

# SPECIES DYNAMICS IN THE TWO-PARAMETER POISSON–DIRICHLET DIFFUSION MODEL

MATTEO RUGGIERO,\* *University of Torino and Collegio Carlo Alberto*

## Abstract

The recently introduced two-parameter infinitely-many-neutral-alleles model extends the celebrated one-parameter version (which is related to Kingman’s distribution) to diffusive two-parameter Poisson–Dirichlet frequencies. In this paper we investigate the dynamics driving the species heterogeneity underlying the two-parameter model. First we show that a suitable normalization of the number of species is driven by a critical continuous-state branching process with immigration. Secondly, we provide a finite-dimensional construction of the two-parameter model, obtained by means of a sequence of Feller diffusions of Wright–Fisher flavor which feature finitely many types and inhomogeneous mutation rates. Both results provide insight into the mathematical properties and biological interpretation of the two-parameter model, showing that it is structurally different from the one-parameter case in that the frequency dynamics are driven by state-dependent rather than constant quantities.

*Keywords:* Alpha diversity; infinite-alleles model; infinite-dimensional diffusion; mutation rate; Poisson–Dirichlet distribution; weak convergence

2010 Mathematics Subject Classification: Primary 60J60

Secondary 60G57; 92D25

## 1. Introduction

The two-parameter infinitely-many-neutral-alleles model is a family of infinite-dimensional diffusion processes, introduced in Petrov (2009) and further investigated in Ruggiero and Walker (2009) and Feng and Sun (2010), which extends the celebrated one-parameter version, formulated in Watterson (1976) and characterized in Ethier and Kurtz (1981). Throughout the paper, we will refer to the one- and two-parameter infinitely-many-neutral-alleles models simply as the one- and two-parameter models. More specifically, let

$$\bar{V}_\infty = \left\{ z \in [0, 1]^\infty : z_1 \geq z_2 \geq \dots \geq 0, \sum_{i=1}^\infty z_i \leq 1 \right\} \tag{1}$$

be the closure of the infinite-dimensional ordered simplex, and define, for constants  $0 \leq \alpha < 1$  and  $\theta > -\alpha$ , the second-order differential operator

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

where  $\delta_{ij}$  denotes the Kronecker delta, acting on a certain dense subalgebra of the space  $C(\bar{V}_\infty)$  of continuous functions on  $\bar{V}_\infty$ . Then the closure of  $\mathcal{B}$  generates a strongly continuous

Received 13 September 2012; revision received 26 April 2013.

\* Postal address: Department of Economics and Statistics, University of Torino, Corso Unione Sovietica 218/bis, 10134 Torino, Italy. Email address: [matteo.ruggiero@unito.it](mailto:matteo.ruggiero@unito.it)

semigroup of contractions on  $C(\overline{V}_\infty)$ , and the sample paths of the associated process are almost surely (a.s.) continuous functions from  $[0, \infty)$  to  $\overline{V}_\infty$ . Such a process can be thought of as describing the temporal evolution of the decreasingly ordered allelic frequencies  $(z_1, z_2, \dots)$  at a particular locus in an ideally infinite population with infinitely many possible types or species. See Feng (2010) for a review of infinitely-many-alleles models. Recently, Feng *et al.* (2011) determined the transition density for the two-parameter case. See also Borodin and Olshanski (2009) for a general construction related to Petrov (2009), and Ruggiero *et al.* (2013) for a partially related model with diffusive parameter  $\theta$ .

As shown in Petrov (2009) and Feng and Sun (2010), the two-parameter model is reversible and ergodic with respect to the Poisson–Dirichlet distribution with parameters  $(\theta, \alpha)$ , henceforth denoted as  $\text{PD}(\theta, \alpha)$ . First introduced in Perman *et al.* (1992) (see also Pitman (1995) and Pitman and Yor (1997)), this extends the one-parameter version  $\text{PD}(\theta) := \text{PD}(\theta, 0)$  due to Kingman (1975), and has found numerous applications in several fields. See, for example, Bertoin (2006) for fragmentation and coalescent theory, Pitman (2006) for excursion theory and combinatorics, Aoki (2008) for economics, Lijoi and Prünster (2009) for Bayesian inference, Teh and Jordan (2009) for machine learning, and Feng (2010) for population genetics. Both these random discrete distributions arise as the decreasingly ordered weights of a Dirichlet process (see Ferguson (1973)), when  $\alpha = 0$ , and of a two-parameter Poisson–Dirichlet (or Pitman–Yor) process (see Pitman (1995)), respectively. Alternatively, they can be constructed by means of the following so-called stick-breaking procedure, also known as the residual allocation model. Consider a sequence of random variables  $(V_1, V_2, \dots)$  obtained by setting

$$V_1 = W_1, \quad V_n = W_n \prod_{i=1}^{n-1} (1 - W_i), \quad W_i \stackrel{\text{ind}}{\sim} \text{Beta}(1 - \alpha, \theta + i\alpha),$$

where  $\stackrel{\text{ind}}{\sim}$  denotes independence,  $0 \leq \alpha < 1$ , and  $\theta > -\alpha$ . The vector  $(V_1, V_2, \dots)$  is said to have the  $\text{GEM}(\theta, \alpha)$  distribution, named after Griffiths, Engen, and McCloskey, while the vector of descending order statistics,  $(V_{(1)}, V_{(2)}, \dots)$ , is said to have the  $\text{PD}(\theta, \alpha)$  distribution. See Feng and Wang (2007) for an infinite-dimensional diffusion process related to GEM distributions.

