DOI: 10.1214/ECP.v14-1508

Elect. Comm. in Probab. **14** (2009), 501–517

ELECTRONIC COMMUNICATIONS in PROBABILITY

COUNTABLE REPRESENTATION FOR INFINITE DIMENSIONAL DIFFUSIONS DERIVED FROM THE TWO-PARAMETER POISSON-DIRICHLET PROCESS

MATTEO RUGGIERO
Department of Economics and Quantitative Methods, University of Pavia, Via S. Felice 5, 27100, Pavia, Italy
email: matteo.ruggiero@unipv.it

STEPHEN G. WALKER
Institute of Mathematics, Statistics and Actuarial Science, University of Kent, CT2 7NZ, Canterbury, UK
email: S.G.Walker@kent.ac.uk

Submitted April 20, 2007, *accepted in final form* November 6, 2009

AMS 2000 Subject classification: 60G57, 60J60, 92D25
Keywords: Two-parameter Poisson-Dirichlet process, population process, infinite-dimensional diffusion, stationary distribution, Gibbs sampler.

Abstract

This paper provides a countable representation for a class of infinite-dimensional diffusions which extends the infinitely-many-neutral-alleles model and is related to the two-parameter Poisson-Dirichlet process. By means of Gibbs sampling procedures, we define a reversible Moran-type population process. The associated process of ranked relative frequencies of types is shown to converge in distribution to the two-parameter family of diffusions, which is stationary and ergodic with respect to the two-parameter Poisson-Dirichlet distribution. The construction provides interpretation for the limiting process in terms of individual dynamics.

1 Introduction

The two-parameter Poisson-Dirichlet process, introduced in [19] and further developed in [20] and [22], provides a family of random probability measures which generalises the Dirichlet process, due to [13], and which has found various applications, among which fragmentation and coalescent theory, excursion theory, combinatorics, machine learning and Bayesian statistics. See, among others, [2], [21], [17], [25] and references therein. A definition of the two-parameter Poisson-Dirichlet process can be given as follows. Let α be a finite non null measure on a complete and separable metric space $\mathbb{X}$, endowed with its Borel sigma algebra $\mathscr{B}(\mathbb{X})$. Call $\theta = \alpha(\mathbb{X})$ the total mass of α, and let $\sigma \in [0,1)$. For $k = 1,2,\ldots$, let V_k be independent Beta$(1-\sigma, \theta + k\sigma)$

random variables, and define a sequence of weights $(q_1, q_2, \dots)$ by

$$q_1 = V_1, \qquad q_i = V_i \prod_{k=1}^{i-1}(1 - V_k), \qquad i \geq 2. \tag{1}$$

The sequence $(q_1, q_2, \dots)$ is said to have GEM(θ, σ) distribution, which generalises the one parameter GEM distribution named after Griffiths, Engen and McCloskey. The sequence of descending order statistics $(q_{(1)}, q_{(2)}, \dots)$ is said to have Poisson-Dirichlet distribution with parameters (θ, σ), denoted here by $\Pi_{\theta,\sigma}$. The GEM(θ, σ) distributed sequence $(q_1, q_2, \dots)$ is also obtained as a size-biased permutation of $(q_{(1)}, q_{(2)}, \dots)$. The case $\Pi_{\theta,0}$ is the (one parameter) Poisson-Dirichlet distribution (see [16]), which is the law of the ranked atoms of a Dirichlet process. Let $(Y_1, Y_2, \dots)$ be i.i.d. observations from the normalised measure $\nu_0 = \alpha/\alpha(\mathbb{X})$, which we assume to be diffuse, and denote with δ_x a point mass at x. A random probability measure μ is said to be a two-parameter Poisson-Dirichlet process with parameters (θ, σ), denoted here $\mu \sim \tilde{\Pi}_{\theta,\sigma}$, if

$$\mu(\cdot) \stackrel{d}{=} \sum_{i=1}^{\infty} q_i \delta_{Y_i}(\cdot). \tag{2}$$

The right-hand side of (2) is known as the stick-breaking representation of the two-parameter Poisson-Dirichlet process, the reason being apparent from (1). This extends the constructive definition of the Dirichlet process, corresponding to $\tilde{\Pi}_{\theta,0}$, which is due to [24] and is obtained from (2) by letting $\sigma = 0$ in (1). [20] provides a prediction scheme which generates a sequence of random variables from a two-parameter Poisson-Dirichlet process. Let ν_0 be as above. Let $(X_1, X_2, \dots) \in \mathbb{X}^\infty$ be such that $X_1 \sim \nu_0$ and, for every $n \geq 1$, given $X_1 = x_1, \dots, X_n = x_n$,

$$X_{n+1} | x_1, \dots, x_n \sim \frac{\theta + \sigma K_n}{\theta + n} \nu_0 + \sum_{j=1}^{K_n} \frac{n_j - \sigma}{\theta + n} \delta_{x_j^*} \tag{3}$$

where $0 \leq \sigma < 1$, $\theta > -\sigma$, K_n is the number of distinct values $(x_1^*, \dots, x_{K_n}^*)$ in the vector $(x_1, \dots, x_n)$, and n_j is the cardinality of the cluster associated with x_j^*. Then $X_1, \dots, X_n$ given μ are i.i.d. μ, where $\mu \sim \tilde{\Pi}_{\theta,\sigma}$. Observe that the rule (3) is non degenerate also when

$$\sigma = -\kappa < 0 \quad \text{and} \quad \theta = m\kappa \qquad \text{for some } \kappa > 0 \text{ and } m = 2, 3, \dots . \tag{4}$$

In this case, the number of distinct values or species in the n-sized vector is bounded above by m, and $\Pi_{m\kappa,-\kappa}$ is m-dimensional symmetric Dirichlet. If $n_0 = \min\{n \in \mathbb{N} : K_n = m\}$, then for all $n > n_0$ the new samples are just copies of past observations. When $\sigma = 0$, (3) reduces to the Blackwell-MacQueen Pólya-urn scheme (see [4]; see also [15]), which generates a sequence of random variables from $\tilde{\Pi}_{\theta,0}$. The Blackwell-MacQueen case is also obtained when (4) holds by taking the limit for m going to infinity for fixed $\theta = m\kappa$.

The two-parameter Poisson-Dirichlet distribution has been recently shown to be the stationary measure of a certain class of diffusion processes taking values in the closure $\overline{\nabla}_\infty$ of the infinite dimensional ordered simplex

$$\nabla_\infty = \left\{ z = (z_1, z_2, \dots) \in [0,1]^\infty : z_1 \geq z_2 \geq \cdots \geq 0, \sum_{i=1}^{\infty} z_i = 1 \right\}. \tag{5}$$

See [18]. More specifically, a class of infinite dimensional diffusions with infinitesimal operator

$$L^{\theta,\sigma} = \sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} - \sum_{i=1}^{\infty} (\theta z_i + \sigma) \frac{\partial}{\partial z_i}, \tag{6}$$

on an appropriately defined domain, is obtained as the limit of certain Markov chains, defined on the space of partitions of the natural numbers, based on the two-parameter generalisation of the Ewens sampling formula due to [19]. When $\sigma = 0$, (6) is the infinitesimal operator of the infinitely-many-neutral-alleles model, studied by [9], which is an unlabeled version of the Fleming-Viot measure-valued diffusion without selection nor recombination, but the diffusion with operator $L^{\theta,\sigma}$ seems to fall outside the class of Fleming-Viot processes. See [12] for a review. Fleming-Viot processes also arise naturally as limits in distribution of certain Markov processes, often referred to as countable constructions or particle processes, which retain local information, i.e. relative to single individuals, rather than pooling it into a probability measure. Examples are [6],[7],[8] and [23].
The aim of this paper is to provide interpretation for (6) in terms of a countable construction of particles, which specifies individual dynamics. By means of simple ideas related to the Gibbs sampler (see, e.g., [14]), we construct a fixed-size right-continuous population process, driven by Pitman's prediction scheme (3), which is reversible with respect to the joint law of a sequence sampled from (3). The associated process of ranked relative frequencies of types is shown to converge in distribution, under suitable conditions, to the diffusion with operator (6).
The paper is organised as follows. In Section 2 the Gibbs sampler is briefly introduced. Section 3 defines the particle process, the associated process of relative frequencies of types, and proves weak convergence. In Section 4 we deal with the stationary properties of both the particle and the simplex-valued diffusion.

