Electronic Journal of Statistics
Vol. 3 (2009) 1556–1566
ISSN: 1935-7524
DOI: 10.1214/09-EJS563

On a Gibbs sampler based random process in Bayesian nonparametrics

Stefano Favaro*

Università degli Studi di Torino
Corso Unione Sovietica 218/bis, 10134, Torino, Italy
e-mail: stefano.favaro@unito.it

Matteo Ruggiero

Università degli Studi di Pavia
Via San Felice 5, 27100, Pavia, Italy
e-mail: matteo.ruggiero@unipv.it

Stephen G. Walker

University of Kent
CT2 7NZ, Canterbury, UK
e-mail: S.G.Walker@kent.ac.uk

Abstract: We define and investigate a new class of measure-valued Markov chains by resorting to ideas formulated in Bayesian nonparametrics related to the Dirichlet process and the Gibbs sampler. Dependent random probability measures in this class are shown to be stationary and ergodic with respect to the law of a Dirichlet process and to converge in distribution to the neutral diffusion model.

Keywords and phrases: Random probability measure, Dirichlet process, Blackwell-MacQueen Pólya urn scheme, Gibbs sampler, Bayesian nonparametrics.

Received December 2009.

Contents

1. Introduction

Measure valued Markov processes arise naturally in modeling the composition of evolving populations and play an important role in a variety of research areas such as population genetics, bioinformatics, machine learning, Bayesian

*Also affiliated to Collegio Carlo Alberto, Moncalieri, Italy.

nonparametrics, combinatorics and statistical physics. In particular, in Bayesian nonparametrics, there has been interest in measure-valued random processes since the seminal paper by Feigin and Tweedie [10], where the law of the Dirichlet process has been characterized as the unique invariant distribution of a certain class of measure-valued Markov chains.

The Dirichlet process is a random probability measure whose introductory properties and characterizations were extensively presented by Ferguson [11] and Ferguson [12] and subsequently investigated by Blackwell [3] and Blackwell and MacQueen [4]. Consider a complete and separable metric space E endowed with the Borel σ–field $\mathscr{E}$ and let $\mathcal{P}(E)$ be the space of probability measures on E with the σ–field $\mathscr{P}$ generated by the topology of the weak convergence. Let also α be a nonnegative finite measure on $(E, \mathscr{E})$ with total mass $a > 0$. A random probability measure P on E is a Dirichlet process with parameter α, henceforth denoted $P \sim \mathscr{D}(\cdot|\alpha)$, if for any $k \geq 2$ and any finite measurable partition $B_1, \ldots, B_k$ of E such that $\alpha(B_j) > 0$ for $j = 1, \ldots, k$, the random vector $(P(B_1), \ldots, P(B_k))$ has Dirichlet distribution with parameter $(\alpha(B_1), \ldots, \alpha(B_k))$. Various characterizations of the Dirichlet process have been proposed in the literature. In particular, a well-known result obtained by Sethuraman [19] characterizes the law of the Dirichlet process as the unique solution of a certain distributional equation on $\mathcal{P}(E)$. Let α be as above, Y an E-valued random variable with distribution $P_0 := \alpha/a$ and θ a random variable independent of Y with Beta distribution with parameter $(1, a)$; Theorem 2.4 in Sethuraman [19] shows that a Dirichlet process P on E with parameter α uniquely satisfies the distributional equation

$$P \stackrel{\mathrm{d}}{=} \theta \delta_Y + (1 - \theta) P, \tag{1.1}$$

where all the random elements on the right-hand side of (1.1) are independent. Equation (1.1) has been widely used in the Bayesian nonparametrics literature in order to provide properties and characterizations of the Dirichlet process and its linear functionals. See the comprehensive review by Lijoi and Prünster [16] and references therein.