Besides sharing the above stick-breaking construction strategy, it is well known that the difference between these two random discrete distributions is structural and does not simply rely on a different parametrization. For example, the distribution  $\text{PD}(\theta)$  can be obtained by ranking and normalizing the jumps of a gamma subordinator, whereas the  $\text{PD}(\theta, \alpha)$  distribution is obtained by performing the same operation on the jumps of a stable subordinator and appropriately mixing over the law of the normalizing factor (see Pitman (2003)). See Section 2. Furthermore, the  $\text{PD}(\theta)$  distribution is obtained as the weak limit of a Dirichlet-distributed vector of frequencies (see Kingman (1975)), while a similar construction for the two-parameter case is not available. For their diffusive counterparts, the properties of the one-parameter model, related to the  $\text{PD}(\theta)$  distribution, are well understood, whereas numerous open questions remain regarding the two-parameter model. In particular, given the above considerations, it is not surprising that a finite-dimensional construction of the process with operator (2), in terms of a sequence of finite-dimensional diffusion processes, is currently available only when  $\alpha = 0$ . To be more precise, consider the usual approximating diffusion for the Wright–Fisher discrete genetic model with  $n$  selectively neutral alleles and symmetric mutation. This corresponds to

the operator

$$\mathcal{B}_n = \frac{1}{2} \sum_{i,j=1}^n z_i (\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} + \frac{1}{2} \sum_{i=1}^n b_i^{(n)}(z) \frac{\partial}{\partial z_i}, \quad (3)$$

acting on a suitable subspace of  $C^2(\nabla_n)$ , with

$$\nabla_n = \left\{ z \in \bar{\nabla}_\infty : z_{n+1} = 0, \sum_{i=1}^n z_i = 1 \right\},$$

with drift components

$$b_i^{(n)}(z) = \frac{\theta}{n-1} (1 - z_i) - \theta z_i, \quad \theta > 0. \quad (4)$$

Ethier and Kurtz (1981) formalized the conditions under which the sequence of processes with operators defined by (3)–(4) converges in distribution to the one-parameter model, with operator obtained by setting  $\alpha = 0$  in (2). As anticipated, a similar construction for the case in which  $0 < \alpha < 1$  and  $\theta > -\alpha$  is currently unavailable. Two different sequential constructions of the two-parameter model are given in Petrov (2009) and Ruggiero and Walker (2009). In Section 3.1 we will argue that, despite offering interesting reads of the two parameter model, neither construction provides particular insight for the interpretation of the species dynamics underlying the infinite-dimensionality structure. In particular, this is due to the fact that both are based on finitely many items. The problem at hand could then be rephrased as that of understanding from which Wright–Fisher-type mechanism, if any, the two-parameter model comes from. While the importance of providing a particle construction lies in the fact that the individual dynamics are dealt with explicitly, the importance of a sequential construction by means of Wright–Fisher-type diffusions lies in the genetic interpretation we would yield from the specification of the mutation rates at the  $n$ th step of the sequence. Such an interpretation is clear in the case of (4), whereby each type has the same chance of mutating (cf. also (14) below), but is somewhat obscure regarding the role of  $\alpha$  in (2), especially in terms of its effect on finitely many types. This role would be, at least partially, clarified by identifying suitable mutation rates, which are basic building blocks of the model and give important information on the reproductive mechanism of the underlying population. Historically, the (chrono)logical process has been the opposite, namely, diffusion approximations were introduced for dealing mathematically in a simpler way with multitype discrete models such as Wright–Fisher processes. But recent advances, stimulated by neighboring research fields, have provided the infinite-dimensional diffusion without identifying its finite-dimensional source, thus leaving an interpretational gap.

Motivated by these considerations, the purpose of this paper is to investigate what lies underneath the infinite dimensionality of the two-parameter model in terms of the forces driving the species dynamics. We pursue this task in two different ways. First, we derive an  $\alpha$ -diversity diffusion for the two-parameter model. This is a continuous-time, continuous-state extension of the corresponding notion for Poisson–Kingman models (see Pitman (2003)), and describes the dynamics of the suitably normalized number of species in the underlying population. In Section 2 we show that such diffusion for the two-parameter model is a critical continuous-state branching process with immigration, and we discuss a corresponding quantity for the one-parameter case. Second, we find explicit transition rates for the mutation process which gives rise to the two-parameter model and provide a sequential construction for the limiting process in terms of finite-dimensional diffusions similar to those given in (3)–(4) for the one-parameter case. In Section 3.1 we collect some brief considerations on the problem and the

fact that the existing constructions do not provide enough insight from a biological point of view. In Section 3.2 we identify mutation rates that yield the convergence result; these turn out to depend on the current species abundances. By means of some additional restrictions, we formalize a sequential construction where each term of the sequence is a Feller diffusion on a finite-dimensional subspace of  $\bar{V}_\infty$ .