2 The Gibbs sampler

The Gibbs sampler (see, e.g., [14]), also known as "heat bath" or "Glauber dynamics", is a special case of the Metroplis-Hastings algorithm, which in turn belongs to the class of Markov chain Monte Carlo (MCMC) procedures. These are often applied to solve integration and optimisation problems in large dimensional spaces. Suppose for example that an integral of some function $f : \mathbb{X} \to \mathbb{R}^d$ with respect to some distribution $\pi \in \mathscr{P}(\mathbb{X})$ is to be evaluated, and Monte Carlo integration turns out to be unfeasible. Then MCMC methods provide a way of constructing a stationary Markov chain with π as the invariant measure. One can then run the chain, discard the first, say, N iterations, and regard the successive output from the chain as approximate correlated samples from π. The size of N is determined according to the convergence properties of the chain.
The Gibbs sampler is one of the most widely used MCMC schemes, and has found a wide range of applications in Bayesian computation. The construction of a Gibbs sampler is as follows. Consider a law $\pi = \pi(\mathrm{d}x_1, \ldots, \mathrm{d}x_n)$ defined on $(\mathbb{X}^n, \mathscr{B}(\mathbb{X}^n))$, and assume that the conditional distributions $\pi(\mathrm{d}x_i | x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ are available for every $1 \leq i \leq n$. Then, given an initial set of values $(x_1^0, \ldots, x_n^0)$, the vector is iteratively updated as follows:

$$\begin{aligned}
x_1^1 &\sim \pi(\mathrm{d}x_1 | x_2^0, \ldots, x_n^0) \\
x_2^1 &\sim \pi(\mathrm{d}x_2 | x_1^1, x_3^0, \ldots, x_n^0) \\
&\vdots \\
x_n^1 &\sim \pi(\mathrm{d}x_n | x_1^1, \ldots, x_{n-1}^1) \\
x_1^2 &\sim \pi(\mathrm{d}x_1 | x_2^1, \ldots, x_n^1),
\end{aligned}$$

and so on. Under some regularity conditions, this algorithm produces a Markov chain with equilibrium law $\pi(\mathrm{d}x_1, \ldots, \mathrm{d}x_n)$. The above updating rule is known as a *deterministic scan*. If instead

the components are updated in a random order, called *random scan*, one also gets reversibility with respect to π.

3 Countable representation

For $n \geq 2$, define a Markov chain on $\mathbb{X}^n$ as follows. Given any initial state of the chain, at each transition an index $1 \leq i \leq n$ is chosen uniformly and the component x_i is updated with a sample of size one from the predictive distribution for x_i derived from the Pitman urn scheme, leaving all other components unchanged. From (3), by the exchangeability of the sequence, this predictive is

$$X_i | \mathbf{x}_{(-i)} \sim \frac{\theta + \sigma K_{n-1,i}}{\theta + n - 1} \nu_0 + \frac{1}{\theta + n - 1} \sum_{j=1}^{K_{n-1,i}} (n_j - \sigma) \delta_{x_j^*} \tag{7}$$

where θ and σ are as above, $\mathbf{x}_{(-i)} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n)$ and $K_{n-1,i}$ denotes the number of distinct values in the subvector $\mathbf{x}_{(-i)}$. We are thus constructing a stationary chain on $\mathbb{X}^n$ via a Gibbs sampler performed on $\mathbf{x} = (x_1, \ldots, x_n)$ by means of a uniform random scan. Embed now the chain in continuous time by superimposing it to a Poisson process of intensity $\lambda_n > 0$, dependent on the vector size, which governs the holding times between successive updates. This simple construction yields a continuous-time pure-jump Markov process corresponding to a contraction semigroup $\{T_n^{\theta,\sigma}(t)\}$, on the set $\hat{C}(\mathbb{X}^n)$ of continuous functions on $\mathbb{X}^n$ which vanish at infinity, given by

$$T_n^{\theta,\sigma}(t) f(\mathbf{x}) = \int_{\mathbb{X}^n} f(\mathbf{y}) \tilde{T}(t, \mathbf{x}, \mathrm{d}\mathbf{y}) \tag{8}$$

where $\tilde{T} : [0, \infty) \times \mathbb{X}^n \times \mathscr{B}(\mathbb{X})^n \to [0, 1]$ is a transition function defined in terms of (7). The infinitesimal generator of the process is

$$\begin{aligned} A_n^{\theta,\sigma} f(\mathbf{x}) = &\sum_{i=1}^{n} \frac{\lambda_n(\theta + \sigma K_{n-1,i})}{n(\theta + n - 1)} \int \left[f(\eta_i(\mathbf{x}|y)) - f(\mathbf{x}) \right] \nu_0(\mathrm{d}y) \\ &+ \sum_{i=1}^{n} \sum_{j=1}^{K_{n-1,i}} \frac{\lambda_n(n_j - \sigma)}{n(\theta + n - 1)} \left[f(\eta_i(\mathbf{x}|x_j^*)) - f(\mathbf{x}) \right] \end{aligned} \tag{9}$$

with domain $\mathscr{D}(A_n^{\theta,\sigma}) = \{f : f \in \hat{C}(\mathbb{X}^n)\}$, where η_i is defined as

$$\eta_i(\mathbf{x}|y) = (x_1, \ldots, x_{i-1}, y, x_{i+1}, \ldots, x_n).$$

It can be easily checked that $\{T_n^{\theta,\sigma}(t)\}$ is also positive, conservative, and strongly continuous in the supremum norm, hence (9) is the generator of a Feller process. Let $\mu_n : \mathbb{X}^n \to \mathscr{P}(\mathbb{X})$, given by

$$\mu_n(t) = \frac{1}{n} \sum_{i=1}^{n} \delta_{x_i(t)}, \tag{10}$$

be the empirical measure associated to the vector $(x_1, \ldots, x_n)$ at time $t \geq 0$. Then $\mu_n(\cdot) := \{\mu_n(t), t \geq 0\}$ defines a measure-valued process with sample paths in the space $D_{\mathscr{P}(\mathbb{X})}([0, \infty))$ of right-continuous functions from $[0, \infty)$ to $\mathscr{P}(\mathbb{X})$ with left limits. For every $m \leq n$ let now

$$\mu^{(m)} = \frac{1}{[n]_m} \sum_{1 \leq j_1 \neq \cdots \neq j_m \leq n} \delta_{x_{j_1}, \ldots, x_{j_m}}$$

where $[n]_m = n(n-1)\dots(n-m+1)$, and define Φ_{ki} and Φ_{k^*i}, both from $\hat{C}(\mathbb{X}^n)$ to $\hat{C}(\mathbb{X}^{n-1})$, respectively as

$$\Phi_{ki}f(\mathbf{x}) = f(\eta_i(\mathbf{x}|x_k)), \qquad 1 \le k \ne i \le n$$

and

$$\Phi_{k^*i}f(\mathbf{x}) = f(\eta_i(\mathbf{x}|x_k^*)), \qquad 1 \le i \le n, \quad 1 \le k \le K_{n-1,i}.$$

Also, let the intensity rate of the Poisson process underlying the holding times be

$$\lambda_n = n(\theta + n - 1) \tag{11}$$

which is positive for $\theta > -\sigma$ and $n \ge 2$. This provides the correct rescaling in (9). Alternatively we could take any $\lambda_n = \mathcal{O}(n^2)$ and get the same result in the limit (see also discussion after equation (14) for the rescaling choice). Then, taking $F \in C(\mathscr{P}(\mathbb{X}))$ to be $F(\mu) = \langle f, \mu^{(n)} \rangle$, with $f \in \mathscr{D}(A_n^{\theta,\sigma})$ and $\langle f, \mu \rangle = \int f \, \mathrm{d}\mu$, the generator of the empirical-measure-valued process $\mu_n(\cdot)$ is

$$\begin{aligned}\mathbb{A}_n^{\theta,\sigma}F(\mu) =& (\theta + \sigma K_{n-1,i}) \sum_{i=1}^{n} \langle P_i f - f, \mu^{(n)} \rangle + \sum_{1 \le k \ne i \le n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle \\ & - \sigma \sum_{i=1}^{n} \sum_{j=1}^{K_{n-1,i}} \langle \Phi_{j^*i} f - f, \mu^{(n)} \rangle\end{aligned}$$

where $Pg(x) = \int g(y) \nu_0(\mathrm{d}y)$, for $g \in \hat{C}(\mathbb{X})$, and $P_i f$ denotes P applied to the i-th coordinate of f. This can be written as the sum $\mathbb{A}_n^{\theta,\sigma}F(\mu) = \mathbb{A}_n^{\theta}F(\mu) + \mathbb{A}_n^{\sigma}F(\mu)$ where

$$\mathbb{A}_n^{\theta}F(\mu) = \theta \sum_{i=1}^{n} \langle P_i f - f, \mu^{(n)} \rangle + \sum_{1 \le k \ne i \le n} \langle \Phi_{ki} f - f, \mu^{(n)} \rangle$$

and

$$\begin{aligned}\mathbb{A}_n^{\sigma}F(\mu) =& \sigma K_{n-1,i} \sum_{i=1}^{n} \langle P_i f - f, \mu^{(n)} \rangle - \sigma \sum_{i=1}^{n} \sum_{j=1}^{K_{n-1,i}} \langle \Phi_{j^*i} f - f, \mu^{(n)} \rangle \\ =& \sigma \sum_{i=1}^{n} K_{n-1,i} \langle P_i f - Q_i^{n,i} f, \mu^{(n)} \rangle.\end{aligned}$$

