider now a particular choice for $\tilde{\sigma}_n$ given earlier, i.e.

$$\tilde{\sigma}_n(x,y) = \frac{1}{n} + \frac{2}{n^2}\,\sigma(x,y) \tag{14}$$

where σ is a bounded symmetric function on E^2. Note that if $\sigma(x,y) \equiv 0$, then (9) reduces to θ, (10) to ν_0 and $\lambda_{n,i}$ to $n(\theta+n-1)/2$, as in the neutral case (cf. Walker *et al.*, 2007). Using also (9), (10) and (14) in (13) yields

$$\begin{aligned}
A^n f(\mathbf{x}) = &\sum_{i=1}^{n}\sum_{j\neq i}^{n}\frac{1}{2}\theta\int[f\eta_i(\mathbf{x}|y)-f(\mathbf{x})]\left(\frac{1}{n}+\frac{2\sigma(y,x_j)}{n^2}\right)\nu_0(\mathrm{d}y)\\
&+\frac{1}{2n}\sum_{1\leq k\neq i\neq j\leq n}[f\eta_i(\mathbf{x}|x_k)-f(\mathbf{x})]\\
&+\frac{1}{n^2}\sum_{1\leq k\neq i\neq j\leq n}\sigma(x_k,x_j)[f\eta_i(\mathbf{x}|x_k)-f(\mathbf{x})]\\
=&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}G_i^{n,\sigma_j}f(\mathbf{x})\\
&+\frac{1}{2}\sum_{1\leq k\neq i\leq n}[f\eta_i(\mathbf{x}|x_k)-f(\mathbf{x})]\\
&+\frac{1}{n^2}\sum_{1\leq k\neq i\neq j\leq n}\sigma(x_k,x_j)[f\eta_i(\mathbf{x}|x_k)-f(\mathbf{x})]
\end{aligned}$$

where

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

and G_i^{n,σ_j} is the operator G^{n,σ_j} applied to the i-th coordinate.

As in Donnelly and Kurtz (1999), define now for $m\leq n$ the probability measure on E^m

$$\mu^{(m)}=\frac{1}{n(n-1)\dots(n-m+1)}\sum_{1\leq i_1\neq\dots\neq i_m\leq n}\delta_{(x_{i_1},\dots,x_{i_m})}$$

and for $f\in B(E^n)$

$$\phi(\mu)=\langle f,\mu^{(n)}\rangle$$

and

$$\mathbb{A}^n\phi(\mu)=\langle A^n f,\mu^{(n)}\rangle$$

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

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

where $\sigma_{k,j}(\cdot,\cdot)$ 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 now that for $f \in B(E^m)$, $m < n$,

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

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

and

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

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

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

given that

$$\langle \Phi_{ki} f, Z_n^{(n)} \rangle = \langle f, Z_n^{(n)} \rangle$$

when x_k is not an argument of f. Hence, when $f \in B(E^m)$, $m < n$,

$$\begin{aligned}
\mathbb{A}^n \phi(\mu) = & \frac{1}{n} \sum_{i=1}^{m} \sum_{j \neq i}^{m} \langle G_i^{n,\sigma_j} f, \mu^{(m)} \rangle \\
& + \frac{n-m}{n} \sum_{i=1}^{m} \langle G_i^{n,\sigma_{m+1}} f, \mu^{(m+1)} \rangle \\
& + \frac{1}{2} \sum_{1 \leq k \neq i \leq m} \langle \Phi_{ki} f - f, \mu^{(m)} \rangle \\
& + \frac{1}{n^2} \sum_{1 \leq k \neq i \neq j \leq m} \Big(\langle \sigma_{kj}(\cdot,\cdot) \Phi_{ki} f, \mu^{(m)} \rangle - \langle \sigma_{kj}(\cdot,\cdot) f, \mu^{(m)} \rangle \Big) \\
& + \frac{n-m}{n^2} \sum_{i=1}^{m} \sum_{j \neq i}^{m} \Big(\langle \sigma_{ij}(\cdot,\cdot) f, \mu^{(m)} \rangle - \langle \sigma_{\cdot j}(\cdot,\cdot) f, \mu^{(m+1)} \rangle \Big) \\
& + \frac{n-m}{n^2} \sum_{1 \leq k \neq i \leq m} \Big(\langle \sigma_{k\cdot}(\cdot,\cdot) \Phi_{ki} f, \mu^{(m+1)} \rangle - \langle \sigma_{k\cdot}(\cdot,\cdot) f, \mu^{(m+1)} \rangle \Big) \\
& + \frac{(n-m)^2}{n^2} \sum_{i=1}^{m} \Big(\langle \sigma_{i\cdot}(\cdot,\cdot) f, \mu^{(m+1)} \rangle - \langle \sigma(\cdot,\cdot) \otimes f, \mu^{(m+2)} \rangle \Big)
\end{aligned} \tag{15}$$