In achieving the two aforementioned goals, we are able to highlight a key difference between the one- and two-parameter models, conveniently summarized by saying that the species dynamics of the former are driven by constant terms, whereas those of the latter are driven by density-dependent quantities.

### 2. Heterogeneity in the two-parameter model

The notion of  $\alpha$ -diversity was introduced in Pitman (2003) for exchangeable partitions induced by random discrete distributions of Poisson–Kingman type. Let  $\text{PK}(\varrho|t)$  denote the distribution of the weights  $(P_i)_{i \geq 1} = (J_i/T)_{i \geq 1}$  determined by the ranked points  $J_i$  of a Poisson process with Lévy density  $\varrho$ , given  $T = t$ . A Poisson–Kingman distribution with Lévy density  $\varrho$  and mixing distribution  $\gamma$  on  $(0, \infty)$ , denoted by  $\text{PK}(\varrho, \gamma)$ , is defined as the mixture

$$\text{PK}(\varrho, \gamma) = \int_0^\infty \text{PK}(\varrho|t)\gamma(dt).$$

For instance, the  $\text{PD}(\theta, \alpha)$  distribution is obtained as a  $\text{PK}(\varrho_\alpha, \gamma_{\theta, \alpha})$  model, for  $0 < \alpha < 1$  and  $\theta > -\alpha$ , where  $\varrho_\alpha$  is the Lévy density of a stable subordinator of index  $\alpha$ ,  $\gamma_{\theta, \alpha}$  is

$$\gamma_{\theta, \alpha}(dt) = \frac{\Gamma(\theta + 1)}{\Gamma(\theta/\alpha + 1)} t^{-\theta} f_\alpha(t) dt,$$

and  $f_\alpha(t)$  is the density of a positive stable random variable of index  $\alpha$ . A given exchangeable random partition of  $\mathbb{N}$  induced by a Poisson–Kingman distribution, i.e. such that its ranked class frequencies have distribution  $\text{PK}(\varrho, \gamma)$ , is said to have  $\alpha$ -diversity  $S$  if and only if there exists a random variable  $S$ , with  $0 < S < \infty$  a.s., such that

$$\lim_{n \rightarrow \infty} \frac{K_n}{n^\alpha} = S \quad \text{a.s.}, \tag{5}$$

where  $K_n$  is the number of classes of the partition restricted to  $\{1, \dots, n\}$ . For instance, in the case of a  $\text{PD}(\theta, \alpha)$  partition, we have  $S = T^{-\alpha}$ , where  $T$  has distribution  $\gamma_{\theta, \alpha}$ . See Pitman (2003, Proposition 13).

The idea of extending the concept of  $\alpha$ -diversity from random distributions on simplices to a continuous-time, continuous-state framework has been formulated in Ruggiero *et al.* (2013), where a certain rescaled and inhomogeneous random walk on the integers, which tracks the dynamics of the number of species in a normalized inverse-Gaussian population, is shown to converge to a certain one-dimensional diffusion process on  $(0, \infty)$ . Here we derive an  $\alpha$ -diversity diffusion for the two-parameter model, with the aim of providing insight into the species dynamics underlying the infinite-dimensional process. The derivation is based on the particle construction given in Ruggiero and Walker (2009), which we briefly recall here for the ease of the reader. Let  $X^{(n)} = (X_1, \dots, X_n)$  be a sample from a two-parameter Poisson–Dirichlet process, or, equivalently (cf. Pitman (1995)), from the generalized Pólya urn scheme

given by  $X_1 \sim P_0$  and

$$\begin{aligned} &\mathbb{P}\{X_{i+1} \in \cdot \mid X_1, \dots, X_i\} \\ &= \frac{\theta + \alpha K_i}{\theta + i} P_0(\cdot) + \frac{1}{\theta + i} \sum_{j=1}^{K_i} (n_j - \alpha) \delta_{X_j^*}(\cdot) \quad \text{for } i = 2, \dots, n - 1. \end{aligned} \tag{6}$$

Here  $P_0$  is a nonatomic probability measure on the space of observables (e.g. a Polish space),  $K_i \leq i$  denotes the number of distinct elements  $(X_1^*, \dots, X_{K_i}^*)$  observed in  $(X_1, \dots, X_i)$ , and  $\delta_{X_j^*}$  is a point mass at  $X_j^*$ . A simple way to make the sample into a Markov chain with fixed marginals is the following. Let  $X^{(n)}$  be updated at discrete times by replacing a uniformly chosen coordinate. Conditionally on  $K_n = k$  at the current state, and exploiting the exchangeability of the sample, the incoming particle will be of a new type with probability  $(\theta + \alpha k_r)/(\theta + n - 1)$  and will be a copy of one still in the vector after the removal with probability  $(n - 1 - \alpha k_r)/(\theta + n - 1)$ , where  $k_r$  is the value of  $k$  after the removal.

The following proposition recalls, in a discrete-parameter version, the relation between the above described particle chain  $\{X^{(n)}(m), m \in \mathbb{N}_0\}$  and the two-parameter model. For notational simplicity, we omit here the details about the domain of the limiting operator (cf. (30)–(32) below). Here and throughout ‘ $\xrightarrow{D}$ ’ denotes convergence in distribution and  $C_A([0, \infty))$  denotes the space of continuous functions from  $[0, \infty)$  to the space  $A$ .