Here, the operator $Q^{n,i}$ is defined, for $g \in \hat{C}(\mathbb{X})$, as

$$Q^{n,i}g(x) = \int g(y) \mu_{n,i}^*(\mathrm{d}y), \qquad \mu_{n,i}^* = \frac{1}{K_{n-1,i}} \sum_{j=1}^{K_{n-1,i}} \delta_{x_j^*}$$

and $Q_j^{n,i} f$ is $Q^{n,i}$ applied to the j-th coordinate of f. Note that when $F(\mu) = \langle f, \mu^{(m)} \rangle$, $m \le n$, $\mathbb{A}_n^{\theta}$ equals

$$\mathbb{A}_n^{\theta}F(\mu) = \theta \sum_{i=1}^{m} \langle P_i f - f, \mu^{(m)} \rangle + \sum_{1 \le k \ne i \le m} \langle \Phi_{ki} f - f, \mu^{(m)} \rangle$$

which, as n tends to infinity, converges to

$$\mathbb{A}^{\theta}F(\mu) = \theta \sum_{i=1}^{m} \langle P_i f - f, \mu^{m} \rangle + \sum_{1 \le k \ne i \le m} \langle \Phi_{ki} f - f, \mu^{m} \rangle.$$

This is twice the generator of the Fleming-Viot process without selection nor recombination and with parent independent mutation with rate $\theta/2$. Note that by taking $\lambda_n' = \lambda_n/2$ instead of (11), yields $\mathbb{A}^\theta/2$. Of course, $\mathbb{A}^\theta$ is also obtained as the infinite population limit of $\mathbb{A}_n^{\theta,\sigma}$ when $\sigma = 0$ (and $F(\mu) = \langle f, \mu^{(m)}\rangle$, $m \le n$). Thus the special case of the Pitman urn scheme with $\sigma = 0$, i.e. the Blackwell-MacQueen urn, provides, via a Gibbs sampler construction, the neutral diffusion model. When $F(\mu)$ is of the form $F(\mu) = g(\langle f_1, \mu\rangle, \ldots, \langle f_m, \mu\rangle)$, for $m \in \mathbb{N}$, $g \in C^2(\mathbb{R}^m)$, and $f_1, \ldots, f_m \in \hat{C}(\mathbb{X})$, we can write $\mathbb{A}_n^{\theta,\sigma}$ as

$$\begin{aligned}\mathbb{A}_n^{\theta,\sigma}F(\mu) = \theta \sum_{i=1}^m [\langle Pf_i, \mu\rangle - \langle f_i, \mu\rangle]\, g_{z_i}(\langle f_1, \mu\rangle, \ldots, \langle f_m, \mu\rangle) \\ + \sum_{1\le k\ne i\le m} [\langle f_i f_j, \mu\rangle - \langle f_i, \mu\rangle\langle f_j, \mu\rangle]\, g_{z_i z_j}(\langle f_1, \mu\rangle, \ldots, \langle f_m, \mu\rangle) \\ + \sigma \sum_{i=1}^m K_{n-1,i}[\langle Pf_i, \mu\rangle - \langle Q^{n,i} f_i, \mu\rangle]\, g_{z_i}(\langle f_1, \mu\rangle, \ldots, \langle f_m, \mu\rangle)\end{aligned}$$

which, again, converges when $\sigma = 0$ to the familiar generator of the neutral diffusion model (cf., e.g., [11]). Now, define the first and second derivatives of $F(\mu)$ as

$$\frac{\partial F(\mu)}{\partial \mu(x)} = \lim_{\varepsilon \downarrow 0} \frac{1}{\varepsilon}[F(\mu + \varepsilon\delta_x) - F(\mu)],$$

$$\frac{\partial^2 F(\mu)}{\partial \mu(x)\partial \mu(y)} = \lim_{\substack{\varepsilon_1 \downarrow 0 \\ \varepsilon_2 \downarrow 0}} \frac{1}{\varepsilon_1 \varepsilon_2}[F(\mu + \varepsilon_1\delta_x + \varepsilon_2\delta_x) - F(\mu)].$$

Then $\mathbb{A}_n^{\theta,\sigma}$ can also be written

$$\begin{aligned}\mathbb{A}_n^{\theta,\sigma}F(\mu) = \int (\theta + \sigma K_{n-1,x}) B\left(\frac{\partial F(\mu)}{\partial \mu(\cdot)}\right)\mu(\mathrm{d}x) \\ + \int\int [\mu(\mathrm{d}x)\delta_x(y) - \mu(\mathrm{d}x)\mu(\mathrm{d}y)]\frac{\partial^2 F(\mu)}{\partial \mu(x)\partial \mu(y)} \\ - \sigma \int K_{n-1,x}\, C^{(n)}\left(\frac{\partial F(\mu)}{\partial \mu(\cdot)}\right)\mu(\mathrm{d}x) + \frac{R_n(F)}{n}\end{aligned} \tag{12}$$

where $K_{n-1,x}$ is analogous to $K_{n-1,i}$ referred to the atom x,

$$Bf(x) = \int_{\mathbb{X}} [f(y) - f(x)]\nu_0(\mathrm{d}y) \tag{13}$$

is the unit rate mutation operator, $C^{(n)}$ is

$$C^{(n)}f(x) = \int_{\mathbb{X}} [f(y) - f(x)]\mu_{n,x}^*(\mathrm{d}y), \tag{14}$$

and $R_n(F)$ is a bounded remainder. The operator $\mathbb{A}_n^{\theta,\sigma}$ does not seem to be well-behaved in the limit, due to the multiplicative term in the $\mathbb{A}_n^\sigma$ part. An inspection of (7), which generates the particles, reveals the heuristics underlying this phenomenon. The probability of sampling a new

species can be split into two terms, $\theta/(\theta+n-1)$ and $\sigma K_{n-1,i}/(\theta+n-1)$. For large n, the two terms are of order n^{-1} and $n^{-1+\sigma}$ respectively, since K_n is of order n^σ (see [20]). With appropriate changes, similar considerations can be made for the empirical part of (7). The point here is that it is seemingly unfeasible to rescale the process with a rate able to retain, in the limit, all terms as well-defined infinite-dimensional genetic mechanisms. For instance, choosing $\lambda_n = \mathcal{O}(n^{2-\sigma})$, yields in the limit a degenerate measure-valued process with constant mutation rate and no resampling. Conversely, letting $\lambda_n = \mathcal{O}(n^\ell)$, with $\ell > 2-\sigma$, in the attempt to preserve the resampling, leads to a process with infinite mutation rate. Note that we have no degrees of freedom on σ, which cannot depend on n by definition of the two-parameter Poisson-Dirichlet process. This makes the characterization of the limit of (10) a difficult task.

A way of overcoming this problem is to restrict the framework. When we have a vector of size $n \geq 1$, let $F(\mu)$ in (12) be given by $F(\mu) = g(\langle\phi_1,\mu\rangle,\ldots,\langle\phi_n,\mu\rangle)$, where $g \in C^2(\mathbb{R}^n)$ and $\phi_j(\cdot) = 1_{x_j^*}(\cdot)$ is the indicator function of x_j^*, $1 \leq j \leq n$. That is $\langle\phi_i,\mu\rangle = \mu(\{x_j^*\}) = z_j$ is the relative frequency of the j-th observed type. Hence we can identify $\mathcal{P}(\mathbb{X})$ with the simplex

$$\Delta_n = \left\{\mathbf{z} = (z_1,\ldots,z_n) \in [0,1]^n : \sum_{i=1}^n z_i = 1\right\}. \tag{15}$$

