Applied Probability Trust (6 June 2008)

THE NEUTRAL POPULATION MODEL AND BAYESIAN NON-PARAMETRICS

STEFANO FAVARO,* *Bocconi University*

MATTEO RUGGIERO,** *University of Pavia*

DARIO SPANÒ,*** *University of Warwick*

STEPHEN G. WALKER,**** *University of Kent*

Abstract

Fleming-Viot processes are a wide class of probability-measure-valued diffusions which often arise as large population limits of so-called particle processes. Here we invert the procedure and show that a countable population process can be derived directly from the neutral diffusion model, with no arbitrary assumptions. We study the atomic structure of the neutral diffusion model, and elicit a finite dimensional particle process from the time-dependent random measure, for any chosen population size. The static properties are consequences of the fact that its stationary distribution is the Dirichlet process, and rely on a new representation for it. The dynamics are derived directly from the transition function of the neutral diffusion model.

Keywords: Neutral diffusion model; particle process; Dirichlet process; Blackwell-MacQueen urn-scheme; transition function.

* Postal address: IMQ, Università Bocconi, via Sarfatti 25, 20136 Milano, Italy. E-mail: stefano.favaro@phd.unibocconi.it

** Postal address: Department of Economics and Quantitative Methods, University of Pavia, Via San Felice 7, 27100 Pavia, Italy. E-mail: matteo.ruggiero@unipv.it

*** Postal address: Department of Statistics, University of Warwick, CV4 7AL Coventry, UK. E-mail: d.spano@warwick.ac.uk

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

2000 Mathematics Subject Classification: Primary 60G57; 60J35
Secondary 60J60; 92D25

1. Introduction

Bayesian nonparametric statistics and population genetics have a common interest in providing suitable countable representations for the law of random probability distributions. The most studied class of random probability measures in Bayesian nonparametrics is the Dirichlet process, whose characterization and properties were presented by Ferguson (1973) and Ferguson (1974) and further investigated by Blackwell (1973) and Blackwell and MacQueen (1973). In order to define the Dirichlet process, let $\mathcal{X}$ be a Polish space, endowed with its Borel σ-algebra $\mathscr{B}(\mathcal{X})$, and denote with $\mathscr{P}(\mathcal{X})$ the space of Borel probability measures on $\mathcal{X}$, endowed with the topology of weak convergence. Let $(\Omega, \mathscr{F}, \mathbb{P})$ be a probability space and let ν be a finite measure on $\mathcal{X}$. A Dirichlet process on $\mathcal{X}$ with parameter ν, denoted by $\mu \sim \Pi(\cdot|\nu)$, is a random probability measure μ such that, for any finite measurable partition $(A_1, \ldots, A_h)$ of $\mathcal{X}$ with $\nu(A_j) > 0$, for $j = 1, \ldots, h$, $(\mu(A_1;\omega), \ldots, \mu(A_h;\omega))$ is a random vector absolutely continuous with respect to the Lebesgue measure on $\mathbb{R}^{h-1}$ with $\sum_{i=1}^{h} \mu(A_i, \omega) = 1$ and with Dirichlet distribution with parameter $(\nu(A_1), \ldots, \nu(A_h))$, $h \geq 2$. As shown by Blackwell and MacQueen (1973), the Dirichlet process with parameter ν can be alternatively defined as the random probability measure μ induced by the so-called Blackwell-MacQueen Pólya-urn sequence, characterized by the following sampling scheme

$$Pr(X_{k+1} \in \cdot\, |X_1, \ldots, X_k) = \frac{\theta\nu_0(\cdot) + \sum_{j=1}^{k} \delta_{X_j}(\cdot)}{\theta + k} \qquad k \geq 1 \qquad (1)$$

where $\nu_0 = \theta^{-1}\nu$ and $\nu(\mathcal{X}) = \theta$. That is a sequence of observations drawn according to (1) is equivalent to a sequence of i.i.d. observations from μ, where $\mu \sim \Pi(\,\cdot\,|\nu)$. See Blackwell and MacQueen (1973). The sample sequence generated by (1) is exchangeable and therefore, by de Finetti representation theorem

$$L_{\mu,k}(A_1 \times \cdots \times A_k) = \prod_{j=1}^{k} \mu(A_j)$$

for every collection $A_1, \ldots, A_k$ of sets of $\mathscr{B}(\mathcal{X})$, where

$$\mu \overset{d}{=} \lim_{k\to\infty} \eta(X_{(k)}) \qquad a.s. \qquad (2)$$

with $\eta_k = \frac{1}{k}\sum_{j=1}^{k}\delta_{X_j}$. In this case, $(\eta_n)_{n\geq 1}$ is a random sequence with coordinates in $\mathscr{P}(\mathcal{X})$.