A first interesting example of the applicability of (1.1) in the research area of functionals of the Dirichlet process is given by Feigin and Tweedie [10], where (1.1) is recognised as the distributional equation for the unique invariant measure of a measure-valued Markov chain $\{P_m, m \geq 0\}$ defined via the recursive identity

$$P_m = \theta_m \delta_{Y_m} + (1 - \theta_m) P_{m-1} \qquad m \geq 1 \tag{1.2}$$

where $P_0 \in \mathcal{P}(E)$ is arbitrary and $\{Y_m, m \geq 1\}$ and $\{\theta_m, m \geq 1\}$ are sequences of E-valued random variables independent and identically distributed respectively as Y and θ above and independent of each other. By investigating the functional Markov chain $\{G_m, m \geq 0\}$, with $G_m := \int_E g(x) P_m(dx)$ for any $m \geq 0$ and for any measurable linear function $g : E \mapsto \mathbb{R}$, Feigin and Tweedie [10] provide properties of the corresponding linear functional G of a Dirichlet process P on E with parameter α, i.e. $G := \int_E g(x) P(dx)$. Further developments of the linear functional Markov chain $\{G_m, m \geq 0\}$ are provided by Guglielmi and Tweedie [14], Jarner and Tweedie [15] and more recently by Erhardsson [6].

Recently, a generalization of (1.1) has been proposed by Favaro and Walker [9]. Let α be as above, with total mass a, and let $\{Y_j, j \geq 1\}$ be a Blackwell-MacQueen Pólya sequence with parameter α (see Blackwell and MacQueen [4]), i.e. $\{Y_j, j \geq 1\}$ is an exchangeable sequence with de Finetti measure given by the law of a Dirichlet process with parameter α. For a fixed integer $n \geq 1$, consider also a random vector $(q_1, \ldots, q_n)$ with Dirichlet$(1, \ldots, 1)$ distribution, where $\sum_{i=1}^n q_i = 1$, and a random variable θ with Beta(n, a) distribution, such that $\{Y_i, i \geq 1\}$, $(q_1, \ldots, q_n)$ and θ are mutually independent. Then Favaro and Walker [9] show that a Dirichlet process P on E with parameter α uniquely satisfies the distributional equation

$$P \stackrel{\mathrm{d}}{=} \theta \sum_{i=1}^{n} q_i \delta_{Y_i} + (1-\theta) P \tag{1.3}$$

where all the random elements on the right-hand side of (1.3) are independent. It can be easily checked that equation (1.3) generalizes (1.1), which can be recovered by setting $n = 1$. In the present paper, by combining the original idea of Feigin and Tweedie [10] and the distributional equation (1.3), we define and investigate a new class of measure valued Markov chains having the Dirichlet process as its unique invariant measure. In particular, (1.3) is recognised as the distributional equation for the unique invariant measure of a certain class of Gibbs sampler based measure-valued Markov chains whose transition functions are driven by the predictive distributions of the Blackwell-MacQueen Pólya urn scheme. An interesting application of this new class of measure valued Markov chains arises in relation to a well-known stochastic model in population genetics, the so-called neutral diffusion model. See Ethier and Kurtz [8]. This application, together with the constructive definition of the Gibbs sampler based Markov chain, provides a further connection between Bayesian nonparametrics and population genetics along the same research lines recently investigated by Walker et al [20] and Ruggiero and Walker [18].

Following these guidelines, in Section 2 we define and investigate the new class of Gibbs sampler based measure-valued chains, and state the convergence in distribution to the neutral diffusion model. A brief review of the essential features of the neutral diffusion model can be found in the Appendix.

2. A Gibbs sampler based measure valued Markov chain

We first state a lemma, whose proof can be found in Wilks [21], Section 7, which will be useful for the next result, and is reported here for ease of reference.

Lemma 2.1. *Let U and V be independent random vectors with Dirichlet distributions with parameters $(\alpha_1, \ldots, \alpha_n)$ and $(\beta_1, \ldots, \beta_n)$ respectively. Let W, independent of U and V, have Beta distribution with parameters $(\sum_{i=1}^n \alpha_i, \sum_{i=1}^n \beta_i)$. Then*

$$WU + (1-W)V \sim \mathit{Dirichlet}(\alpha_1 + \beta_1, \ldots, \alpha_n + \beta_n).$$

Consider now a sample $(X_1, \dots, X_n)$ of size $n \geq 1$ from a Dirichlet process P with parameter θP_0, for $\theta > 0$ and $P_0 \in \mathcal{P}(E)$. Then it is well known that the Dirichlet process enjoys the conjugacy property (see Ferguson [11]), meaning that conditional on the data $(X_1, \dots, X_n)$, P is still a Dirichlet process with updated parameter, namely

$$P|\, X_1, \dots, X_n \sim \mathscr{D}\left(\cdot|\theta P_0 + \sum\nolimits_{i=1}^{n} \delta_{X_i}\right). \tag{2.1}$$