Note that g has $n - K_n$ null arguments when there are K_n different types in the vector. In this case we regard Δ_{K_n} as a subspace of Δ_n and $g(z_1,\ldots,z_{K_n},0,\ldots,0)$ as $C(\Delta_n)$-valued rather than $C(\Delta_{K_n})$-valued, since K_n is a function of $(x_1,\ldots,x_n)$. Within this more restricted framework, (12) reduces to the operator

$$\begin{aligned}\mathcal{A}_n^{\theta,\sigma} = &\sum_{i=1}^{K_n}(\theta+\sigma\tilde{K}_{n-1,i})\left(\sum_{j=1}^{K_n} b_{ji}^{(K_n)} z_j\right)\frac{\partial}{\partial z_i} + \sum_{i,j=1}^{K_n} z_i(\delta_{ij}-z_j)\frac{\partial^2}{\partial z_i \partial z_j} \\ &- \sigma\sum_{i=1}^{K_n}\tilde{K}_{n-1,i}\left(\sum_{j=1}^{K_n} c_{ji}^{(K_n)} z_j\right)\frac{\partial}{\partial z_i} + \frac{R_n}{n}\end{aligned} \tag{16}$$

with domain

$$\mathcal{D}(\mathcal{A}_n^{\theta,\sigma}) = \{g \in C^2(\Delta_n)\}. \tag{17}$$

Here $\tilde{K}_{n-1,i}$ denotes the number of non null components in the vector $(z_1,\ldots,z_{K_n})$ after z_i is updated to $z_i - n^{-1}$. Furthermore, $(\theta+\sigma\tilde{K}_{n-1,i})b_{ji}^{(K_n)}$ is the intensity of a mutation from type j to type i when there are K_n different types, with $b_{ii}^{(K_n)} = -\sum_{j\neq i} b_{ij}^{(K_n)}$, and $-\sigma\tilde{K}_{n-1,i}c_{ji}^{(K_n)}$ is the analog for the operator (14).

Remark 3.1. Recall from the introduction that the prediction scheme (3) is non degenerate also when $\sigma = -\kappa < 0$ and $\theta = m\kappa$ for some $\kappa > 0$ and $m \geq 2$. It can be easily seen that in this case (16) becomes

$$\begin{aligned}\tilde{\mathcal{A}}_n^{\theta,\sigma} = &\sum_{i=1}^{K_n}(\kappa(m-\tilde{K}_{n-1,i})\left(\sum_{j=1}^{K_n} b_{ji}^{(K_n)} z_j\right)\frac{\partial}{\partial z_i} + \sum_{i,j=1}^{K_n} z_i(\delta_{ij}-z_j)\frac{\partial^2}{\partial z_i \partial z_j} \\ &+ \kappa\sum_{i=1}^{K_n}\tilde{K}_{n-1,i}\left(\sum_{j=1}^{K_n} c_{ji}^{(K_n)} z_j\right)\frac{\partial}{\partial z_i} + \frac{R_n}{n}\end{aligned}$$

which, for n tending to infinity, since $\tilde{K}_{n-1,i}$ eventually reaches m with probability one, converges to

$$\tilde{\mathscr{A}}^{\theta,\sigma} = \theta \sum_{i=1}^{m} \left(\sum_{j=1}^{m} c_{ji}^{m} z_j \right) \frac{\partial}{\partial z_i} + \sum_{i,j=1}^{m} z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j}.$$

This is the neutral-alleles-model with m types, which can be dealt with as in [9]. In particular, for m going to infinity and $\theta = m\kappa$ kept fixed, one obtains, under appropriate conditions, the infinitely-many-neutral-alleles-model, whose stationary distribution is the one parameter Poisson-Dirichlet distribution. This is consistent with the fact that the same limit applied to (3) yields the Blackwell-MacQueen urn scheme. □

When the mutation is governed by (13) we have $b_{ij} = \nu_0(\{j\}) - \delta_{ij}$ (cf., e.g., [11]). Also, from (14) we have $c_{ij} = \mu_i^*(\{j\}) - \delta_{ij} = \tilde{K}_{n-1,i}^{-1} - \delta_{ij}$. When the distribution ν_0 of the allelic type of a mutant is diffuse, these parameters yield

$$\begin{aligned} \mathscr{A}_n^{\theta,\sigma} = &- \theta \sum_{i=1}^{K_n} z_i \frac{\partial}{\partial z_i} + \sum_{i,j=1}^{K_n} z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} \\ &+ \sigma \sum_{i=1}^{K_n} \tilde{K}_{n-1,i} \left\{ \sum_{j=1}^{K_n} \left(-\delta_{ij} - \left(\frac{1}{\tilde{K}_{n-1,i}} - \delta_{ij} \right) \right) z_j \right\} \frac{\partial}{\partial z_i} + \frac{R_n}{n} \end{aligned}$$

which in turn equals

$$\mathscr{A}_n^{\theta,\sigma} = \sum_{i,j=1}^{K_n} z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} - \sum_{i=1}^{K_n} (\theta z_i + \sigma) \frac{\partial}{\partial z_i} + \frac{R_n}{n}. \tag{18}$$

Alternatively we could take the mutation to be symmetric, that is $b_{ji}^{(K_n)} = (K_n - 1)^{-1}$, for $j \neq i$, so that

$$\sum_{1 \le j \le K_n} b_{ji}^{(K_n)} z_j = - \sum_{1 \le j \neq i \le K_n} b_{ij}^{(K_n)} z_j + \sum_{1 \le j \neq i \le K_n} b_{ji}^{(K_n)} z_j = \frac{1 - z_i}{K_n - 1} - z_i$$

This choice yields a different operator $\mathscr{A}_n^{\theta,\sigma}$ but is equivalent in the limit for $n \to \infty$.

Proposition 3.2. *Let $Z^{(n)}(\cdot)$ be a Δ_n-valued process with infinitesimal operator $\mathscr{A}_n^{\theta,\sigma}$ defined by (17) and (18). Then $Z^{(n)}(\cdot)$ is a Feller Markov process with sample paths in $D_{\Delta_n}([0,\infty))$.*

Proof. Denote with $\mathscr{P}_n^{\theta,\sigma}$ the joint distribution of an n-sized vector whose components are sequentially sampled from (3). Let $X^{(n)}(\cdot)$ be the Markov process corresponding to (8). Then $X^{(n)}(\cdot)$ has marginal distributions $\mathscr{P}_n^{\theta,\sigma}$ (see also Corollary 4.2 below). Also, given $\mathbf{x} \in \mathbb{X}^n$, from the exchangeability of the $\mathscr{P}_n^{\theta,\sigma}$-distributed vector it follows that $\tilde{T}(t, \mathbf{x}, B) = \tilde{T}(t, \tilde{\pi}\mathbf{x}, \tilde{\pi}B)$, for every permutation $\tilde{\pi}$ of $\{1, \dots, n\}$ and $B \in \mathscr{B}(\mathbb{X})^n$. By Lemma 2.3.2 of [5], $X^{(n)}(t)$ is an exchangeable Feller process on $\mathbb{X}^n$. The result now follows from Proposition 2.3.3 of [5]. □

The remainder of the section is dedicated to prove the existence of a suitably defined limiting process, which will coincide with that in [18], and the weak convergence of the process of ranked frequencies. In the following section we will then show that the limiting process is stationary and ergodic with respect to the two-parameter Poisson-Dirichlet distribution.

For every $\mathbf{z} \in \Delta_n$ with K_n positive components, define $\rho_n : \Delta_n \to \nabla_\infty$ as

$$\rho_n(\mathbf{z}) = (z_{(1)}, \dots, z_{(K_n)}, 0, 0, \dots) \tag{19}$$

where $z_{(1)} \geq \cdots \geq z_{(K_n)}$ are the the descending order statistics of $\mathbf{z}$ and ∇_∞ is (5). Let also

$$\nabla_n = \left\{ \mathbf{z} = (z_1, z_2, \dots) \in \nabla_\infty : z_{n+1} = 0 \right\}$$

and define the operator

$$\mathscr{B}_n^{\theta,\sigma} = \sum_{i,j=1}^{K_n} z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} - \sum_{i=1}^{K_n} (\theta z_i + \sigma) \frac{\partial}{\partial z_i} + \frac{R_n}{n},$$

with R_n as in (18) and domain

$$\mathscr{D}(\mathscr{B}_n^{\theta,\sigma}) = \{ g \in C(\nabla_n) : \, g \circ \rho_n \in C^2(\Delta_n) \}.$$

Proposition 3.3. *The closure in $C(\nabla_n)$ of $\mathscr{B}_n^{\theta,\sigma}$ generates a strongly continuous, positive, conservative, contraction semigroup $\{\mathscr{T}_n(t)\}$ on $C(\nabla_n)$. Given $\nu_n \in \mathscr{P}(\Delta_n)$, let $Z^{(n)}(\cdot)$ be as in Proposition 3.2, with initial distribution ν_n. Then $\rho_n(Z^{(n)}(\cdot))$ is a strong Markov process corresponding to $\{\mathscr{T}_n(t)\}$ with initial distribution $\nu_n \circ \rho_n^{-1}$ and sample paths in $D_{\nabla_n}([0,\infty))$.*