In Population Genetics, the Ferguson-Dirichlet process arises as the stationary distribution of a measure-valued diffusion process which describes the evolution of the allele frequencies of a population of genes under the hypothesis of neutral, non-recurrent, parent independent mutation (Ethier and Kurtz, 1994). Such a process, known as *neutral diffusion model*, has continuous sample paths which are functions from $[0,\infty)$ to $\mathscr{P}(\mathcal{X})$, and is characterized in terms of the infinitesimal generator

$$\mathbb{A}\varphi(\mu) = \sum_{i=1}^{m}\langle B_i f, \mu^m\rangle + \frac{1}{2}\sum_{1\leq k\neq i\leq m}\big(\langle \Phi_{ki} f, \mu^{m-1}\rangle - \langle f, \mu^m\rangle\big) \tag{3}$$

where the domain $\mathscr{D}(\mathbb{A})$ is taken to be the set of all bounded functions on $\mathscr{P}(\mathcal{X})$ of the form $\varphi(\mu) = \langle f, \mu^m\rangle$, for f a bounded measurable function on $\mathcal{X}^m$, $\langle f,\mu\rangle$ denoting $\int f\mathrm{d}\mu$ and μ^m being an m-fold product measure. Here B_i is the mutation operator

$$Bf(x) = \frac{1}{2}\theta\int\big[f(z) - f(x)\big]\nu_0(\mathrm{d}z) \tag{4}$$

applied to the i-th argument of f, where $\theta \in \mathbb{R}^+$ and $\nu_0 \in \mathscr{P}(\mathcal{X})$ is a non atomic probability measure. Also, $\Phi_{ki} f(x_1,\ldots,x_m) = f(x_1,\ldots,x_{i-1},x_k,x_{i+1},\ldots,x_m)$.

The transition density of the neutral diffusion model is provided by Ethier and Griffiths (1993) in terms of a mixture of Dirichlet processes, showing an interesting connection with the Bayesian framework. This is given by

$$P(t,\mu,\mathrm{d}\nu) = \sum_{m=0}^{\infty} d_m(t)\int_{\mathcal{X}^n}\Pi\Big(\mathrm{d}\nu\Big|\theta\nu_0 + \sum_{i=1}^{m}\delta_{x_i}\Big)\mu^m(\mathrm{d}x_1,\ldots,\mathrm{d}x_m) \tag{5}$$

where μ^m denotes the m-fold product measure $\mu\times\cdots\times\mu$ and $\Pi(\,\cdot\,|\theta\nu_0 + \sum_{i=1}^{m}\delta_{x_i})$ denotes a posterior Dirichlet process, conditional on the observations $(X_1 = x_1,\ldots,X_m = x_m)$ each sampled from μ. That is, the prior process $\Pi(\,\cdot\,|\theta\nu_0)$ is updated after observing $(X_1,\ldots,X_m)$ by means of Bayes' theorem, yielding $\Pi(\,\cdot\,|\theta\nu_0 + \sum_{i=1}^{m}\delta_{x_i})$ (see Ferguson, 1973). In (5), $d_m(t) = Pr(D(t) = m)$, where $\{D(t), t\geq 0\}$ is a death process with rate

$$\lambda_m = \frac{1}{2}m(\theta + m - 1) \tag{6}$$

and such that $D(0) = \infty$ almost surely. Tavaré (1984), for example, computed that for $m\in\mathbb{N}$

$$d_m(t) = \sum_{n=m}^{\infty}(-1)^{n-m}\binom{n}{m}(\theta+m)_{(n-1)}n!^{-1}\gamma_{n,t,\theta} \tag{7}$$

where

$$\gamma_{n,t,\theta} = (\theta + 2n - 1)e^{-\lambda_n t}$$

and

$$d_0(t) = 1 - \sum_{n=1}^{\infty}(-1)^{n-1}(\theta)_{(n-1)}n!^{-1}\gamma_{n,t,\theta}.$$

Here, $a_{(n)} = a(a+1)\dots(a+n-1)$ for $n \in \mathbb{N}$, with $a_{(0)} = 1$. We will also use $a_{[n]} = a(a-1)\dots(a-n+1)$ for $n \in \mathbb{N}$, with $a_{[0]} = 1$.

Further connections with the Bayesian nonparametric framework are established in some recent results. Walker, Hatjispyros and Nicoleris (2007) provide a construction of the neutral diffusion model via its transition function using ideas on Gibbs sampler based Markov processes. Ruggiero and Walker (2008) propose a construction of the Fleming-Viot process with selection based on a generalised Blackwell-MacQueen Pólya urn scheme, obtained from a Bayesian hierarchical mixture model (see Lo, 1984).

Fleming-Viot processes, introduced by Fleming and Viot (1979) and which include the neutral diffusion model, are generally viewed as limit approximations of the behavior of finite populations of say k alleles, as k goes to infinity. The model of reproduction of the k-alleles population is often represented by a k-dimensional particle processe $\{(X_1(t), \dots, X_k(t)), t \geq 0\}$ with sample paths on the space $D_{\mathcal{X}^k}[0,\infty)$ of càdlàg functions from $[0,\infty)$ to $\mathcal{X}^k$. In this case $\{(X_1(t), \dots, X_k(t)), t \geq 0\}$ is a countable representation of the Fleming-Viot process $\{\mu(t), t \geq 0\}$ in the sense that the process of allele frequencies $\{\eta_k(t), t \geq 0\}$, where at every $t \geq 0$

$$\eta_k(t) = \frac{1}{k}\sum_{j=1}^{k}\delta_{X_j(t)}$$

converges in distribution (in the Skorohod topology) to $\{\mu_t, t \geq 0\}$ as k grows to infinity. A general theory for a countable representation of Fleming-Viot processes is provided by Donnelly and Kurtz (1996) and Donnelly and Kurtz (1999).