Theorem 2.2 below provides a representation for the posterior process (2.1) in terms of a stochastic equation. To this end, recall that a sample of size n from a Dirichlet process P has marginal distribution given by the Blackwell-MacQueen Pólya urn scheme, that is

$$\mathcal{M}_n = P_0 \prod_{j=1}^{n-1} \frac{\theta P_0 + \sum_{1 \leq i \leq j} \delta_{X_i}}{\theta + j}, \tag{2.2}$$

with θ and P_0 as above.

Theorem 2.2. *Let α_n, H_n, $X^{(n)} = (X_1, \dots, X_n)$ and $W^{(n)} = (W_{1,n}, \dots, W_{n,n})$ be mutually independent, where*

$$\begin{aligned} \alpha_n &\sim Beta(n, \theta) \\ H_n &\sim \mathscr{D}(\cdot|\alpha), \qquad \alpha = \theta P_0 \\ X^{(n)} &\sim \mathcal{M}_n \\ W^{(n)} &\sim Dirichlet(1, \dots, 1) \end{aligned}$$

with $\mathcal{M}_n$ as in (2.2). Define

$$Q_n = \alpha_n Z_n + (1 - \alpha_n) H_n \tag{2.3}$$

where

$$Z_n = \sum_{i=1}^{n} W_{i,n} \delta_{X_i}. \tag{2.4}$$

Then Q_n given $(X_1, \dots, X_n)$ is a Dirichlet process with parameter $\alpha + \sum_{i=1}^{n} \delta_{X_i}$.

Proof. Denote by $(x_1, \dots, x_n)$ the observed values of $(X_1, \dots, X_n)$. Then it is sufficient to prove that for any $k \in \mathbb{N}$ and any measurable partition $A_1, \dots, A_k$ of E, conditionally on $(x_1, \dots, x_n)$ we have $(Z_n(A_1), \dots, Z_n(A_k))$ has Dirichlet distribution with parameters $(\sum_{i=1}^{n} \delta_{x_i}(A_1), \dots, \sum_{i=1}^{n} \delta_{x_i}(A_k))$. Then the result follows from Lemma 2.1. Let $V_1, V_2, \dots, V_n$ be independent and such that $V_i \sim \mathrm{Beta}(1, n-i)$ for $i = 1, \dots, n$ (so that $V_n = 1$ a.s.). If we define $W_{1,n} = V_1$ and $W_{i,n} = V_i \prod_{j=1}^{i-1}(1 - V_j)$ for $i = 2, \dots, n$, then it can be easily checked that $W_{1,n}, W_{2,n}, \dots, W_{n,n}$ are identically distributed with $W_{i,n} \sim \mathrm{Beta}(1, n-1)$ for $i = 1, \dots, n$ and $(W_{1,n}, \dots, W_{n,n}) \sim \mathrm{Dirichlet}(1, \dots, 1)$. Observe that by construction we have $1 - \sum_{i=1}^{n-1} W_{i,n} = \prod_{j=1}^{n-1}(1 - V_j)$ and $\sum_{i=1}^{n} W_{i,n} = 1$. It follows from Lemma 2.1 that, conditionally on x_n we have

$$V_n(\delta_{x_n}(A_1), \dots, \delta_{x_n}(A_k)) \sim \text{Dirichlet}(\delta_{x_n}(A_1), \dots, \delta_{x_n}(A_k)).$$

By induction we can write

$$\sum_{i=1}^{n} W_{i,n}(\delta_{x_i}(A_1), \dots, \delta_{x_i}(A_k)) = \sum_{i=1}^{n-1} W_{i,n}(\delta_{x_i}(A_1), \dots, \delta_{x_i}(A_k)) + \left(1 - \sum_{i=1}^{n-1} W_{i,n}\right)[V_n(\delta_{x_n}(A_1), \dots, \delta_{x_n}(A_k))].$$

and by repeated application of Lemma 2.1 it can be easily checked that, conditionally on $(x_1, \dots, x_n)$

$$(Z_n(A_1), \dots, Z_n(A_k)) \sim \text{Dirichlet}\left(\sum_{i=1}^{n} \delta_{x_i}(A_1), \dots, \sum_{i=1}^{n} \delta_{x_i}(A_k)\right).$$