Proof. Let $\{\mathscr{S}_n(t)\}$ be the Feller semigroup corresponding to $Z^{(n)}(\cdot)$. Then the proof is the same as that of Proposition 2.4 of [9]. In particular, it can be shown that $\{\mathscr{S}_n(t)\}$ maps the set of permutation-invariant continuous functions on Δ_n into itself. This, together with the observation that for every such f there is a unique $g \in C(\nabla_n)$ such that $g = f \circ \rho_n^{-1}$ and $g \circ \rho_n = f$, allows to define a strongly continuous, positive, conservative, contraction semigroup $\{\mathscr{T}_n(t)\}$ on $C(\nabla_n)$ by $\mathscr{T}_n(t)f = [\mathscr{S}_n(t)(f \circ \rho_n)] \circ \rho_n^{-1}$. Then $\rho_n(Z^{(n)}(\cdot))$ inherits the strong Markov property from $Z^{(n)}(\cdot)$, and is such that $\mathbb{E}[f(\rho_n(Z^{(n)}(t+s)))|\rho_n(Z^{(n)}(u)), u \leq s] = \mathscr{T}_n(t)f(\rho_n(Z^{(n)}(s)))$. □

Define now the operator

$$\mathscr{B}^{\theta,\sigma} = \sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} - \sum_{i=1}^{\infty} (\theta z_i + \sigma) \frac{\partial}{\partial z_i} \tag{20}$$

with domain defined as

$$\begin{aligned} \mathscr{D}(\mathscr{B}^{\theta,\sigma}) = \{ &\text{subalgebra of } C(\overline{\nabla}_\infty) \text{ generated by functions } \varphi_m : \overline{\nabla}_\infty \to [0,1], \\ &\text{where } \varphi_1 \equiv 1 \text{ and } \varphi_m(\mathbf{z}) = \textstyle\sum_{i \geq 1} z_i^m, \; m \geq 2 \}. \end{aligned} \tag{21}$$

Here $\overline{\nabla}_\infty$ is the closure of ∇_∞, namely

$$\overline{\nabla}_\infty = \left\{ \mathbf{z} = (z_1, z_2, \dots) \in [0,1]^\infty : z_1 \geq z_2 \geq \cdots \geq 0, \; \sum_{i=1}^{\infty} z_i \leq 1 \right\}$$

which is compact, so that the set $C(\overline{\nabla}_\infty)$ of real-valued continuous functions on $\overline{\nabla}_\infty$ with the supremum norm $\left|\left| f \right|\right| = \sup_{\mathbf{x} \in \overline{\nabla}_\infty} |f(\mathbf{x})|$ is a Banach space. Functions φ_m are assumed to be evaluated in ∇_∞ and extended to $\overline{\nabla}_\infty$ by continuity. We will need the following result, whose proof can be found in the Appendix.

Proposition 3.4. *For $M \geq 1$, let L_M be the subset of $\mathscr{D}(\mathscr{B}^{\theta,\sigma})$ given by polynomials with degree not higher than M. Then $\mathscr{B}^{\theta,\sigma}$ maps L_M into L_M.*

Then we have the following.

Proposition 3.5. *Let $\mathscr{B}^{\theta,\sigma}$ be defined as in (20) and (21). The closure in $C(\overline{\nabla}_\infty)$ of $\mathscr{B}^{\theta,\sigma}$ generates a strongly continuous, positive, conservative, contraction semigroup $\{\mathscr{T}(t)\}$ on $C(\overline{\nabla}_\infty)$.*

Proof. For $f \in C(\overline{\nabla}_\infty)$, let $r_n f = f|_{\nabla_n}$ be the restriction of f to ∇_n. Then for every $g \in \mathscr{D}(\mathscr{B}^{\theta,\sigma})$ we have $|\mathscr{B}_n^{\theta,\sigma} r_n g - r_n \mathscr{B}^{\theta,\sigma} g| \leq n^{-1} R_n$, where R_n is bounded. Hence

$$\left|\left| \mathscr{B}_n^{\theta,\sigma} r_n g - r_n \mathscr{B}^{\theta,\sigma} g \right|\right| \to 0, \qquad g \in \mathscr{D}(\mathscr{B}^{\theta,\sigma}) \tag{22}$$

as $n \to \infty$. From Proposition 3.3 and the Hille-Yosida Theorem it follows that $\mathscr{B}_n^{\theta,\sigma}$ is dissipative for every $n \geq 1$. Hence (22), together with the fact that $\left|\left| r_n g - g \right|\right| \to 0$ for $n \to \infty$ for all $g \in C(\overline{\nabla}_\infty)$, implies that $\mathscr{B}^{\theta,\sigma}$ is dissipative. Furthermore, $\mathscr{D}(\mathscr{B}^{\theta,\sigma})$ separates the points of $\overline{\nabla}_\infty$. Indeed $\varphi_m(\mathbf{z})$ is the $(m-1)$-th moment of a random variable distributed according to $\nu_{\mathbf{z}} = \sum_{i\geq 1} z_i \delta_{z_i} + (1 - \sum_{i\geq 1} z_i)\delta_0$, for $\mathbf{z} \in \overline{\nabla}_\infty$, and $\varphi_m(\mathbf{z}) = \varphi_m(\mathbf{y})$ for $m \geq 2$ implies all moments are equal, hence $\mathbf{z} = \mathbf{y}$. The Stone-Weierstrass theorem then implies that $\mathscr{D}(\mathscr{B}^{\theta,\sigma})$ is dense in $C(\overline{\nabla}_\infty)$. Proposition 3.4, together with Proposition 1.3.5 of [10], then implies that the closure of $\mathscr{B}^{\theta,\sigma}$ generates a strongly continuous contraction semigroup $\{\mathscr{T}(t)\}$ on $C(\overline{\nabla}_\infty)$. Also, since $\mathscr{B}^{\theta,\sigma}\varphi_1 = \mathscr{B}^{\theta,\sigma} 1 = 0$, $\{\mathscr{T}(t)\}$ is conservative. Finally, (22) and Theorem 1.6.1 of [10] imply the strong semigroup convergence

$$\left|\left| \mathscr{T}_n(t) r_n g - r_n \mathscr{T}(t) g \right|\right| \to 0, \qquad g \in C(\overline{\nabla}_\infty) \tag{23}$$

uniformly on bounded intervals, from which the positivity of $\{\mathscr{T}(t)\}$ follows. □

We are now ready to prove the convergence in distribution of the process of ranked relative frequencies of types.

Theorem 3.6. *Given $\nu_n \in \mathscr{P}(\Delta_n)$, let $\{Z^{(n)}(\cdot)\}$ be a sequence of Markov processes such that, for every $n \geq 2$, $Z^{(n)}(\cdot)$ is as in Proposition 3.2 with initial distribution ν_n and sample paths in $D_{\Delta_n}([0,\infty))$. Also, let $\rho_n : \Delta_n \to \nabla_\infty$ be as in (19), $Y_n(\cdot) = \rho_n(Z^{(n)}(\cdot))$ be as in Proposition 3.3, and $\{\mathscr{T}(t)\}$ be as in Proposition 3.5. Given $\nu \in \mathscr{P}(\overline{\nabla}_\infty)$, there exists a strong Markov process $Y(\cdot)$, with initial distribution ν, such that*

$$\mathbb{E}(f(Y(t+s))|Y(u), u \leq s) = \mathscr{T}(t) f(Y(s)), \qquad f \in C(\overline{\nabla}_\infty),$$

and with sample paths in $D_{\overline{\nabla}_\infty}([0,\infty))$. If also $\nu_n \circ \rho_n^{-1} \Rightarrow \nu$, then $Y_n(\cdot) \Rightarrow Y(\cdot)$ in $D_{\overline{\nabla}_\infty}([0,\infty))$ as $n \to \infty$.

Proof. The result follows from (23) and Theorem 4.2.11 of [10]. □

Remark 3.7. The statement of Theorem 3.6 can be strengthened. From [18] it follows that the sample paths of $Y(\cdot)$ belong to $C_{\overline{\nabla}_\infty}([0,\infty))$ almost surely. Then [3] (cf. Chapter 18) implies that $\rho_n(Z^{(n)}(\cdot)) \Rightarrow Y(\cdot)$ in the uniform topology. □

4 Stationarity

Denote with

$$\mathscr{P}_n^{\theta,\sigma}(\mathrm{d}\mathbf{x}) = \nu_0(\mathrm{d}x_1)\prod_{i=2}^{n}\frac{(\theta+\sigma K_{i-1})\nu_0(\mathrm{d}x_i)+\sum_{k=1}^{K_{i-1}}(n_k-\sigma)\delta_{x_k^*}(\mathrm{d}x_i)}{\theta+i-1} \tag{24}$$

the joint law of an n-sized sequential sample from the Pitman urn scheme (3), and with $p_n(\mathrm{d}x_i|\mathbf{x}_{(-i)})$ the conditional distribution in (7).