Here we consider the opposite problem. Given a measure-valued diffusion, and in particular given the neutral diffusion model, we investigate how a particle process should be in order to be a suitable representation for a finite-population extract from the limiting diffusion $\{\eta_\infty(t), t \geq 0\}$. The main point is of course what *suitable* means. A reasonable criterion seems be that the defining properties of the particle process be derived only from the intrinsic features of the neutral diffusion model itself, with no further arbitrary assumptions. In our case, the static properties of the particle process will be consequences of the fact that the stationary distribution of the neutral diffusion model is the Dirichlet process. The dynamic properties will be derived directly from the transition function (5) and its implications. Here the focus is on the fact that the Dirichlet process provides random probability measures which are purely atomic. This suggests that instead of adopting the usual approach by proposing a population process and show that this converges in distribution to the measure-valued diffusion, we can invert the procedure and derive the population process directly from the diffusion. That is, we investigate the properties of the atoms which give the time-dependent random measure, and show that for any chosen population size $k \geq 1$ we can elicit k atoms from the random measure; then their properties automatically define a particle process, each atom being a particle, with sample paths in $D_{\mathcal{X}^k}[0, \infty)$. When k grows to infinity the infinite population process, summarized by its empirical distribution, is equivalent to the neutral diffusion model. We can thus talk of a population process underlying the neutral diffusion model, in the sense that all properties of the former are derived by the latter. Such constructive approach brings new evidence, once again, of the key role played by Blackwell-MacQueen urn schemes in explaining the fundamental structure of Ferguson-Dirichlet populations.

The paper is organized as follows. Section 2 states the result in the finite case, which is the most relevant here, that is for an arbitrary population size, which determines the size of the finite dimensional particle process. The proof is derived via several lemmas and propositions. In particular, Lemma 1 below provides a representation for the Dirichlet process which is used to elicit the particles from the random measure, and Propositions 1, 2 and 3, with the aid of some technical results, show the dynamics of the particle process. Section 3 provides some discussion and deals with the infinite population case.

2. The particle process

Before stating the main result, we give the following lemma, which beside having a key role in the construction, provides some intuition into the problem. The lemma provides a representation of a random probability measure which is from a Dirichlet process. As recalled in the introduction (see (1)), the Dirichlet process has been characterised via the Blackwell-MacQueen Pólya urn scheme, the relation between the sequence of draws and the random measure being (2). Denote with $\mathscr{P}_k(\theta,\nu_0)$ the joint distribution of a sequence $(Y_1,\dots,Y_k)$ drawn from (1). It is easy to check that for every $k \geq n$, the marginal distribution of n variables of the vector $(Y_1,\dots,Y_k)$ is $\mathscr{P}_n(\theta,\nu_0)$. The lemma elicits k atoms from the random measure, and these atoms have joint distribution $\mathscr{P}_k(\theta,\nu_0)$.

Lemma 1. *For arbitrary $k \geq 1$, let $(Y_1,\dots,Y_k) \sim \mathscr{P}_k(\theta,\nu_0)$, and let $w_1,\dots,w_k$ be k independent random varariables distributed according to a Beta distribution function with parameters $(1,\theta+k-i)$. Let $\mu \sim \Pi(\,\cdot\,|\theta\nu_0)$ independent of the w_i's and define the random element*

$$\mu_k = \sum_{i=1}^{k} p_i \delta_{Y_i} + \Big(1 - \sum_{i=1}^{k} p_i\Big)\mu \tag{8}$$

where $p_1 = w_1$ and $p_i = w_i \prod_{j=1}^{i-1}(1-w_j)$ for $i = 2,\dots,k$. Then $\mu_k \sim \Pi(\,\cdot\,|\theta\nu_0)$.

Proof. From the definition of Dirichlet process on $\mathcal{X}$ it follows that it suffices to prove the result for an h-dimensional vector $(\mu(A_1),\dots,\mu(A_h))$, for any finite measurable partition $A_1,\dots,A_h$ of $\mathcal{X}$ and any $h \geq 1$.

For all $k \geq 1$ we have that $1 - \sum_{i=1}^{k} p_i = \prod_{i=1}^{k}(1-w_i)$. Using the constructive definition of the random variables $p_1,\dots,p_k$ we have

$$\begin{aligned}
&\sum_{i=1}^{k} p_i \Big(\delta_{Y_i}(A_1),\dots,\delta_{Y_i}(A_h)\Big) + \Big(1 - \sum_{i=1}^{k} p_i\Big)(\mu(A_1),\dots,\mu(A_h)) = \\
&= \sum_{i=1}^{k-1} p_i \Big(\delta_{Y_i}(A_1),\dots,\delta_{Y_i}(A_h)\Big) + \Big(1 - \sum_{i=1}^{k-1} p_i\Big) \times \\
&\qquad \times \Big[w_k\Big(\delta_{Y_k}(A_1),\dots,\delta_{Y_k}(A_h)\Big) + (1-w_k)(\mu(A_1),\dots,\mu(A_h))\Big].
\end{aligned}$$