Hence, by Lemma 2.1 it follows that, conditionally on $(x_1, \dots, x_n)$, the vector $(Q_n(A_1), \dots, Q_n(A_k))$ has distribution

$$\text{Dirichlet}\left(\theta P_0(A_1) + \sum_{i=1}^{n} \delta_{x_i}(A_1), \dots, \theta P_0(A_k) + \sum_{i=1}^{n} \delta_{x_i}(A_k)\right)$$

giving the result. □

We now provide a dynamic version of (2.3), denoted $Q_n(\cdot) = \{Q_n(k), k \geq 1\}$, obtained by means of Gibbs sampling techniques, and investigate the properties of the resulting random element. The dynamics on (2.3) are induced via (2.4) as follows. Update iteratively a component X_i, say, of $(X_1, \dots, X_n)$, according to its full conditional distribution

$$\mathcal{M}_n(X_i \in \cdot \mid X_1, \dots, X_{i-1}, X_{i+1}, \dots, X_n) = \frac{\theta P_0(\cdot) + \sum_{k \leq n, k \neq i} \delta_{x_k}(\cdot)}{\theta + n - 1} \tag{2.5}$$

where X_i is selected at random with uniform distribution. It is clear that we are performing a *random scan* Gibbs sampler on $(X_1, \dots, X_n)$. The following result identifies the limit in distribution of $Q_n(\cdot)$, if appropriately rescaled and for n tending to infinity, to be the celebrated neutral diffusion model. This is a random element taking values in the space of continuous functions from $\mathbb{R}_+$ to the space $\mathcal{P}(E)$ of probability measures. See the Appendix for a brief review.

Theorem 2.3. *Assume E is compact. Let $S(\cdot)$ be a neutral diffusion model with initial distribution $\nu \in \mathcal{P}(\mathcal{P}(E))$, let $Q_n(\cdot)$ be as above, and let $\tilde{Q}_n(\cdot)$ be the subsequence obtained by retaining only one every n^2 iterations of $Q_n(\cdot)$. If $\tilde{Q}_n(\cdot)$ has initial distribution $\nu_n \in \mathcal{P}(\mathcal{P}(E))$ and $\nu_n \Rightarrow \nu$, then $\tilde{Q}_n(\cdot) \Rightarrow S(\cdot)$ in $C_{\mathcal{P}(E)}[0, \infty]$.*

Proof. Note first that considering a subsequence of $Q_n(\cdot)$ is equivalent to embedding $Q_n(\cdot)$ in continuous time and rescaling appropriately the process by means of the distribution of the waiting times between discontinuities. Let the waiting times be negative exponential with parameter λ_n and denote the resulting process $S_n(\cdot) = \{S_n(t), t \geq 0\}$. The infinitesimal operator for the process $X^{(n)}(\cdot) = \{(X_1^{(n)}(t), \dots, X_n^{(n)}(t)), t \geq 0\}$ can be written

$$A^n f(x) = \sum_{i=1}^{n} \frac{\lambda_n}{n} \int \big[f(\eta_i(x|y)) - f(x)\big] \frac{\theta P_0(dy) + \sum_{k\neq i}^{n} \delta_{x_k}(dy)}{\theta + n - 1} \tag{2.6}$$

for $f \in C(E)$, with $C(E)$ denoting continuous functions on E, and $\eta_i(x|y) = (x_1, \dots, x_{i-1}, y, x_{i+1}, \dots, x_n)$. Taking $\lambda_n = n(\theta + n - 1)/2$ (more generally, choose $\lambda_n \in O(n^2/2)$, which yields the same result in the limit for $n \to \infty$) we have

$$A^n f(x) = \sum_{i=1}^{n} B_i f(x) + \frac{1}{2} \sum_{1\leq k\neq i\leq n} \big[f(\eta_i(x|x_k)) - f(x)\big]$$