Proposition 4.1. *For $n \geq 1$, let $X^{(n)}(\cdot)$ be the $\mathbb{X}^n$-valued particle process with generator (9). Then $X^{(n)}(\cdot)$ is reversible with respect to $\mathscr{P}_n^{\theta,\sigma}$.*

Proof. Let $q(\mathbf{x},\mathrm{d}\mathbf{y})$ denote the infinitesimal transition kernel given by $q(\mathbf{x},\mathrm{d}\mathbf{y}) = \lim_{t\downarrow 0} t^{-1}\tilde{T}(t,\mathbf{x},\mathrm{d}\mathbf{y})$. When $\tilde{T}(t,\mathbf{x},\mathrm{d}\mathbf{y})$ is as in (8) and λ_n as in (11), we have

$$\begin{aligned}
\mathscr{P}_n^{\theta,\sigma}(\mathrm{d}\mathbf{x})q(\mathbf{x},\mathrm{d}\mathbf{y}) =& \mathscr{P}_n^{\theta,\sigma}(\mathrm{d}\mathbf{x})\frac{1}{n}\sum_{i=1}^{n}\lambda_n\, p_n(\mathrm{d}y_i|\mathbf{x}_{(-i)})\prod_{k\neq i}\delta_{x_k}(y_k) \\
=& \frac{1}{n}\sum_{i=1}^{n}\lambda_n\mathscr{P}_{n-1}^{\theta,\sigma}(\mathrm{d}\mathbf{x}_{-i})p_n(\mathrm{d}x_i|\mathbf{x}_{(-i)})p_n(\mathrm{d}y_i|\mathbf{x}_{(-i)})\prod_{k\neq i}\delta_{x_k}(y_k) \\
=& \frac{1}{n}\sum_{i=1}^{n}\lambda_n\mathscr{P}_{n-1}^{\theta,\sigma}(\mathrm{d}\mathbf{y}_{-i})p_n(\mathrm{d}x_i|\mathbf{y}_{-i})p_n(\mathrm{d}y_i|\mathbf{y}_{-i})\prod_{k\neq i}\delta_{y_k}(x_k) \\
=& \mathscr{P}_n^{\theta,\sigma}(\mathrm{d}\mathbf{y})\frac{1}{n}\sum_{i=1}^{n}\lambda_n p_n(\mathrm{d}x_i|\mathbf{y}_{-i})\prod_{k\neq i}\delta_{y_k}(x_k)
\end{aligned} \tag{25}$$

which is $\mathscr{P}_n^{\theta,\sigma}(\mathrm{d}\mathbf{y})q(\mathbf{y},\mathrm{d}\mathbf{x})$, giving the result. □

Integrating out with respect to $\mathbf{x}$ both sides of (25) immediately yields the following.

Corollary 4.2. *Let $X^{(n)}(\cdot)$ be the $\mathbb{X}^n$-valued particle process with generator (9). Then $X^{(n)}(\cdot)$ has invariant law $\mathscr{P}_n^{\theta,\sigma}$.*

We turn now to the stationary properties of the infinite-dimensional process of Theorem 3.6.

Proposition 4.3. *Let $Y(\cdot)$ be as in Theorem 3.6. Then $Y(\cdot)$ has at most one stationary distribution.*

Proof. See Appendix. □

Theorem 4.4. *Let $Y(\cdot)$ be as in Theorem 3.6. Then the two-parameter Poisson-Dirichlet distribution $\Pi_{\theta,\sigma}$ is an invariant law for $Y(\cdot)$.*

Proof. In view of Corollary 4.2 assume, without loss of generality, that $\mathscr{P}_n^{\theta,\sigma}$ is the initial law of $X^{(n)}(\cdot)$. Hence, for every $n \geq 1$ and every $t \geq 0$, $X^{(n)}(t)$ is an n-sized i.i.d. sample from $\mu \sim \tilde{\Pi}_{\theta,\sigma}$, given μ. But $n^{-1}\sum_{i\geq 1}\delta_{x_i(t)} \Rightarrow \mu$ almost surely, with $\mu \sim \tilde{\Pi}_{\theta,\sigma}$ (see [1], Lemma 2.15). Also, $\mu = \sum_{j\geq 1} q_j\delta_{Y_j}$ almost surely, where Y_j are i.i.d. samples from a common diffuse distribution ν_0, and $(q_1,q_2,\ldots) \sim \mathrm{GEM}(\theta,\sigma)$, hence $(q_{(1)},q_{(2)},\ldots) \sim \Pi_{\theta,\sigma}$ (see [20], Proposition 11 and subsequent discussion). It follows that $q_{(j)}$ is the frequency of the j-th largest species in an infinite-sized sample from (3), from which $Y(t) \sim \Pi_{\theta,\sigma}$, for every $t \geq 0$. Recall now from Proposition

3.5 that $\mathscr{D}(\mathscr{B}^{\theta,\sigma})$ separates points of $\overline{\nabla}_\infty$. Then Theorems 3.4.5 and 4.1.6 of [10] respectively imply that $\mathscr{D}(\mathscr{B}^{\theta,\sigma})$ is separating and that $Y(\cdot)$ is the only solution of the $C_{\overline{\nabla}_\infty}([0,\infty))$-martingale problem for $(\mathscr{B}^{\theta,\sigma},\nu)$. The fact that $\Pi_{\theta,\sigma}$ is an invariant law for $Y(\cdot)$ is then implied by Lemma 4.9.1 of [10]. □

Remark 4.5. Let $V_k, k \geq 1$ be as in (1) and note that $V_1 \to 1$ in mean square for θ and σ jointly converging to zero, since

$$\mathbb{E}_{\theta,\sigma}(V_1) = \frac{1-\sigma}{1+\theta}, \qquad \mathrm{Var}_{\theta,\sigma}(V_1) = \frac{(1-\sigma)(\theta+\sigma)}{(1+\theta)^2(2+\theta)}.$$

Then for $\theta, \sigma = 0$, the distribution $\Pi_{\theta,\sigma}$ puts all of its mass to the point of $\overline{\nabla}_\infty$ given by $(1,0,0,\ldots)$. □

The following proposition shows that the limiting diffusion is ergodic.

Proposition 4.6. *Let $Y(\cdot)$ be as in Theorem 3.6. Then $Y(\cdot)$ is ergodic in the sense that*

$$\left|\left|\mathscr{T}(t)g - \int_{\overline{\nabla}_\infty} g\,\mathrm{d}\mu\right|\right| \to 0, \qquad g \in C(\overline{\nabla}_\infty) \tag{26}$$

as $t \to \infty$, where μ is the unique stationary distribution.

Proof. See Appendix. □

Since the two-parameter Poisson-Dirichlet distribution is concentrated on ∇_∞ (cf., e.g., [22]), it follows that (26) can be modified to

$$\left|\left|\mathscr{T}(t)g - \int_{\nabla_\infty} g\,\mathrm{d}\Pi_{\theta,\sigma}\right|\right| \to 0, \qquad g \in C(\overline{\nabla}_\infty).$$

Hence eventually the process ends up in ∇_∞ with probability one for any initial state belonging to $\overline{\nabla}_\infty$.

Acknowledgements

The authors are grateful to an anonymous referee for valuable comments which greatly improved the paper.