Then, it follows by induction that conditionally on $(Y_1,\dots,Y_k)$,

$$\sum_{i=1}^{k} p_i \Big(\delta_{Y_i}(A_1),\dots,\delta_{Y_i}(A_h)\Big) + \Big(1 - \sum_{i=1}^{k} p_i\Big)(\mu(A_1),\dots,\mu(A_h)) \tag{9}$$

is a random variable distributed according to a Dirichlet distribution function with parameters $(\theta\nu_0(A_1) + \sum_{j=1}^{k}\delta_{Y_j}(A_1),\dots,\theta\nu_0(A_h) + \sum_{j=1}^{k}\delta_{Y_j}(A_h))$. The result follows integrating out $(Y_1,\dots,Y_k)$. □

The sample path of the neutral diffusion model $\{\mu(t), t \geq 0\}$ at stationarity is such that at each time point the state of the process is a random probability measure distributed according to a Dirichlet process. From Lemma 1 it follows that a representation alternative to (1)-(2) of a Dirichlet process is given by (8), which can thus be used, once indexed by time, to describe every instant state of the neutral diffusion model. Given the almost sure discreteness of the Dirichlet process (see Introduction) the connection between two states of the process at different time points, say without loss of generality 0 and $t > 0$, can be expressed according to how many atoms $\mu_k(0)$ and $\mu_k(t)$ share, for arbitrary $k \geq 1$, where for any $t \geq 0$

$$\mu_k(t) = \sum_{i=1}^{k} p_i(t)\delta_{Y_i(t)} + \Big(1 - \sum_{i=1}^{k} p_i(t)\Big)\mu(t)$$

Thus, the change in time of $Y_1, \ldots, Y_k$ in (8) provides an approximation of the change undergone by μ_k. The vector $Y_1, \ldots, Y_k$, whose joint distribution is $\mathscr{P}_k(\theta, \nu_0)$, is then a natural candidate for a finite-dimensional particle process whose components in every instant are from the population μ_k. Since the dynamics of the particle process reflect to a certain extent those of the measure-valued process, $Y_1(0), \ldots, Y_k(0)$ will remain fixed at their state during the interval $[0, t)$ so long as $Y_1, \ldots, Y_k$ remain atoms of $\mu_k(s)$ for $0 \leq s < t$. When one of the atoms drops out, the state of this $\mathcal{X}^k$-valued random process changes, so it is componentwise piecewise constant with jumps. We are then interested in the distribution of interarrival times between jumps, that is the holding times between any atom change. We will show that the atoms change one at a time, and the holding times are exponential with parameter λ_k given in (6). Once again we remark that these results on the dynamic properties of the particle process will rely only on the transition function (5) of the neutral diffusion model, with no further assumptions.

The next theorem, which is the main result of the paper, formalizes the above heuristics. It will be proved by means of several lemmas in the remainder of the section.

Theorem 1. *For any arbitrary $k \geq 1$, let $(\mu_k(t), t \geq 0)$ be the neutral diffusion model with infinitesimal generator (3). Then, $((Y_1(t), \ldots, Y_k(t)), t \geq 0)$ is a k-dimensional particle process with sample paths in $D_{\mathcal{X}^k}[0, \infty)$ and jumps at exponential times of* *parameter λ_k, given by (6), such that at each jump at most one coordinate at a time is updated according to (1).*

First we state two results that will be needed later. The first can be found in Walker, Hatjispyros and Nicoleris (2007) (cf. Result [A] and [B], pag. 72-73).

Lemma 2. *Let $d_m(t)$ be (7). Then*

$$\sum_{m=k}^{\infty} \frac{m_{[k]}}{(\theta+m)_{(k)}} d_m(t) = e^{-\lambda_k t} \tag{10}$$

and

$$\sum_{m=k-1}^{\infty} \frac{m_{[k-1]}}{(\theta+m)_{(k)}} d_m(t) = \frac{e^{-\lambda_{k-1}t} - e^{-\lambda_k t}}{2(\lambda_k - \lambda_{k-1})}. \tag{11}$$

The following lemma provides a useful result that will be used later.

Lemma 3. *Let $\theta > 0$ and $m, n \in \mathbb{N}$, with $n \leq m$. Then*

$$\sum_{n=1}^{m} \frac{\Gamma(\theta+m-n)}{\Gamma(1+m-n)} = \frac{\Gamma(\theta+m)}{\theta\Gamma(m)}.$$

Proof.

$$\begin{aligned}
\sum_{n=1}^{m} \frac{\Gamma(\theta+m-n)}{\Gamma(1+m-n)} &= \frac{\Gamma(\theta+m)}{\Gamma(1+m)} \sum_{n=1}^{m} \frac{m_{[n]}}{(\theta+m-1)_{[n]}} \\
&= \frac{\Gamma(\theta+m)}{\Gamma(1+m)(\theta+m-1)_{[m]}} \left[\sum_{n=1}^{m-1} (\theta+m-1-n)_{[m-n]} m_{[n]} + m_{[m]} \right] \\
&= \frac{\Gamma(\theta+m)}{\Gamma(1+m)(\theta+m-1)_{[m]}} m(\theta+m-1)_{[m-1]} \\
&= \frac{\Gamma(\theta+m)}{\theta\Gamma(m)}.
\end{aligned}$$

□

We have now all the ingredients to show that the interarrival times between successive jumps, that is single atom updates, are exponential with parameter λ_k. This will be proved by means of the following three propositions.