where $B_i f$ is the operator

$$Bg(x) = \frac{1}{2}\theta \int \big[g(y) - g(x)\big] P_0(dy), \qquad g \in C(E)$$

acting on f as a function of its i-th argument. For $m \leq n$ define

$$W_{j_1,\dots,j_m,n}(t) = W_{j_1,n}(t) \frac{W_{j_2,n}(t)}{1 - W_{j_1,n}(t)} \cdots \frac{W_{j_m,n}(t)}{1 - \sum_{l=1}^{m-1} W_{j_l,n}(t)} \tag{2.7}$$

for $1 \leq j_1 \neq \cdots \neq j_m \leq n$, with $(W_{1,1}, \dots, W_{1,n})$ as in Theorem 2.2. Here (2.7) is simply the probability of picking m elements of n without replacement, once the i-th element is assigned a weight W_i. Define also the probability measure on E^m given by

$$S_n^{(m)}(t) = \alpha_n Z_n^{(m)}(t) + (1 - \alpha_n) H_n^m$$

where

$$Z_n^{(m)}(t) = \sum_{1\leq j_1\neq\cdots\neq j_m\leq n} W_{j_1,\dots,j_m,n}(t) \delta_{(x_{j_1}(t),\dots,x_{j_m}(t))}$$

and H_n^m denotes the m-fold product measure $H_n \times \cdots \times H_n$. Finally, let $\phi_m \in C(\mathcal{P}(E))$ be, for $f \in C(E^m)$,

$$\phi_m(S_n(t)) = \langle f, S_n^{(m)}(t)\rangle \tag{2.8}$$

where $\langle f, \mu\rangle = \int f d\mu$. The infinitesimal generator of the process $S_n(\cdot)$ for test functions of type (2.8) and general λ_n can be written

$$\mathbb{A}^n \phi_n(S_n) = \frac{\lambda_n}{n} \sum_{i=1}^{n} \int_{E\times\Delta_n} [\phi_n(\tilde{\eta}_i(S_n|z, W^{(n)})) - \phi_n(S_n)] \Gamma(n) dW^{(n)} \tag{2.9}$$

$$\times \frac{\theta P_0(dz) + \sum_{k\neq i} \delta_{x_k}(dz)}{\theta + n - 1}$$

where λ_n is the Poisson rate driving the waiting times, the Dirichlet$(1,\dots,1)$ distributed vector $W^{(n)} := (W_{1,n},\dots,W_{n,n})$ is defined on the $(n-1)$-dimensional simplex

$$\Delta_n = \left\{ (W_{1,n},\dots,W_{n,n}) :\, W_{i,n} \geq 0,\, i = 1,\dots,n,\, \sum_{i=1}^n W_{i,n} = 1 \right\}$$

and $\tilde{\eta}_i(S_n|z, W^{(n)})$ is S_n with z instead of x_i and $W^{(n)}$ instead of the current weights. Letting $\lambda_n = n(\theta+n-1)/2$ as above yields

$$\begin{aligned}\mathbb{A}^n\phi_n(S_n) =\ & \frac{\theta}{2}\sum_{i=1}^n \int_{E\times\Delta_n} [\phi_n(\tilde{\eta}_i(S_n|z, W^{(n)})) - \phi_n(S_n)]\Gamma(n)dW^{(n)}P_0(dz)\\ & \qquad + \frac{1}{2}\sum_{1\leq k\neq i\leq n}\int_{\Delta_n} [\phi_n(\tilde{\eta}_i(S_n|x_k, W^{(n)})) - \phi_n(S_n)]\Gamma(n)dW^{(n)}\\ =\ & \sum_{i=1}^n [\langle M_i f, QS_n^{(n)}\rangle - \langle f, S_n^{(n)}\rangle] + \frac{1}{2}\sum_{1\leq k\neq i\leq n} [\langle \Phi_{ki} f, QS_n^{(n)}\rangle - \langle f, S_n^{(n)}\rangle]\end{aligned} \tag{2.10}$$

where $Mf(x) = (\theta/2)\int_E f(y)P_0(dy)$ and M_i is M applied to the i-th coordinate of f, and $QS_n^{(n)}$ is defined as