where

$$\sigma_{h\cdot}(\cdot,\cdot)f = \sigma(x_h, x_{m+1})f(x_1,\dots,x_m)$$

and

$$\sigma(\cdot,\cdot)\otimes f = \sigma(x_{m+1}, x_{m+2})f(x_1,\dots,x_m).$$

Note that in the fifth term we have σ_{ij} since with the operator Φ_{ki} the particle at level i in f is now x_k, which is also the difference between the two f in the last term, which justifies the different dimension of integration.

Given now that

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

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

$$\langle G_i f, \mu^{(m+1)}\rangle = \langle G_i f, \mu^{(m)}\rangle$$

due to the fact that the $(m+1)$-th dimension in (16) vanishes in the limit.

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

$$\begin{aligned}\mathbb{A}\phi(\mu) = &\sum_{i=1}^{m}\langle G_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\le k\ne i\le m}\langle \Phi_{ki}f - f, \mu^m\rangle \\ &+ \sum_{i=1}^{m}\Big(\langle \sigma_{i\cdot}(\cdot,\cdot)f, \mu^{(m+1)}\rangle - \langle \sigma(\cdot,\cdot)\otimes f, \mu^{(m+2)}\rangle\Big)\end{aligned} \tag{16}$$

which is the generator of the Fleming-Viot process with diploid fertility selection. For relative compactness conditions and weak convergence see Donnelly and Kurtz (1999).

Observe that for $\sigma \equiv 0$, i.e. when there is no selection, (16) reduces to the generator of the neutral Fleming-Viot, whose stationary distribution is the Dirichlet process. This is coherent with the generalisation of the predictive distribution described in Section 3, since $\sigma \equiv 0$ reduces the new predictive to the the Blackwell-MacQueen case.

6 Stationary distribution

As stated in the introduction, it was shown by Ethier and Kurtz (1994) that the measure-valued process with generator (16) has stationary distribution given by (3). In this section we provide a different proof of this result, based on the construction of the previous sections. In particular, in Section 4 the use of the Gibbs sampler together with the hierarchical framework introduced in Section 2, 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 particle, when the population size goes to infinity.

Theorem 6.1. *Let E be a locally compact complete separable metric space, and let $\{\mu_t, t \geq 0\}$ be the Fleming-Viot process on $\mathscr{P}(E)$ with generator given by* (16)*. Then*

$$\Pi_\infty(\mathrm{d}\mu) = C \exp\left\{\int_{E^2} \sigma(x,y)\, \mu(\mathrm{d}x)\mu(\mathrm{d}y)\right\} \Pi_{\theta,\nu_0}(\mathrm{d}\mu) \tag{17}$$

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

Proof. Since the transition density of the E^n-valued particle process is given by (8), it follows that $(x_1, \ldots, x_n | \, \mu, y_1 = 1, \ldots, y_n = 1)$ are i.i.d. μ and $\mu \sim \Pi_n$, where Π_n is (11), from which

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i}$$

has distribution Π_∞ (see for example Aldous, 1985). That is, for fixed $t \geq 0$ the limiting distribution for large n 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 $(y_1 = 1, y_2 = 1, \ldots)$. Also, as noted in Section 5, the measure valued process is constant between two consecutive jumps of the particle process, and this implies that Π_n is the de Finetti measure of the particles in every $t \geq 0$.