Let $\{\mu(t), t \geq 0\}$ be a neutral diffusion model, so that the transitions of $\mu(t)$ are described by (5). The form of the transition function yields that conditionally on the starting state $\mu(0)$, the arrival state $\mathrm{d}\mu(t)$ after a time interval t is obtained as follows. An m-sized sample $(X_1, \dots, X_m)$ is drawn from $\mu(0)$, where the sample size m is governed by a death process D_t starting from infinity, so that the probability of sampling m variables from $\mu(0)$ for an interval of lag t is $d_m(t)$ (see (7)). Then $\mathrm{d}\mu(t)$ is sampled from a posterior Dirichlet process, conditionally on the vector $(X_1, \dots, X_m)$ (see Introduction). Hence the m-sized vector sampled from the starting state $\mu(0)$ carries m atoms of information about $\mu(0)$, which are taken into account when sampling $\mu(t)$.

We exploit these intrinsic features of the transition function (5) for computing the probability that respectively none, one or two atoms of $\mu_k(0)$ among those in $(Y_1, \dots, Y_k)$ drop in the interval $\mathrm{d}t$. These three cases will be examined separately in Proposition 1, 2 and 3 below.

Proposition 1. *Let $\{\mu(t), t \geq 0\}$ be a neutral diffusion model with transition function (5), and suppose the time interval $[0, s]$ is of infinitesimal length. Then the probability of $(Y_1(0), \dots, Y_k(0))$ being atoms of $\mu(s)$ is $e^{-\lambda_k s}$, where λ_k is (6).*

Proof. Call $n_1, \dots, n_k$ the multiplicity of $Y_1(0), \dots, Y_k(0)$ respectively in an m-sized sample from $\mu(0)$, where $\mu(0)$ is given by (8). A necessary condition for $Y_1(0), \dots, Y_k(0)$ to be in the m-sized sample from $\mu(0)$, and hence possibly be atoms of $\mu(s)$, is that m be not smaller than k, and that $\sum_{i=1}^k n_i \leq m$. Hence we have to integrate: over the random weights $p_1, \dots, p_k$ associated to the atoms $Y_1, \dots, Y_k$, whose distribution is derived by the stick-breaking procedure, also known as residual allocation model, in Lemma 1; over all possible combinations of multiplicities of atom draws in a sample of size m, so that $n_1 \in \{1, \dots, m\}$, $n_2 \in \{1, \dots, m - n_1\}$, and so on up to $n_k \in \{1, \dots, m - \sum_{i=1}^k n_i\}$, so that $\sum_{i=1}^k n_i \leq m$; and over the sample size for $m \geq k$. Hence we have that the probability of $(Y_1(0), \dots, Y_k(0))$ being atoms in $\mu(s)$ is

$$
\begin{aligned}
&P(\{Y_1(0), \dots, Y_k(0)\} \in Y_\infty(s)) = \\
&= \sum_{m=k}^{\infty} d_m(s) \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_k=1}^{m-n_1-\dots-n_{k-1}} \binom{m}{n_1, n_2, \dots, n_k} \\
&\quad \times \int_0^1 \cdots \int_0^1 \prod_{i=1}^{k} \left(w_i^{n_i} \prod_{j=1}^{i-1} (1-w_j)^{n_i} \right) (1-w_i)^{m - \sum_{h=1}^k n_h}
\end{aligned}
$$

$$\times \prod_{\ell=1}^{k} (\theta + k - \ell)(1 - w_\ell)^{\theta+k-\ell-1} \mathrm{d}w_1 \dots \mathrm{d}w_k$$

which simplifies to

$$\sum_{m=k}^{\infty} \theta_{(k)} d_m(s) \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_k=1}^{m-n_1-\ldots-n_{k-1}} \binom{m}{n_1, n_2, \ldots, n_k}$$
$$\times \int_0^1 \cdots \int_0^1 \prod_{i=1}^{k} w_i^{n_i} (1 - w_i)^{\theta+m-\sum_{h=1}^{i} n_h + k - i - 1} \mathrm{d}w_1 \dots \mathrm{d}w_k. \quad (12)$$

By solving the integrals as incomplete Beta densities, the previous equals

$$\sum_{m=k}^{\infty} \theta_{(k)} d_m(s) \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_k=1}^{m-n_1-\ldots-n_{k-1}} \binom{m}{n_1, n_2, \ldots, n_k}$$
$$\times \prod_{i=1}^{k} \frac{\Gamma(n_i + 1)\Gamma(\theta + m - \sum_{h=1}^{i} n_h + k - i)}{\Gamma(\theta + m - \sum_{h=1}^{i-1} n_h + k - i)}$$

and simplifying the product with the multinomial coefficient gives

$$\sum_{m=k}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_k=1}^{m-n_1-\ldots-n_{k-1}} \frac{\Gamma(\theta + m - \sum_{h=1}^{k} n_h)}{\Gamma(m - \sum_{h=1}^{k} n_h + 1)}.$$

Applying Lemma 3 yields

$$\sum_{m=k}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \sum_{n_1=1}^{m} \cdots \sum_{n_{k-1}=1}^{m-n_1-\ldots-n_{k-2}} \frac{\Gamma(\theta + m - \sum_{h=1}^{k-1} n_h)}{\theta\, \Gamma(m - \sum_{h=1}^{k-1} n_h)}.$$