$$\Gamma(n)\int_{\Delta_n} S_n^{(n)} dW^{(n)}$$

with $W^{(n)} = (W_{1,n},\dots,W_{n,n})$ as in (2.9). Note that when $f \in C(E^m)$, $m < n$, for $i = m+1,\dots,n$ we have $M_i f = f$ and $\Phi_{ki} f = f$, and when $i \leq m$ and $k > m$ we have $\langle M_i f, \mu\rangle = \langle f, \mu\rangle$ and $\langle \Phi_{ki} f, \mu\rangle = \langle f, \mu\rangle$ for every $\mu \in \mathcal{P}(E)$. Hence

$$\sum_{i=m+1}^n [\langle M_i f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle] = (n-m)[\langle f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle]$$

and

$$\begin{aligned}\sum_{i=m+1}^n\sum_{k=1}^n [\langle \Phi_{ki} f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle] + \sum_{i=1}^m\sum_{k=m+1}^n [\langle \Phi_{ki} f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle]\\ = (n-m)(n+m)[\langle f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle].\end{aligned}$$

It follows that for $f \in C(E^m)$, $m < n$, (2.10) becomes

$$\begin{aligned}\mathbb{A}^n\phi_m(S_n) = & \sum_{i=1}^m [\langle M_i f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle]\\ & + \frac{1}{2}\sum_{i=1}^m\sum_{k=1}^m [\langle \Phi_{ki} f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle]\\ & + \frac{1}{2}(n+m+2)(n-m)[\langle f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle].\end{aligned} \tag{2.11}$$

Now, we can write

$$\langle f, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle \tag{2.12}$$
$$= \alpha_n \sum_{1\le i_1\neq\ldots\neq i_m\le n} f(x_{i_1},\ldots,x_{i_m})\left[\Gamma(n)\int_{\Delta_m} W_{i_1,\ldots,i_m,n} dW^{(m)} - W_{i_1,\ldots,i_m,n}\right]$$

where $W_{i_1,\ldots,i_m,n}$ is (2.7). It can be checked that the expectation of the term in square brackets is zero, since $W^{(n)} = (W_{i_1}\ldots W_{i_m}) \sim \text{Dirichlet}(1,\ldots,1)$, and its second moment equals the second moment of $W_{i_1,\ldots,i_m,n}$. From the proof of Lemma 2.1 it follows that $W_{i_1,\ldots,i_m,n}$ is a product of m independent Beta random variables such that

$$\xi_1 = W_{i_1} \sim \text{Beta}(1, n-1)$$
$$\xi_2 = \frac{W_{i_2}}{1-W_{i_1}} \sim \text{Beta}(1, n-2)$$
$$\vdots$$
$$\xi_m = \frac{W_{i_m}}{1-\sum_{j=1}^{m-1} W_{i_j}} \sim \text{Beta}(1, n-m).$$

Hence the second moment of $W_{i_1,\ldots,i_m,n}$ is just

$$\mathbb{E}(\xi_1^2\cdots\xi_m^2) = \frac{2^m}{(n+1)_{[m]}n_{[m]}} \tag{2.13}$$

where $n_{[m]} = n(n-1)\cdots(n-m+1)$, so that as $n\to\infty$ the term in square brackets and (2.12) converge to zero in mean square. Note that this also implies that $\langle g, QS_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle \to \langle g, S_n^{(m)}\rangle - \langle f, S_n^{(m)}\rangle$ as $n\to\infty$. Furthermore we have $\mathbb{E}(\xi_1\cdots\xi_m) = 1/n_{[m]}$, which, together with (2.13), implies that, for large n, $Z_n^{(m)}$ behaves like

$$\frac{1}{n_{[m]}} \sum_{1\le i_1\neq\cdots\neq i_m\le n} \delta_{(x_{i_1},\ldots,x_{i_m})}.$$