Since further the $C_{\mathscr{P}(E)}[0, \infty)$ martingale problem for $\mathbb{A}$ is well posed (cf. Ethier and Kurtz, 1993), the result follows from Lemma 4.9.1 of Ethier and Kurtz (1986). □

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

7 Haploid case

The Fleming-Viot process with haploid fertility selection has got generator given by

$$
\begin{aligned}
&\frac{1}{2}\sum_{1\le i\ne j\le m}\big(\langle \Phi_{ij}f,\mu^{m-1}\rangle - \langle f,\mu^m\rangle\big) + \sum_{j=1}^{m}\langle G_j f,\mu^m\rangle \\
&+\sum_{j=1}^{m}\big(\langle \sigma_j(\cdot)f,\mu^m\rangle - \langle \sigma(\cdot)\otimes f,\mu^{m+1}\rangle\big)
\end{aligned}
\tag{18}
$$

where $\sigma_j(\cdot)$ denotes $\sigma(x_j)$ 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

$$
\Pi(d\mu) = Ce^{2\langle\sigma,\mu\rangle}\Pi_{\theta,\nu_0}(d\mu) \tag{19}
$$

where $\langle\sigma,\mu\rangle = \int \sigma(x)\mu(\mathrm{d}x)$. Note that (19) is a special case of (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 the following Dirichlet process mixture model:

$$
\begin{aligned}
&p_n(y_1=1,\ldots,y_n=1|\,x_1,\ldots,x_n) = \prod_{i=1}^{n}\tilde\sigma_n(x_i) \\
&x_1,\ldots,x_n|\,\mu \overset{i.i.d.}{\sim} \mu \\
&\mu \sim \Pi_{\theta,\nu_0}.
\end{aligned}
\tag{20}
$$

From this we obtain

$$
p(x_1,\ldots,x_n|y_1=1,\ldots,y_n=1) \propto p(x_1,\ldots,x_n)\prod_{i=1}^{n}\tilde\sigma_n(x_i)
$$

from which the $(n-1)$ predictive density for

$$
p(x_1,\ldots,x_n|y_1=1,\ldots,y_n=1) = q_n(x_1,\ldots,x_n)
$$

can be written

$$
q_n(\mathrm{d}x|x_1,\ldots,x_{j-1},x_{j+1},\ldots,x_n) = \frac{\theta_n\,\nu_n(\mathrm{d}x) + \sum_{i\ne j}^{n}\tilde\sigma_n(x_i)\,\delta_{x_i}(\mathrm{d}x)}{\theta_n + \sum_{i\ne j}^{n}\tilde\sigma_n(x_i)} \tag{21}
$$

where

$$
\theta_n = \theta\int \tilde\sigma_n(x)\,\nu_0(\mathrm{d}x) \tag{22}
$$

and

$$\nu_n(\mathrm{d}x) = \frac{\tilde{\sigma}_n(x)\,\nu_0(\mathrm{d}x)}{\int \tilde{\sigma}_n(x)\,\nu_0(\mathrm{d}x)}. \tag{23}$$

Defining a Markov particle process as in Section 5, where now the transition density for the new particle is given by (21), and letting

$$\lambda_{n,j} = \frac{n\left(\theta_n + \sum_{i\neq j}\tilde{\sigma}_n(x_i)\right)}{2}$$

we obtain

$$\sum_{j=1}^{n} \frac{1}{2}\theta_n \int [f\eta_j(\mathbf{x}|y) - f(\mathbf{x})]\nu_n(\mathrm{d}y) + \sum_{1\leq i\neq j\leq n} \frac{1}{2}\tilde{\sigma}_n(x_i)[f\eta_j(\mathbf{x}|x_i) - f(\mathbf{x})].$$