Appendix

Proof of Proposition 3.4

Denote with ∂_i the partial derivative with respect to z_i. For φ_m we have

$$\mathscr{B}^{\theta,\sigma}\varphi_m = \sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j)\partial_{ij}\varphi_m - \sum_{i=1}^{\infty}(\theta z_i + \sigma)\partial_i\varphi_m \tag{27}$$

$$
\begin{aligned}
&= \sum_{i=1}^{\infty} z_i \partial_{ii}\varphi_m - \left[\sum_{i\neq j} z_i z_j \partial_{ij}\varphi_m + \sum_{i=1}^{\infty} z_i^2 \partial_{ii}\varphi_m \right] - \sum_{i=1}^{\infty} (\theta z_i + \sigma)\partial_i \varphi_m \\
&= [m(m-1) - m\sigma]\varphi_{m-1} - [m(m-1) + m\theta]\varphi_m.
\end{aligned}
$$

For $\varphi_{m_1} \cdots \varphi_{m_k}$ we have

$$
\mathscr{B}^{\theta,\sigma}(\varphi_{m_1} \cdots \varphi_{m_k}) = \sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j)\partial_{ij}(\varphi_{m_1} \cdots \varphi_{m_k}) - \sum_{i=1}^{\infty} (\theta z_i + \sigma)\partial_i(\varphi_{m_1} \cdots \varphi_{m_k}).
$$

The first term on the right-hand side equals

$$
\begin{aligned}
&\sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j)\partial_j \sum_{h=1}^{k} \partial_i \varphi_{m_h} \prod_{\ell\neq h} \varphi_{m_\ell} \\
&= \sum_{h=1}^{k} \sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j) \left[\partial_{ij}\varphi_{m_h} \prod_{\ell\neq h} \varphi_{m_\ell} + \sum_{q\neq h} \partial_i \varphi_{m_h} \partial_j \varphi_{m_q} \prod_{\ell\neq h,q} \varphi_{m_\ell} \right] \\
&= \sum_{h=1}^{k} \left[\sum_{i=1}^{\infty} z_i \partial_{ii}\varphi_{m_h} \prod_{\ell\neq h} \varphi_{m_\ell} + \sum_{i=1}^{\infty} z_i \sum_{q\neq h} \partial_i \varphi_{m_h} \partial_i \varphi_{m_q} \prod_{\ell\neq h,q} \varphi_{m_\ell} \right. \\
&\quad - \sum_{i\neq j}^{\infty} z_i z_j \partial_{ij}\varphi_{m_h} \prod_{\ell\neq h} \varphi_{m_\ell} - \sum_{i=1}^{\infty} z_i^2 \partial_{ii}\varphi_{m_h} \prod_{\ell\neq h} \varphi_{m_\ell} \\
&\quad \left. - \sum_{i,j=1}^{\infty} z_i z_j \sum_{q\neq h} \partial_i \varphi_{m_h} \partial_j \varphi_{m_q} \prod_{\ell\neq h,q} \varphi_{m_\ell} \right] \\
&= \sum_{h=1}^{k} \left[m_h(m_h-1)\varphi_{m_h-1} \prod_{\ell\neq h} \varphi_{m_\ell} + \sum_{q\neq h} m_h m_q \varphi_{m_h+m_q-1} \prod_{\ell\neq h,q} \varphi_{m_\ell} \right. \\
&\qquad \left. - m_h(m_h-1) \prod_{\ell=1}^{k} \varphi_{m_\ell} - \sum_{q\neq h} m_h m_q \varphi_{m_h} \varphi_{m_q} \prod_{\ell\neq h,q} \varphi_{m_\ell} \right].
\end{aligned}
$$

As for the second term, we have

$$
\begin{aligned}
\sum_{i=1}^{\infty} (\theta z_i + \sigma)\partial_i(\varphi_{m_1} \cdots \varphi_{m_k}) &= \sum_{i=1}^{\infty} (\theta z_i + \sigma) \sum_{h=1}^{k} \partial_i \varphi_{m_h} \prod_{\ell\neq h} \varphi_{m_\ell} \\
&= \theta \sum_{h=1}^{k} m_h \prod_{\ell=1}^{k} \varphi_{m_\ell} + \sigma \sum_{h=1}^{k} m_h \varphi_{m_h-1} \prod_{\ell\neq h} \varphi_{m_\ell}.
\end{aligned}
$$

Hence

$$
\begin{aligned}
\mathscr{B}^{\theta,\sigma}(\varphi_{m_1} \cdots \varphi_{m_k}) = \sum_{h=1}^{k} \Bigg[& m_h(m_h-1)\varphi_{m_h-1} \prod_{\ell\neq h} \varphi_{m_\ell} \\
& + \sum_{q\neq h} m_h m_q \varphi_{m_h+m_q-1} \prod_{\ell\neq h,q} \varphi_{m_\ell} - \sum_{h=1}^{k} m_h(m_h-1) \prod_{\ell=1}^{k} \varphi_{m_\ell}
\end{aligned} \tag{28}
$$

$$
\begin{aligned}
&- \sum_{q\neq h} m_h m_q \prod_{\ell=1}^{k} \varphi_{m_\ell} \Bigg] - \theta \sum_{h=1}^{k} m_h \prod_{\ell=1}^{k} \varphi_{m_\ell} - \sigma \sum_{h=1}^{k} m_h \varphi_{m_h-1} \prod_{\ell\neq h} \varphi_{m_\ell} \\
&= \sum_{h=1}^{k} [m_h(m_h-1) - m_h\sigma] \varphi_{m_h-1} \prod_{\ell\neq h} \varphi_{m_\ell} + \sum_{h=1}^{k} \sum_{q\neq h} m_h m_q \varphi_{m_h+m_q-1} \prod_{\ell\neq h,q} \varphi_{m_\ell} \\
&- \sum_{h=1}^{k} \left\{ \left[m_h(m_h-1) + m_h\theta \right] + \sum_{q\neq h} m_h m_q \right\} \prod_{\ell=1}^{k} \varphi_{m_\ell}
\end{aligned}
$$

so that $\mathscr{B}^{\theta,\sigma} : L_M \to L_M$.

The two following proofs are modifications of proofs in [9] adapted for the two-parameter case.

Proof of Proposition 4.3

If μ is a stationary distribution for the Markov process with generator $\mathscr{B}^{\theta,\sigma}$ we have

$$
\int_{\overline{\nabla}_\infty} \mathscr{B}^{\theta,\sigma} f \, \mathrm{d}\mu = 0 \tag{29}
$$

for every $f \in \mathscr{D}(\mathscr{B}^{\theta,\sigma})$ (cf. [10], Proposition 4.9.2). From (27) we have

$$
\int_{\overline{\nabla}_\infty} \mathscr{B}^{\theta,\sigma} \varphi_2 \mathrm{d}\mu = \int_{\overline{\nabla}_\infty} [(2-2\sigma) - (2+2\theta)\varphi_2] \mathrm{d}\mu
$$

so that

$$
\int_{\overline{\nabla}_\infty} \varphi_2 \mathrm{d}\mu = \frac{1-\sigma}{1+\theta} \tag{30}
$$

and assuming $\int_{\overline{\nabla}_\infty} \varphi_{m-1} \mathrm{d}\mu$ is determined, we have

$$
\begin{aligned}
\int_{\overline{\nabla}_\infty} \mathscr{B}^{\theta,\sigma} \varphi_m \mathrm{d}\mu &= \int_{\overline{\nabla}_\infty} \Big([m(m-1) - m\sigma] \varphi_{m-1} - [m(m-1) + m\theta] \varphi_m \Big) \mathrm{d}\mu \\
&= [m(m-1) - m\sigma] \int_{\overline{\nabla}_\infty} \varphi_{m-1} \mathrm{d}\mu - [m(m-1) + m\theta] \int_{\overline{\nabla}_\infty} \varphi_m \mathrm{d}\mu
\end{aligned}
$$

from which

$$
\int_{\overline{\nabla}_\infty} \varphi_m \mathrm{d}\mu = \frac{m-1-\sigma}{m-1+\theta} \int_{\overline{\nabla}_\infty} \varphi_{m-1} \mathrm{d}\mu = \frac{(1-\sigma)_{(m-1)}}{(1+\theta)_{(m-1)}}.
$$

Hence $\int_{\overline{\nabla}_\infty} \varphi_m \mathrm{d}\mu$ is the $(m-1)$-th moment of a Beta$(1-\sigma, \theta+\sigma)$. In general, assume degree$(\varphi_{m_1} \cdots \varphi_{m_k}) = M$ and all $\int_{\overline{\nabla}_\infty} \varphi_{m'_1} \cdots \varphi_{m'_\ell} \mathrm{d}\mu$, with $\ell \leq k$, such that degree$(\varphi_{m'_1} \cdots \varphi_{m'_\ell}) = M-1$ are determined. From (28) we have

$$
\int_{\overline{\nabla}_\infty} \mathscr{B}^{\theta,\sigma} (\varphi_{m_1} \cdots \varphi_{m_k}) \mathrm{d}\mu
$$

$$= \sum_{h=1}^{k} C_{\sigma,h} \int_{\overline{\nabla}_\infty} \varphi_{m_h-1} \prod_{\ell \neq h} \varphi_{m_\ell} \mathrm{d}\mu + \sum_{q \neq h} C'_{h,q} \int_{\overline{\nabla}_\infty} \varphi_{m_h+m_q-1} \prod_{\ell \neq h,q} \varphi_{m_\ell} \mathrm{d}\mu$$
$$- C''_{\theta,1,\ldots,k} \int_{\overline{\nabla}_\infty} \varphi_{m_1} \cdots \varphi_{m_k} \mathrm{d}\mu$$