Finally, note that $E(\alpha_n) = n/(n+\theta)$ and $\text{Var}(\alpha_n) = n\theta/[(n+\theta)^2(n+1+\theta)]$ so that $\alpha_n \to 1$ in mean square for $n\to\infty$. It follows that $S_n^{(m)}$ converges to the product measure μ^m, where $\mu = \lim_n n^{-1}\sum_{i\ge1}\delta_{x_i}$, and that (2.11) converges uniformly on $\mathcal{P}(E)$ to

$$\mathbb{A}\phi_m(\mu) = \sum_{i=1}^m [\langle M_i f, \mu^m\rangle - \langle f, \mu^m\rangle] + \frac{1}{2}\sum_{1\le k\neq i\le m} [\langle \Phi_{ki} f, \mu^m\rangle - \langle f, \mu^m\rangle]$$

which is (A.2). From Theorem 1.6.1 of Ethier and Kurtz [7], the uniform convergence of $\mathbb{A}^n\phi_m(S_n)$ to $\mathbb{A}\phi_m(\mu)$ implies convergence of the corresponding semigroups on $C(\mathcal{P}(E))$. Call ν_n the distribution of $S_n(0)$ and ν that of $S(0)$. It can be easily checked by means of Lemma 2.3.2 and Lemma 2.3.3 in Dawson [5] that $X_n(\cdot)$ is an exchangeable Feller process. It follows that the closure of (A.2) in $C(\mathcal{P}(E))$ generates a Feller semigroup on $C(\mathcal{P}(E))$, which by means of Theorem 4.2.5 in Ethier and Kurtz [7] implies that if $\nu_n \Rightarrow \nu$ then $S_n(\cdot) \Rightarrow S(\cdot)$ in $D_{\mathcal{P}(E)}[0,\infty)$. Since the limiting process has continuous sample paths with probability one, from Billingsley [2], Section 18, it follows that the convergence holds in $C_{\mathcal{P}(E)}[0,\infty)$. □

The following proposition shows that the Dirichlet process is an invariant measure and an equilibrium distribution for $Q_n(\cdot)$.

Proposition 2.4. *The law of a Dirichlet process $\mathscr{D}(\cdot|\theta P_0)$ is an invariant measure for $Q_n(\cdot)$ for any $n \geq 1$. Moreover, the law of $Q_n(\cdot) = \{Q_n(k), k \geq 1\}$ converges, as $k \to \infty$, to the law of a Dirichlet process $\mathscr{D}(\cdot|\theta P_0)$ for every initial distribution.*

Proof. Assume $Q_n(0) = Q_n$, with Q_n as in (2.3). Then $Q_n(0)|x^n(0) \sim \mathscr{D}(\cdot|\theta P_0 + \sum_{i=1}^n \delta_{x_i(0)})$, where $X^n(0)$ has distribution (2.2). At the following transition, say at $t \geq 0$, a randomly chosen x_i is updated with a sample from (2.5) and by means of the exchangeability of the vector $X^{(n)}$, the distribution of $X^{(n)}(t)$ is still (2.2). This follows from the fact that the updating rule is a Gibbs sampler (see Gelfand and Smith [13]). Since the vector of weights $W^{(n)}(t)$ is resampled independently, it follows that $Q_n(t)|x^n(t) \sim \mathscr{D}(\cdot|\theta P_0 + \sum_{i=1}^n \delta_{x_i(t)})$ for $t \geq 0$. The first statement then follows by Corollary 1.1 in Antoniak [1]. Note now that from Theorem 4 in Roberts and Rosenthal [17] it follows that the transition function of $Q_n(\cdot)$ converges to the stationary distribution (2.2) in total variation. The second statement is now implied by Theorem 2.2 together with Corollary 1.1 in Antoniak [1], since when $X^n(\cdot)$ is in steady state $Q_n(t)|x^n(t) \sim \mathscr{D}(\cdot|\theta P_0 + \sum_{i=1}^n \delta_{x_i(t)})$. □

From Proposition 2.4 it follows that $\mathscr{D}(\cdot|\theta P_0)$ is the unique invariant distribution of $Q_n(\cdot)$.