**Proposition 1.** (Ruggiero and Walker (2009).) *Let  $Z(\cdot)$  be the two-parameter model corresponding to the operator  $\mathcal{B}$  as in (2). Also, let  $\{X^{(n)}(m), m \in \mathbb{N}_0\}$  be the particle chain described above, and define  $Y^{(n)}(\cdot) = \{Y^{(n)}(t), t \geq 0\}$  by  $Y^{(n)}(t) = \eta(X^{(n)}(\lfloor n^2 t \rfloor))$ , where  $\eta(x^{(n)}) = (z_1, \dots, z_n, 0, 0, \dots)$  if  $z_i$  is the relative size of the  $i$ th largest cluster in  $x^{(n)}$ . Then*

$$Y^{(n)}(\cdot) \xrightarrow{D} Z(\cdot) \quad \text{in } C_{\overline{\mathbb{V}}_\infty}([0, \infty))$$

as  $n \rightarrow \infty$ .

Hence, the Markov chain  $\{X^{(n)}(m), m \in \mathbb{N}_0\}$ , once appropriately transformed and rescaled, provides a Moran-type particle construction of the two-parameter model. Denote by  $\{K_n(m), m \in \mathbb{N}_0\}$  the chain which keeps track of the number of distinct types in  $X^{(n)}(m)$ , and let  $M_{1,n}$  denote the number of types in  $X^{(n)}(m)$  with only one representative. The transition probabilities for  $K_n(m)$ , denoted for short by

$$p(k, k') = \mathbb{P}\{K_n(m + 1) = k' \mid K_n(m) = k\},$$

are given by

$$p(k, k') = \begin{cases} \left(1 - \frac{M_{1,n}}{n}\right) \frac{\theta + \alpha k}{\theta + n - 1} & \text{if } k' = k + 1, \\ \frac{M_{1,n}}{n(\theta + n - 1)} (n - 1 - \alpha(k - 1)) & \text{if } k' = k - 1, \\ 1 - p(k, k + 1) - p(k, k - 1) & \text{if } k' = k, \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

for  $1 \leq k \leq n$ . Here  $M_{1,n}/n$  is the probability of removing a cluster of size 1, and  $k = 1$  and  $k = n$  imply that  $p(1, 0) = 0$  and  $p(n, n + 1) = 0$ , respectively. Since (7) need not be

Markovian, we use an approximation of  $M_{1,n}$  based on the following asymptotic result. From (5) and Lemma 3.11 of Pitman (2006), the number  $M_{1,n}$  of clusters of size 1 observed in the sample is such that

$$\frac{M_{1,n}}{n^\alpha} \rightarrow \alpha S \quad \text{a.s.,}$$

so that  $M_{1,n} \approx \alpha k$ . This yields

$$p(k, k') = \begin{cases} \left(1 - \frac{\alpha k}{n}\right) \frac{\theta + \alpha k}{\theta + n - 1} + o(n^{-1+\alpha}) & \text{if } k' = k + 1, \\ \frac{\alpha k}{n(\theta + n - 1)}(n - 1 - \alpha(k - 1)) + o(n^{-1+\alpha}) & \text{if } k' = k - 1, \\ 1 - p(k, k + 1) - p(k, k - 1) & \text{if } k' = k, \\ 0 & \text{otherwise.} \end{cases} \quad (8)$$

The following theorem identifies the  $\alpha$ -diversity diffusion for the two-parameter model. Denote by  $C_0([0, \infty))$  the space of continuous functions on  $[0, \infty)$  vanishing at  $\infty$ .

**Theorem 1.** *Let  $\{K_n(m), m \in \mathbb{N}_0\}$  be a Markov chain on  $\mathbb{N}$  with transition probabilities as in (8) for  $0 < \alpha < 1$  and  $\theta > -\alpha$ , and define  $\{\tilde{K}_n(t), t \geq 0\}$  by letting  $\tilde{K}_n(t) = K_n(\lfloor n^{1+\alpha}t \rfloor)/n^\alpha$ . Also, let  $\{S_{\theta,\alpha}(t), t \geq 0\}$  be a diffusion process on  $[0, \infty)$  driven by the stochastic differential equation*

$$dS_{\theta,\alpha}(t) = \theta dt + \sqrt{2\alpha S_{\theta,\alpha}(t)} dB(t), \quad (9)$$

where  $B(t)$  is a standard Brownian motion. If  $\tilde{K}_n(0) \xrightarrow{D} S_{\theta,\alpha}(0)$  then

$$\{\tilde{K}_n(t), t \geq 0\} \xrightarrow{D} \{S_{\theta,\alpha}(t), t \geq 0\} \quad \text{in } C_{[0,\infty)}([0, \infty))$$

as  $n \rightarrow \infty$ .