Take now $\theta' = \theta + 1$ and $m' = m - 1$, so that the last ratio in the previous is

$$\frac{\Gamma(\theta' + m' - \sum_{h=1}^{k-1} n_h)}{\theta \Gamma(m' - \sum_{h=1}^{k-1} n_h + 1)}$$

and apply again Lemma 3 to get

$$\sum_{m=k}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \sum_{n_1=1}^{m} \cdots \sum_{n_{k-2}=1}^{m-n_1-\ldots-n_{k-3}} \frac{\Gamma(\theta' + m' - \sum_{h=1}^{k-2} n_h)}{\theta\, \theta'\, \Gamma(m' - \sum_{h=1}^{k-2} n_h)}$$
$$= \sum_{m=k}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \sum_{n_1=1}^{m} \cdots \sum_{n_{k-2}=1}^{m-n_1-\ldots-n_{k-3}} \frac{\Gamma(\theta + m - \sum_{h=1}^{k-2} n_h)}{\theta(1+\theta)\Gamma(m - \sum_{h=1}^{k-2} n_h - 1)}.$$

Repeat the procedure other $k-2$ times, taking $\theta'' = \theta' + 1$, $m'' = m' + 1$ and so on, yielding

$$\sum_{m=k}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \sum_{n_1=1}^{m} \cdots \sum_{n_{k-3}=1}^{m-n_1-\ldots-n_{k-4}} \frac{\Gamma(\theta + m - \sum_{h=1}^{k-3} n_h)}{\theta(1+\theta)(2+\theta)\Gamma(m - \sum_{h=1}^{k-3} n_h - 1)}$$

$$\vdots$$

$$= \sum_{m=k}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \frac{\Gamma(\theta+m)}{\theta(\theta+1)\dots(\theta+k-1)\Gamma(m-k+1)}$$

$$= \sum_{m=k}^{\infty} \frac{m_{[k]}}{(\theta+m)_{(k)}} d_m(s)$$

which by means of (10) gives the result. □

The following proposition gives the probability that one atom update occurs in an infinitesimal lag.

Proposition 2. *Let $\{\mu(t), t \geq 0\}$ be a neutral diffusion model with transition function (5), and suppose the time interval $[0, s]$ is of infinitesimal length. The probability that exactly $k-1$ particles of the vector $(Y_1(0), \dots, Y_k(0))$ are atoms in $\mu(s)$ is $\lambda_k s + o(s)$.*

Proof. Consider the setting of the proof of Proposition 1. If the atom that changes is Y_j, $1 \leq j \leq k$, in order to compute the probability of the statement it suffices to set $n_j = 0$ in (12), so that there are no values of $Y_j(0)$ in the m-sized sample from $\mu(0)$ (hence no piece of information about $\mu(0)$ corresponding to the atom $Y_j(0)$ pass to $\mu(s)$). Hence the probability that one atom drops out is

$$\sum_{j=1}^{k} P(\{Y_1(0), \dots, Y_{j-1}(0), Y_{j+1}(0), \dots, Y_k(0)\} \in Y_\infty(s), Y_j \notin Y_\infty(s)) =$$

$$= \sum_{j=1}^{k} \sum_{m=k-1}^{\infty} \theta_{(k)} d_m(s)$$

$$\times \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_{j-1}=1}^{m-\sum_{\ell=1}^{j-2} n_\ell} \sum_{n_{j+1}=1}^{m-\sum_{\ell=1}^{j-1} n_\ell} \cdots \sum_{n_k=1}^{m-\sum_{\ell\neq j}^{k-1} n_\ell} \binom{m}{n_1, n_2, \dots, n_{j-1}, n_{j+1}, \dots, n_k}$$

$$\times \int_0^1 \dots \int_0^1 \prod_{i\neq j}^{k} w_i^{n_i} (1-w_i)^{\theta+m-\sum_{h\neq j}^{i} n_h+k-i-1} (1-w_j)^{\theta+m-\sum_{h=1}^{j-1} n_h+k-j-1} \mathrm{d}w_1 \dots \mathrm{d}w_k$$

Proceeding as in Proposition 1, and simplifying with the multinomial coefficient the Gamma functions resulting from the integrals, yields

$$\sum_{j=1}^{k} \sum_{m=k-1}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s)$$

$$\times \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_{j-1}=1}^{m-\sum_{\ell=1}^{j-2} n_\ell} \sum_{n_{j+1}=1}^{m-\sum_{\ell=1}^{j-1} n_\ell} \cdots \sum_{n_k=1}^{m-\sum_{\ell\neq j}^{k-1} n_\ell} \frac{\Gamma(\theta+m-\sum_{h\neq j}^{k} n_h)}{\Gamma(m-\sum_{h\neq j}^{k} n_h+1)}$$