for appropriate constants $C_{\sigma,h}, C'_{h,q}, C''_{\theta,1,\ldots,k}$ depending on $\theta, \sigma, m_1, \ldots, m_k$, so that

$$\int_{\overline{\nabla}_\infty} \varphi_{m_1} \cdots \varphi_{m_k} \mathrm{d}\mu$$
$$= \sum_{h=1}^{k} \frac{C_{\sigma,h}}{C''_{\theta,1,\ldots,k}} \int_{\overline{\nabla}_\infty} \varphi_{m_h-1} \prod_{\ell \neq h} \varphi_{m_\ell} \mathrm{d}\mu + \sum_{q \neq h} \frac{C'_{h,q}}{C''_{\theta,1,\ldots,k}} \int_{\overline{\nabla}_\infty} \varphi_{m_h+m_q-1} \prod_{\ell \neq h,q} \varphi_{m_\ell} \mathrm{d}\mu$$

is determined. It follows that $\int_{\overline{\nabla}_\infty} \varphi_{m_1} \cdots \varphi_{m_k} \mathrm{d}\mu$ is uniquely determined for all $m_1, \ldots, m_k \geq 2$ and $k \geq 1$, and so is μ.

Proof of Proposition 4.6

It suffices to show the result for all $\varphi_{m_1} \ldots \varphi_{m_k}$, for $k \geq 1$, $\varphi_m(\mathbf{z}) = \sum_{i \geq 1} z_i^m$ and $\mathbf{z} \in \overline{\nabla}_\infty$, and use the fact that the algebra generated by $(1, \varphi_2, \varphi_3, \ldots)$ is dense in $C(\overline{\nabla}_\infty)$, as shown in the proof of Proposition 3.5. From (27) we have

$$\mathscr{B}^{\theta,\sigma} \varphi_2 = (2 - 2\sigma) - (2 + 2\theta)\varphi_2$$

and letting $\{\mathscr{T}(t)\}$ be the semigroup of Proposition 3.5 for $f \in \mathscr{D}(\mathscr{B}^{\theta,\sigma})$ we have that $\mathrm{d}T(t)f/\mathrm{d}t = \mathscr{B}^{\theta,\sigma} T(t)f$ for $t \geq 0$ (cf. [10], Proposition 1.1.5). Hence

$$\frac{\mathrm{d}\mathscr{T}(t)\varphi_2}{\mathrm{d}t} = 2(1 - \sigma) - 2(1 + \theta)\mathscr{T}(t)\varphi_2$$

from which the general solution is

$$\mathscr{T}(t)\varphi_2 = 2(1 - \sigma)e^{-2(1+\theta)t} \int_0^t e^{2(1+\theta)s} \mathrm{d}s + \kappa e^{-2(1+\theta)t}$$

where κ is a constant and the initial condition $\mathscr{T}(0)\varphi_2 = \varphi_2$ yields $\kappa = \varphi_2$ from which

$$\mathscr{T}(t)\varphi_2 = \frac{1 - \sigma}{1 + \theta} e^{-2(1+\theta)t} [e^{2(1+\theta)t} - 1] + \varphi_2 e^{-2(1+\theta)t}$$

which converges to $(1 - \sigma)/(1 + \theta)$ uniformly on $z \in \overline{\nabla}_\infty$ as $t \to \infty$. This and (30) imply $\left|\left|\mathscr{T}(t)\varphi_2 - \int_{\overline{\nabla}_\infty} \varphi_2 \mathrm{d}\mu\right|\right| \to 0$ as $t \to \infty$, where μ here is the stationary distribution $\tilde{\Pi}_{\theta,\sigma}$. Take now $f_M = \varphi_{m_1} \ldots \varphi_{m_k}$ such that degree$(\varphi_{m_1} \cdots \varphi_{m_k}) = M$, and assume (26) holds for all $f_{M-1} = \varphi_{m'_1} \ldots \varphi_{m'_\ell}$, $\ell \leq k$, such that degree$(\varphi_{m'_1} \cdots \varphi_{m'_\ell}) = M - 1$. From (28) we have

$$\mathscr{B}^{\theta,\sigma} f_M = f_{M-1} - \kappa' f_M \tag{31}$$

so that, as above,

$$\mathscr{T}(t)f_M = e^{-\kappa' t} \int_0^t e^{\kappa' s} \mathscr{T}(s) f_{M-1} \mathrm{d}s + e^{-\kappa' t} f_M.$$

Also, from (29) applied to (31) we have $\int f_{M-1}\mathrm{d}\mu - \kappa' \int f_M \mathrm{d}\mu = 0$ so that

$$\int_{\overline{\nabla}_\infty} f_M \mathrm{d}\mu = e^{-\kappa' t} \int_0^t e^{\kappa' s} \int_{\overline{\nabla}_\infty} f_{M-1}\mathrm{d}\mu\,\mathrm{d}s + e^{-\kappa' t} \int_{\overline{\nabla}_\infty} f_M \mathrm{d}\mu.$$

Then we have

$$\left|\left|\mathscr{T}(t)f_M - \int_{\overline{\nabla}_\infty} f_M \mathrm{d}\mu\right|\right| \le \kappa'' e^{-\kappa' t} + e^{-\kappa' t} \int_0^t e^{\kappa' s} \left|\left|\mathscr{T}(s)f_{M-1} - \int_{\overline{\nabla}_\infty} f_{M-1}\mathrm{d}\mu\right|\right| \mathrm{d}s$$

giving the result.

References

[1] Aldous, D. (1985). Exchangeability and related topics. *École d'été de probabilités de Saint-Flour XIII. Lecture Notes in Math.* **1117**. Springer, Berlin. MR0883646

[2] Bertoin, J. (2006). *Random fragmentation and coagulation processes*. Cambridge University Press, Cambridge. MR2253162

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

[4] Blackwell, D. and 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 Math.* **1541**. Springer, Berlin. MR1242575

[6] Donnelly, P. and Kurtz, T.G. (1996). A countable representation of the Fleming-Viot measure-valued diffusion. *Ann. Probab.* **24**, 698–742. MR1404525

[7] Donnelly, P. and Kurtz, T.G. (1999a). Genealogical processes for Fleming-Viot models with selection and recombination. *Ann. Appl. Probab.* **9**, 1091–1148. MR1728556

[8] Donnelly, P. and Kurtz, T.G. (1999b). Particle representation for measure-valued population models. *Ann. Probab.* **27**, 166–205. MR1681126

[9] Ethier, S.N. and Kurtz, T.G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.* **13**, 429–452. MR0615945

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

[11] Ethier, S.N. and Kurtz, T.G. (1992). On the stationary distribution of the neutral diffusion model in population genetics. *Ann. Appl. Probab.* **2**, 24–35. MR1143391

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

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

[14] GELFAND, A.E. and SMITH, A.F.M. (1990). Sampling-based approaches to calculating marginal densities. *J. Amer. Statist. Assoc.* **85**, 398–409. MR1141740

[15] ISHWARAN, H. and JAMES, L. (2001). Gibbs sampling methods for stick-breaking priors. *J. Amer. Statist. Assoc.* **96**, 161–173. MR1952729

[16] KINGMAN, J.F.C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser B* **37**, 1–22. MR0368264

[17] LIJOI, A. and PRÜNSTER, I. (2009). Models beyond the Dirichlet process. To appear in Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. (Eds.), *Bayesian Nonparametrics*, Cambridge University Press.

[18] PETROV, L. (2009). Two-parameter family of diffusion processes in the Kingman simplex. *Funct. Anal. Appl.*, in press.

[19] PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory and Related Fields* **102**, 145–158. MR1337249

[20] PITMAN, J. (1996). Some developments of the Blackwell-MacQueen urn scheme. In Ferguson, T.S., Shapley, L.S. and MacQueen, J.B., *Statistics, probability and game theory*, IMS Lecture Notes - Monograph Series **30**, 245–267. Hayward, CA: Institute of Mathematical Statistics. MR1481784

[21] PITMAN, J. (2006). Combinatorial stochastic processes. *École d'été de probabilités de Saint-Flour XXXII. Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

[22] PITMAN, J. and Yor, M. (1997). The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900. MR1434129

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

[24] SETHURAMAN, J. (1994). A constructive definition of the Dirichlet process prior. *Statist. Sinica* **2**, 639–650. MR1309433

[25] TEH, Y.W. and JORDAN, M.I. (2009). Bayesian Nonparametrics in Machine Learning. To appear in Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. (Eds.), *Bayesian Nonparametrics*, Cambridge University Press.