Consider now a particular choice for $\tilde{\sigma}_n$, that is

$$\tilde{\sigma}_n(x) = 1 + \frac{2}{n}\sigma(x)$$

where σ is a bounded nonnegative measurable function on E; this yields

$$\begin{aligned} A^n f(\mathbf{x}) =& \sum_{j=1}^{n} \frac{1}{2}\theta \int [f\eta_j(\mathbf{x}|y) - f(\mathbf{x})]\{1 + 2\sigma(y)/n\}\nu_0(\mathrm{d}y) \\ &+ \frac{1}{2}\sum_{1\leq i\neq j\leq n} [f\eta_j(\mathbf{x}|x_i) - f(\mathbf{x})] + \sum_{1\leq i\neq j\leq n} n^{-1}\sigma(x_i)[f\eta_j(\mathbf{x}|x_i) - f(\mathbf{x})]. \end{aligned}$$

Proceeding as in Section 5 we can derive the generator for the process of the empirical measures in the n-dimensional case

$$\begin{aligned} \mathbb{A}^n\phi(\mu) =& \sum_{j=1}^{n}\langle G_j^n f, \mu^{(n)}\rangle + \frac{1}{2}\sum_{1\leq i\neq j\leq n}\langle \Phi_{ij}f - f, \mu^{(n)}\rangle \\ &+ \sum_{1\leq i\neq j\leq n} n^{-1}\sigma(x_i)\langle \Phi_{ij}f - f, \mu^{(n)}\rangle \end{aligned} \tag{24}$$

where

$$G^n f(x) = \frac{1}{2}\theta \int [f(z) - f(x)]\{1 + 2\sigma(z)/n\}\nu_0(\mathrm{d}z). \tag{25}$$

When $f \in B(S^m)$, $m < n$

$$\begin{aligned} \mathbb{A}^n\phi(\mu) =& \sum_{j=1}^{m}\langle G_j^n f, \mu^{(m)}\rangle + \frac{1}{2}\sum_{1\leq i\neq j\leq m}\langle \Phi_{ij}f - f, \mu^{(m)}\rangle \\ &+ \frac{1}{n}\sum_{1\leq i\neq j\leq m}\left(\langle \sigma_i(\cdot)\Phi_{ij}f, \mu^{(m)}\rangle - \langle \sigma_i(\cdot)f, \mu^{(m)}\rangle\right) \\ &+ \frac{n-m}{n}\sum_{j=1}^{m}\left(\langle \sigma_j(\cdot)f, \mu^{(m)}\rangle - \langle \sigma(\cdot)\otimes f, \mu^{(m+1)}\rangle\right). \end{aligned}$$

The limiting operator hence is

$$\mathbb{A}\phi(\mu) = \sum_{j=1}^{m} \langle G_j f, \mu^m \rangle + \frac{1}{2} \sum_{1 \le i \ne j \le m} \langle \Phi_{ij} f - f, \mu^m \rangle \qquad (26)$$
$$+ \sum_{j=1}^{m} \left(\langle \sigma_j(\cdot) f, \mu^m \rangle - \langle \sigma(\cdot) \otimes f, \mu^{m+1} \rangle \right).$$

where G^n has been replaced by G, defined in (1).

Following an analogous procedure to that used in the proof of Theorem 6.1, one can show that the stationary distribution of the Fleming-Viot process with generator (26) is (19), as we know from Ethier and Kurtz (1994) and Ethier and Shiga (2000).

8 Discussion

This paper provides an explicit construction of the Fleming-Viot process with fertility selection in a Bayesian nonparametric framework. The construction shows how to tackle a specific class of population genetics problems with instruments that are widely used in Bayesian settings, like urn schemes, the Gibbs sampler and the Dirichlet process mixture model, hence pointing out useful connections between the two fields. We believe that the used techniques simplify the investigation of Fleming-Viot processes, especially for what regards the elicitation of the stationary distribution, and shed new light on the connections between the empirical measure of the particle and the de Finetti measure underlying the model, thus clarifying the role of the individuals with respect to the measure-valued process from a Bayesian viewpoint.

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.

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.

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

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.

Lo, A.Y. (1984). On a class of Bayesian nonparametric estimates I. Density estimates. *Ann. Statist.* **12**, 351–357.

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