*Proof.* Denote by  $U_n$  the semigroup induced by (8). For notational brevity, here we do not distinguish between  $(n, k)$  and  $(n - 1, k - 1)$ , since they are asymptotically equivalent. Then, for  $f \in C_0([0, \infty))$ , we can write

$$(U_n - I)f\left(\frac{k}{n^\alpha}\right) = \left[f\left(\frac{k+1}{n^\alpha}\right) - f\left(\frac{k}{n^\alpha}\right)\right] \left(1 - \frac{\alpha k}{n}\right) \frac{\theta + \alpha k}{\theta + n} + \left[f\left(\frac{k-1}{n^\alpha}\right) - f\left(\frac{k}{n^\alpha}\right)\right] \frac{\alpha k(n - \alpha k)}{n(\theta + n)} + o(n^{-1+\alpha}).$$

By means of a Taylor expansion we obtain

$$(U_n - I)f\left(\frac{k}{n^\alpha}\right) = \frac{1}{n^\alpha} C_{\theta,\alpha,k,n}^{(1)} f'\left(\frac{k}{n^\alpha}\right) + \frac{1}{2n^{2\alpha}} C_{\theta,\alpha,k,n}^{(2)} f''\left(\frac{k}{n^\alpha}\right) + o(n^{-1+\alpha}),$$

where

$$C_{\theta,\alpha,k,n}^{(1)} = \left(1 - \frac{\alpha k}{n}\right) \frac{\theta + \alpha k}{\theta + n} - \frac{\alpha k}{n} \left(\frac{n - \alpha k}{\theta + n}\right) = \frac{\theta}{\theta + n} + o(n^{-1})$$

and

$$C_{\theta,\alpha,k,n}^{(2)} = \left(1 - \frac{\alpha k}{n}\right) \frac{\theta + \alpha k}{\theta + n} + \frac{\alpha k}{n} \left(\frac{n - \alpha k}{\theta + n}\right) = \frac{2\alpha k}{\theta + n} + o(n^{-1+\alpha}).$$

Using (5), it follows that

$$\sup_{s \in [0, \infty)} |\mathcal{L}f(s) - n^{1+\alpha}(U_n - I)f(s)| \rightarrow 0 \quad \text{as } n \rightarrow \infty, \tag{10}$$

where  $\mathcal{L}f(s) = \theta f'(s) + \alpha s f''(s)$  is the infinitesimal operator corresponding to (9). Equation (10) holds for every  $f$  belonging to an appropriate restriction  $\mathcal{D}(\mathcal{L})$  of  $C_0([0, \infty))$  (to be formalized in Proposition 2 below). Under these conditions, Theorem 1.6.5 of Ethier and Kurtz (1986) implies that

$$\sup_{s \in [0, \infty)} |U(t)f(s) - U_n(\lfloor n^{1+\alpha}t \rfloor)f(s)| \rightarrow 0, \quad f \in C_0([0, \infty)), \tag{11}$$

as  $n \rightarrow \infty$  and for all  $t \geq 0$ , where  $U$  is the semigroup operator corresponding to  $\mathcal{L}$ . The assertion of the theorem with  $C_{[0, \infty)}([0, \infty))$  replaced by  $D_{[0, \infty)}([0, \infty))$  now follows from (11) and Theorem 4.2.6 of Ethier and Kurtz (1986). Finally, the convergence holds in  $C_{[0, \infty)}([0, \infty))$  since the limit probability measure puts mass 1 on  $C_{[0, \infty)}([0, \infty))$ , and the Skorokhod topology relative to  $C_{[0, \infty)}([0, \infty))$  coincides with the uniform topology of  $C_{[0, \infty)}([0, \infty))$  (cf. Billingsley (1986, Section 18)). This completes the proof.

Hence, the dynamic heterogeneity of the two-parameter model is described by a nonnegative diffusion obtained with a space–time rescaling which depends on the parameter  $\alpha$ . Note that  $S_{\theta, \alpha}(\cdot)$  in (9) can be seen as a critical continuous-state branching process with immigration (see Kawazu and Watanabe (1971) and Li (2006)), obtained, for example, as the diffusion approximation of a Galton–Watson branching process with immigration, with a unitary mean number of offspring per individual. Here  $\theta > 0$  is interpreted as the immigration rate; the  $\theta < 0$  case has been treated in Göing-Jaeschke and Yor (2003).

The next proposition, which provides the complete boundary behavior of the process driven by (9) and formalizes its well-defined nature, is not new and is included for formal completeness. Let  $\mathcal{L}$  be the second-order differential operator

$$\mathcal{L} = \theta \frac{d}{ds} + \alpha s \frac{d^2}{ds^2}, \quad 0 < \alpha < 1, \theta > -\alpha. \tag{12}$$

**Proposition 2.** *The process  $\{S_{\theta, \alpha}(t), t \geq 0\}$  driven by (9) has the following boundary behavior: the boundary  $s = 0$  is absorbing for  $\theta \leq 0$ , instantaneously reflecting for  $0 < \theta < \alpha$ , and entrance for  $\theta \geq \alpha$ ; the boundary  $s = \infty$  is natural and nonattracting for  $\theta \leq \alpha$ , and natural and attracting for  $\theta > \alpha$ . Moreover,  $S_{\theta, \alpha}(t)$  is null recurrent for  $\theta = \alpha$  and transient for  $\theta \neq \alpha$ . For  $\mathcal{L}$  as in (12), define*

$$\mathcal{D}(\mathcal{L}) = \{f \in C_0([0, \infty)) \cap C^2((0, \infty)): \mathcal{L}f \in C_0([0, \infty))\}$$

and

$$\mathcal{D}_{\theta, \alpha}(\mathcal{L}) = \begin{cases} f \in \mathcal{D}(\mathcal{L}) & \text{if } \theta \geq \alpha, \\ f \in \mathcal{D}(\mathcal{L}): \lim_{x \rightarrow 0} x^{\theta/\alpha} f'(x) = 0 & \text{if } 0 < \theta < \alpha, \\ f \in \mathcal{D}(\mathcal{L}): \lim_{x \rightarrow 0} \mathcal{L}f(x) = 0 & \text{if } -\alpha < \theta \leq 0. \end{cases}$$