Applying $k-1$ times Lemma 3 we obtain

$$\begin{aligned}
&\sum_{j=1}^{k} \sum_{m=k-1}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \\
&\quad \times \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_{j-1}=1}^{m-\sum_{\ell=1}^{j-2} n_\ell} \sum_{n_{j+1}=1}^{m-\sum_{\ell=1}^{j-1} n_\ell} \cdots \sum_{n_{k-1}=1}^{m-\sum_{\ell\neq j}^{k-2} n_\ell} \frac{\Gamma(\theta+m-\sum_{h\neq j}^{k-1} n_h)}{\theta\Gamma(m-\sum_{h\neq j}^{k-1} n_h)} \\
&= \sum_{j=1}^{k} \sum_{m=k-1}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \\
&\quad \times \sum_{n_1=1}^{m} \sum_{n_2=1}^{m-n_1} \cdots \sum_{n_{j-1}=1}^{m-\sum_{\ell=1}^{j-2} n_\ell} \sum_{n_{j+1}=1}^{m-\sum_{\ell=1}^{j-1} n_\ell} \cdots \sum_{n_{k-2}=1}^{m-\sum_{\ell\neq j}^{k-3} n_\ell} \frac{\Gamma(\theta+m-\sum_{h\neq j}^{k-2} n_h)}{\theta(1+\theta)\Gamma(m-\sum_{h\neq j}^{k-2} n_h-1)} \\
&\vdots \\
&= \sum_{j=1}^{k} \sum_{m=k-1}^{\infty} \theta_{(k)} \frac{\Gamma(m+1)}{\Gamma(\theta+m+k)} d_m(s) \frac{\Gamma(\theta+m)}{\theta(\theta+1)\ldots(\theta+k-2)\Gamma(m-k+2)} \\
&= k(\theta+k-1) \sum_{m=k-1}^{\infty} \frac{m_{[k-1]}}{(\theta+m)_{(k)}} d_m(s),
\end{aligned}$$

from which, using (11) and the definition of λ_k, we get

$$\lambda_k \frac{e^{-\lambda_{k-1}s} - e^{-\lambda_k s}}{\lambda_k - \lambda_{k-1}} = \lambda_k s + o(s)$$

which gives the result. □

Before stating the third proposition, we need one last technical result.

Lemma 4. *Let $d_m(s)$ be (7). Then*

$$\sum_{m=k-2}^{\infty} \frac{m_{[k-2]}}{(\theta+m)_{(k)}} d_m(s) = \\ = \frac{(\lambda_{k-1}-\lambda_{k-2})e^{-\lambda_k s} - (\lambda_k-\lambda_{k-2})e^{-\lambda_{k-1}s} + (\lambda_k-\lambda_{k-1})e^{-\lambda_{k-2}s}}{4(\lambda_k-\lambda_{k-1})(\lambda_k-\lambda_{k-2})(\lambda_{k-1}-\lambda_{k-2})}$$

Proof. Denote

$$G(t) = \sum_{m=k-2}^{\infty} \frac{m_{[k-2]}}{(\theta+m)_{(k)}} d_m(s);$$

from a result in Ethier and Griffiths (1993), pag. 1585, it follows that

$$\frac{\mathrm{d}G(s)}{\mathrm{d}s} + \lambda_k G(s) = \frac{1}{2} \sum_{m=k-2}^{\infty} \frac{m_{[k-2]}}{(\theta+m)_{(k-1)}} d_m(s)$$

and we know from Walker, Hatjispyros and Nicoleris (2007) (cf. Result B pag. 73) that

$$\sum_{m=k-2}^{\infty} \frac{m_{[k-2]}}{(\theta+m)_{(k-1)}} d_m(s) = \frac{e^{-\lambda_{k-2}s} - e^{-\lambda_{k-1}s}}{2(\lambda_{k-1}-\lambda_{k-2})}.$$

The general solution of the differential equation is

$$G(s) = \frac{e^{-\lambda_{k-2}s}}{4(\lambda_k - \lambda_{k-2})(\lambda_{k-1}-\lambda_{k-2})} - \frac{e^{-\lambda_{k-1}s}}{4(\lambda_k - \lambda_{k-1})(\lambda_{k-1}-\lambda_{k-2})} + Ce^{-\lambda_k s}$$

and using the initial condition $G(0) = 0$ we obtain

$$C = \frac{1}{4(\lambda_k - \lambda_{k-1})(\lambda_k - \lambda_{k-2})}$$

from which the result follows. □

The last proposition states that the probability of two atom updates occurring in an infinitesimal time lag is negligible.

Proposition 3. *Let $\{\mu(t), t \geq 0\}$ be a neutral diffusion model with transition function (5), and suppose the time interval $[0,s]$ is of infinitesimal length. The probability that only $k-2$ particles of the vector $(Y_1(0),\ldots,Y_k(0))$ are atoms in $\mu(s)$ is $o(s)$.*

Proof. The event of two particles changing in $[0,s]$ means that $Y_j(0), Y_h(0)$, for $1 \leq j \neq h \leq k$, are not selected in the m-sized sample from $\mu(0)$ and thus do not compare as atoms in $\mu(s)$. Similarly to Proposition 2, we set $n_j = n_h = 0$, and integrate out the indices, obtaining

$$\begin{aligned}
&\sum_{1\leq j\neq h\leq k} P(\{Y_i(0),\ i\neq j,h\} \in Y_\infty(s), \{Y_j(0), Y_h(0)\} \notin Y_\infty(s)) \\
&= \sum_{1\leq j\neq h\leq k}^{k} \sum_{m=k-1}^{\infty} \theta_{(k)} d_m(s) \sum_{(*)} \binom{m}{n_1, n_2, \ldots, n_{j-1}, n_{j+1}, \ldots, n_{h-1}, n_{h+1}, \ldots, n_k} \\
&\times \prod_{i\neq j,h}^{k} \int_0^1 w_i^{n_i} (1-w_i)^{\theta+m-\sum_{\ell\neq j,h}^{i} n_\ell + k-i-1} (1-w_j)^{\theta+m-\sum_{\ell\neq j,h}^{j-1} n_\ell + k-j-1} \mathrm{d}w_i \\
&\times \int_0^1 (1-w_j)^{\theta+m-\sum_{\ell\neq h}^{j-1} n_\ell + k-j-1} \mathrm{d}w_j \int_0^1 (1-w_h)^{\theta+m-\sum_{\ell\neq j}^{h-1} n_\ell + k-h-1} \mathrm{d}w_h
\end{aligned}$$