Appendix: Background on the neutral diffusion model

The neutral diffusion model, also known as Fleming-Viot process, is a diffusion taking values in the space $\mathcal{P}(E)$ of probability measures on $(E, \mathscr{E})$. A review can be found in Ethier and Kurtz [8]. Assume the individuals of an infinite population evolve subject to mutation and resampling (or random genetic drift). Assume also the mutation is parent-independent, i.e. it does not depend on the genotype of the parent, and the mutation process is driven by the operator

$$Bg(y) = \frac{1}{2}\theta \int [g(z) - g(y)] P_0(dz) \tag{A.1}$$

for $g \in C(E)$, $\theta > 0$ and P_0 a non atomic probability measure on $(E, \mathscr{E})$. Since P_0 is non atomic, every mutant has a type which has never been observed with probability one. Then the neutral diffusion model is characterized by the infinitesimal generator

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

where the domain $\mathscr{D}(\mathbb{A})$ can be taken as the algebra generated by functions of the type $\varphi(\mu) = \langle f, \mu^m \rangle$, where $\langle f, \mu \rangle = \int f d\mu$, $f \in C(E^m)$, and μ^m denotes an m-fold product measure. Also, B_i is (A.1) applied to the i-th component of f, and $\Phi_{ki} : B(E^m) \to B(E^{m-1})$ is defined, for every $1 \leq k \neq i \leq m$, as $\Phi_{ki} f(x) = f(\eta_i(x|x_k))$, with $\eta_i(x|y)$ as in (2.6).

References

[1] Antoniak, C.E., 1974. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems. *Ann. Statist.* **2**, 1152–1274. MR0365969

[2] Billingsley, P., 1968. *Convergence of probability measures*. Wiley, New York. MR0233396

[3] Blackwell, D., 1973. Discreteness of Ferguson selections. *Ann. Statist.* **1** 356–358. MR0348905

[4] Blackwell, D., MacQueen, J.B., 1973. Ferguson distributions via Pólya urn schemes. *Ann. Statist.* **1**, 353–355. MR0362614

[5] Dawson, D.A., 1993. *Measure-valued Markov processes*. Ecole d'Eté de Probabilités de Saint Flour XXI. Lecture Notes in Mathematics **1541**. Springer, Berlin. MR1242575

[6] Erhardsson, T., 2008. Nonparametric Bayesian inference for integrals with respect to an unknown finite measure. *Scand. J. Statist.* **35**, 354–368. MR2418747

[7] Ethier, S.N., Kurtz, T.G., 1986. *Markov processes: characterization and convergence*. Wiley, New York. MR0838085

[8] Ethier, S.N., Kurtz, T.G., 1993. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.* **31**, 345–386. MR1205982

[9] Favaro S., Walker S. G., 2008. A generalized constructive definition for Dirichlet processes. *Statist. Probab. Lett.* **78** 2836–2838. MR2465128

[10] Feigin, P. D., Tweedie, R. L., 1989. Linear functionals and Markov chains associated with Dirichlet process. *Math. Proc. Cambridge Philos. Soc.* **105** 579–585. MR0985694

[11] Ferguson, T.S., 1973. A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230. MR0350949

[12] Ferguson, T. S., 1974. Prior distributions on spaces of probability measures. *Ann. Statist.* **2**, 615–629. MR0438568

[13] Gelfand, A.E., Smith, A.F.M., 1990. Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409. MR1141740

[14] Guglielmi, A. Tweedie, R. L., 2001. Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process. *Bernoulli* **7** 573–592. MR1849368

[15] Jarner, S. F., Tweedie, R. L., 2002. Convergence rates and moments of Markov chains associated with the mean of Dirichlet processes. *Stochastic Process. Appl.* **101** 257–271. MR1931269

[16] Lijoi, A. and Prünster, I., 2009. Distributional properties of means of random probability measures. *Statist. Surveys* **3**, 47–95.

[17] Roberts, G.O., Rosenthal, J.S., 2004. General state space Markov chains and MCMC algorithms. *Probab. Surv.*, **1**, 20–71. MR2095565

[18] Ruggiero, M. and Walker, S. G. (2009). Bayesian nonparametric construction of the Fleming-Viot process with fertility selection. *Statist. Sinica*, **19** 707–720. MR2514183

[19] Sethuraman, J., 1994. A constructive definition of Dirichlet priors. *Statist. Sinica.* **4** 639–650. MR1309433

[20] Walker, S. G., Hatjispyros, S. J. and Nicoleris, T. (2007). A Fleming-Viot process and Bayesian nonparametrics. *Ann. Appl. Probab.* **17**, 67–80. MR2292580

[21] Wilks, S., 1962. *Mathematical Statistics*. John Wiley, New York. MR0144404