Then  $\{(f, \mathcal{L}f): f \in \mathcal{D}_{\theta, \alpha}(\mathcal{L})\}$  generates a Feller semigroup on  $C_0([0, \infty))$ .

*Proof.* The first assertion follows from Ikeda and Watanabe (1989, Example IV.8.2) and Karlin and Taylor (1981, Table 15.6.2). The second assertion follows from the first assertion and

Ikeda and Watanabe (1989, Theorem VI.3.1). The third assertion follows from the first assertion, together with the fact that  $\exp\{\int_1^x (\theta/\alpha y) dy\} = x^{\theta/\alpha}$  and Theorem 8.1.1 and Corollary 8.1.2 of Ethier and Kurtz (1986). This completes the proof.

The lack of positive recurrence immediately determines the nonstationarity of the process  $\{S_{\theta,\alpha}(t), t \geq 0\}$ .

We conclude this section with a brief discussion of the corresponding process for the one-parameter model. Although the notion of  $\alpha$ -diversity is given for Poisson–Kingman models with  $0 < \alpha < 1$  (cf. Pitman (2003)), a result analogous to Theorem 1 can be derived for the one-parameter case, for  $\alpha = 0$ . The limit corresponding to (5) when  $\alpha = 0$ , given in Korwar and Hollander (1973), is

$$\lim_{n \rightarrow \infty} \frac{K_n}{\log n} = \theta \quad \text{a.s.}$$

Hence, we expect the process for the normalized number of species to converge to a constant process, i.e.

$$\left\{ \frac{K_n(\lfloor c_n t \rfloor)}{\log n}, t \geq 0 \right\} \xrightarrow{\mathbb{P}} S_{\theta,0}(t) \equiv \theta$$

for some  $c_n \rightarrow \infty$ . Setting  $\alpha = 0$  in (7) and proceeding similarly as in the proof of Theorem 1, we obtain

$$\begin{aligned} (U_n - I)f\left(\frac{k}{\log n}\right) &= \frac{1}{\log n} f'\left(\frac{k}{\log n}\right) \left[ \left(1 - \frac{w}{n}\right) \frac{\theta}{\theta + n} - \frac{w}{n} \left(\frac{n}{\theta + n}\right) \right] \\ &\quad + \frac{1}{2 \log^2 n} f''\left(\frac{k}{\log n}\right) \left[ \left(1 - \frac{w}{n}\right) \frac{\theta}{\theta + n} + \frac{w}{n} \left(\frac{n}{\theta + n}\right) \right] \\ &\quad + o((n \log n)^{-1}), \end{aligned}$$

where  $w$  stands for the fact that

$$M_{1,n} \xrightarrow{D} W \sim \text{Poisson}(\theta);$$

see Arratia *et al.* (1992). Hence, in the limit the argument of the derivatives is constant, and  $c_n(U_n - I)f(k/\log n)$ , with  $c_n = n \log n$ , converges to 0. It follows that the dynamics of the number of species underlying the infinitely-many-alleles model are driven by the constant process  $S_{\theta,0}(t) \equiv \theta$  in the one-parameter case, and by the diffusion process  $S_{\theta,\alpha}(t)$  on  $[0, \infty)$  with state-dependent volatility in the two-parameter case. The structural difference between the one- and two-parameter models is therefore also confirmed from this dynamic viewpoint. A similar difference between the two cases will be found in Section 3.2 using a different approach.

### 3. Finite-dimensional construction of the two-parameter model

#### 3.1. Preliminary remarks

In the introduction it was mentioned that two different sequential constructions of the two-parameter model have been provided in Petrov (2009) and Ruggiero and Walker (2009). In this section we briefly outline why these offer only partial insight into the dynamics underlying the two-parameter model from a biological perspective, motivating the need for further investigation.

The abovementioned constructions are respectively given by a sequence taking values in the space of partitions of  $\mathbb{N}$  and by the Moran-type particle representation outlined in Section 2.