where $(*)$ denotes the set of frequencies n_i, for $1 \leq i \leq k$ and $i \neq h, k$, such that each n_i runs from 1 to $m - \sum_{\ell \neq j,h}^{i-1} n_\ell$. Proceeding as in Proposition 1 we obtain

$$k(k-1)(\theta+k-1)(\theta+k-2) \sum_{m=k-2}^{\infty} \frac{m_{[k-2]}}{(\theta+m)_{(k)}} d_m(s).$$

By Lemma 4 the previous equals, up to a multiplicative constant,

$$\begin{aligned} &\lambda_k(e^{-\lambda_{k-2}s} - e^{-\lambda_{k-1}s}) + \lambda_{k-1}(e^{-\lambda_k s} - e^{-\lambda_{k-2}s}) + \lambda_{k-2}(e^{-\lambda_{k-1}s} - e^{-\lambda_k s}) \\ &= \Big[\lambda_k(\lambda_{k-1} - \lambda_{k-2}) + \lambda_{k-1}(\lambda_{k-2} - \lambda_k) + \lambda_{k-2}(\lambda_k - \lambda_{k-1})\Big] s + o(s) = o(s) \end{aligned}$$

which gives the result. □

Propositions 1, 2 and 3 imply that the interarrival times of the particle process are governed by a Poisson process with parameter λ_k, and that one particle at a time drops out of the k-dimensional time-dependent vector. Say that Y_i is such particle. Then, from Lemma 1 and the exchangeability of a sequence drawn according to (1), it follows that the incoming particle is a sample from

$$\frac{\theta}{\theta+k-1}\,\nu_0 + \frac{1}{\theta+k-1} \sum_{j \neq i} \delta_{Y_j}. \tag{13}$$

This is due to the fact that conditionally on $\mu(t)$, the removed particle will be replaced by another variable in the infinite sequence from the Blackwell-MacQueen urn that characterizes $\mu(t)$. Integrating out $\mu(t)$, the incoming variable will still be from the Blackwell-MacQueen urn, but conditionally on the other $k-1$ particles, and its law will be the predictive distribution (13). This completes the proof of Theorem 1.

3. Discussion and infinite population limit

We have constructed a particle process which is directly derived by the properties of neutral diffusion model. The key of the derivation is the representation of a Dirichlet process as μ_k in (8), as proved in Lemma 1. Then, given k atoms $(Y_1(0), \ldots, Y_k(0))$ of the starting state $\mu(0)$ of the neutral diffusion model, we can describe a particle process as follows. The state of the particle process remains constant until the first time t such that one of the particles is no longer an atom of $\mu(t)$. The computation of the probabilities that all k particles are still atoms of $\mu(t)$ and that one of the k particles is no longer an atom of $\mu(t)$ yields the distribution of the interarrival time of the particle process until the following renewal. When one of the particles is no longer an atom of the random measure, not having been sampled from the starting state, it is substituted with another atom of $\mu(0)$ which differs from the other $k-1$, and hence is another observation from the Blackwell-MacQueen urn.

When the population size of the particle process grows to infinity, in Lemma 1 we have that the sum of weights $\sum_{i=1}^{k} p_i$ tends to one, and the second term in

$$\mu_k = \sum_{i=1}^{k} p_i \delta_{Y_i} + \Big(1 - \sum_{i=1}^{k} p_i\Big)\mu$$

vanishes. Then $\mu_\infty = \text{weak-lim}_{k\to\infty} \mu_k$ is still a Dirichlet process, but unlike for finite k, the particle process now fully characterises every instant state of the neutral diffusion model, as we have an infinite sequence of observations from $\mu(t)$, conditionally on $\mu(t)$, which provides full information on the distribution. From this setting it is now trivial to derive all usual infinite population results for the neutral diffusion model, like the weak convergence in the Skorohod space of the process of empirical measures of the particles.

Acknowledgements

This work was partially carried out while M. Ruggiero was affiliated to the Department of Statistics and Applied Mathematics, University of Turin, Italy, and supported by CRT Foundation, Turin, and by MIUR, grant 2006/133449. D. Spanò is partly supported by EPSRC-CRiSM.

References

Blackwell, D. (1973). Discreteness of Fersuson selections. *Ann. Statist.* **1**, 356–358.

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

Donnelly, P. and Kurtz, T.G. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.* **24**, 698–742.

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 Griffiths, R.C. (1993). The transition function of a Fleming-Viot process. *Ann. Probab.* **21**, 1571–1590.

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.

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

Ferguson, T.S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629.

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

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

Ruggiero, M. and Walker, S.G. (2008). Bayesian nonparametric construction of the Fleming-Viot process with fertility selection. To appear in *Statist. Sinica*.

Tavaré, S. (1984). Line-of-descent and genealogical processes, and their applications in population genetic models. *Theoret. Population Biol.* **26**, 119–164.

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