Both cases are based on a dynamic system of finitely many exchangeable particles and exhibit right-continuous sample paths, whereas (3) (with an appropriate domain) characterizes an  $n$ -dimensional diffusion process. Another notable feature of these constructions is the assumption that the distribution that generates the mutant types is nonatomic and, thus, selects types which appear for the first time with probability 1 (in the framework of Petrov (2009) this amounts to saying that a new box is occupied with probability 1). In particular, such a feature turns out to be the key for proving the weak convergence of the sequences to the two-parameter model (see, e.g. Ruggiero and Walker (2009) after Remark 3.1). Such an assumption of nonatomicity cannot be applied in a construction similar to (3)–(4) because the mass of the distribution must concentrate on the enumerated types, in order to keep the maximum amount of species constant in time. To be more precise, note first that the drift coefficients in the Wright–Fisher operator (3) are determined as

$$b_i^{(n)}(z) = \sum_{j \neq i} q_{ji}^{(n)}(z)z_j - \sum_{j \neq i} q_{ij}^{(n)}(z)z_i. \tag{13}$$

Here  $q_{ij}$  is the intensity of a mutation from type  $i$  to type  $j$ , and diagonal elements are  $q_{ii} = -\sum_{j \neq i} q_{ij}$ , so that  $(q_{ij})_{i,j=1,\dots,n}$  is a square matrix with nonnegative off-diagonal elements and row sums equal to 0. In general, the mutation rate  $q_{ij}^{(n)}(z)$  can be thought of as state dependent, but in many interesting cases only the dependence on  $n$  is needed. The drift (4) for example is obtained by taking parent-independent symmetric mutations with rates

$$q_{ij}^{(n)} = \frac{\theta}{n-1}, \quad i \neq j, \tag{14}$$

whereby, when a type  $i$  mutates, the new type will be any of the other  $n - 1$  types with equal chances, and  $\theta$  controls how often, on average, mutations occur. In this case the mutant type is chosen with uniform probability, and the mutant type distribution is discretely supported. Such a derivation of the one-parameter model can be extended to have nonsymmetric mutation (see, for example, Ethier and Kurtz (1981, Theorem 3.4)), but the difference is not relevant for our purposes.

The two existing constructions for the two-parameter model therefore feature finitely many objects, potentially of infinitely many types, and a diffuse mutant type distribution, while the desired construction should feature infinitely many objects of finitely many types and a discretely supported mutant type distribution.

From a mathematical point of view, we ideally seek mutation rates  $q_{ij}^{(n)}(z)$  yielding, through (13), the  $i$ th component limit drift term

$$b_i^{(n)}(z) \longrightarrow -\theta z_i - \alpha, \tag{15}$$

where the convergence is uniform, and the  $b_i^{(n)}(z)$ s satisfy the boundary conditions

$$b_i^{(n)}(z) \geq 0 \quad \text{if } z_i = 0, \quad b_i^{(n)}(z) \leq 0 \quad \text{if } z_i = 1, \tag{16}$$

for  $z \in \Delta_n$  and

$$\Delta_n = \left\{ z \in [0, 1]^n : z_i \geq 0, \sum_{i=1}^n z_i = 1 \right\}. \tag{17}$$

However, obtaining (15) and (16) jointly is clearly not possible, since, when  $z_i = 0$ , the drift should be nonnegative for all  $n$ , but strictly negative in the limit. Since condition (16) is crucial

for the well-definedness of the  $n$ th term of the sequence, an alternative strategy is to relax (15) to the weaker condition

$$\sum_{i=1}^n b_i^{(n)}(z) \frac{\partial f(z)}{\partial z_i} \rightarrow - \sum_{i=1}^{\infty} (\theta z_i + \alpha) \frac{\partial f(z)}{\partial z_i} \tag{18}$$

uniformly as  $n \rightarrow \infty$  for a sufficiently large set of functions  $f(z)$ . Obtaining rates  $q_{ij}^{(n)}(z)$  which yield drift terms satisfying (18), together with some additional restrictions concerning the volatility and the state space of the process, will then suffice to provide the desired convergence.

### 3.2. Sequential construction

Let  $n \geq 2$  throughout the section, and let  $\Delta_n$  be as in (17). Consider a sequence of real numbers  $\{\varepsilon_n\}_{n \in \mathbb{N}}$  satisfying

$$0 < \varepsilon_n < \frac{1}{n} \text{ for all } n, \quad \varepsilon_n = o(n^{-1}), \tag{19}$$

and define the compact subspace of  $\Delta_n$  by

$$\Delta_{n,\varepsilon_n} = \left\{ z \in [0, 1]^n : z_i \geq \varepsilon_n, \sum_{i=1}^n z_i = 1 \right\},$$

where  $z \in \Delta_{n,\varepsilon_n}$  implies that  $z_i \in [\varepsilon_n, 1 - (n - 1)\varepsilon_n] \neq \emptyset$  for all  $i$ . (See Figure 2 below.) Consider the second-order differential operator

$$\mathcal{A}_n = \frac{1}{2} \sum_{i,j=1}^n a_{ij}^{(n)}(z) \frac{\partial^2}{\partial z_i \partial z_j} + \frac{1}{2} \sum_{i=1}^n b_i^{(n)}(z) \frac{\partial}{\partial z_i} \tag{20}$$

with domain

$$\mathcal{D}(\mathcal{A}_n) = \{f : f \in C^2(\Delta_{n,\varepsilon_n})\}, \tag{21}$$

where

$$C^2(\Delta_{n,\varepsilon_n}) = \{f \in C(\Delta_{n,\varepsilon_n}) : \text{there exists } \tilde{f} \in C^2(\mathbb{R}^n), \tilde{f}|_{\Delta_{n,\varepsilon_n}} = f\}.$$

The covariance components in (20) are specified to be

$$\begin{aligned} a_{ij}^{(n)}(z) &= (z_i - \varepsilon_n)(\delta_{ij}(1 - n\varepsilon_n) - (z_j - \varepsilon_n)) \\ &= \begin{cases} (z_i - \varepsilon_n)(1 - (n - 1)\varepsilon_n - z_i) & \text{if } i = j, \\ -(z_i - \varepsilon_n)(z_j - \varepsilon_n) & \text{if } i \neq j. \end{cases} \end{aligned} \tag{22}$$

These can be seen as Wright–Fisher-type covariance terms restricted to  $\Delta_{n,\varepsilon_n}$ , since they vanish at  $z_i = \varepsilon_n$  and  $z_i = 1 - (n - 1)\varepsilon_n$ . Additionally, consider the state-dependent mutation rates

$$q_{ij}^{(n)}(z) = \frac{\theta}{n - 1} + \frac{2\alpha j}{z_i n(n + 1)} \left[ 1 - \exp\left\{ \frac{-2(z_i - \varepsilon_n)}{\varepsilon_n} \right\} \right], \quad i \neq j. \tag{23}$$

Before providing some considerations concerning the form of  $q_{ij}^{(n)}(z)$ , note that (13) yields the drift components

$$\begin{aligned} b_i^{(n)}(z) &= \frac{\theta}{n - 1} (1 - z_i) - \theta z_i + \frac{2\alpha i}{n(n + 1)} \sum_{j=1}^n \left[ 1 - \exp\left\{ \frac{-2(z_j - \varepsilon_n)}{\varepsilon_n} \right\} \right] \\ &\quad - \alpha \left[ 1 - \exp\left\{ \frac{-2(z_i - \varepsilon_n)}{\varepsilon_n} \right\} \right]. \end{aligned} \tag{24}$$

Here the first two terms of  $b_i^{(n)}(z)$  equal

$$\frac{\theta}{n-1}(1 - (n-1)\varepsilon_n - z_i) - \theta(z_i - \varepsilon_n),$$

and  $z_i = 1 - (n-1)\varepsilon_n$  implies that  $z_j = \varepsilon_n$  for all  $j \neq i$ . Using the last two observations in (24) shows that

$$b_i^{(n)}(z) > 0 \quad \text{if } z_i = \varepsilon_n, \quad b_i^{(n)}(z) < 0 \quad \text{if } z_i = 1 - (n-1)\varepsilon_n, \tag{25}$$

so that  $b_i^{(n)}(z)$  satisfies (16) restricted to  $\Delta_{n,\varepsilon_n}$ .

In order to provide some intuition about (23), we have to separately consider the constant and frequency-dependent terms. The former term attributes equal chances of mutation to all species regardless of their abundance, as in the one-parameter model. To evaluate the effect of  $q_{ij}^{(n)}(z)$  as a deviation from (14), recall that the limit operator  $\mathcal{B}$  (see (2)) acts on functions defined on  $\bar{\mathcal{V}}_\infty$  (see (1)), where the frequencies have been ordered. The same ordering operation will be done before taking the limit of  $\mathcal{A}_n$ , with the formal appearance of the operator unchanged, so that it is correct to think in terms of ranked frequencies. In light of this, the term  $j$  in the numerator of the second term in  $q_{ij}^{(n)}(z)$  can be interpreted as an approximate indication of the size of the frequency  $z_j$ , where a larger  $j$  value implies a lower  $z_j$ . Hence, mutations from  $i$  to  $j$  occur more frequently if  $z_j$  is relatively low, implying a redistributive effect. This has to be interpreted as a conditional mechanism, related to the probability of directing the mass of a type- $i$  individual to some species  $j \neq i$ , conditional on the fact that such an individual mutates. In order to evaluate the unconditional chances of mutation of type- $i$  individuals, consider the rescaled state-dependent term in  $q_{ij}^{(n)}(z)$ , namely,

$$z_i^{-1} \left[ 1 - \exp \left\{ \frac{-2(z_i - \varepsilon_n)}{\varepsilon_n} \right\} \right]. \tag{26}$$

Recall that the range of values of  $z_i$  is determined by  $n$  through  $\varepsilon_n$  and grows to  $[0, 1]$  as  $n$  diverges, and note that the clear nonmonotonicity of the quantity in (26) is displayed on such a range only for large enough  $n$ . Figure 1 provides a qualitative comparison of (26) as a function of  $z_i$  for  $n_1 < n_2 < n_3$  and  $\varepsilon_n = n^{-1.1}$ , so that (19) holds. The plot highlights the contribution of the rescaled state-dependent term of  $q_{ij}^{(n)}(z)$  with respect to the one-parameter mutation rate (14). The behavior of  $q_{ij}^{(n)}(z)$  for  $z_i$  relatively far from  $\varepsilon_n$  can be interpreted in terms of a reinforcement mechanism similar to that featured by the  $PD(\theta, \alpha)$  distribution (see Lijoi *et al.* (2007)). It can indeed be observed that the probability that a further sample from (6) is an already observed species is not allocated proportionally to the current frequencies. The ratio of probabilities assigned by (6) to any pair of species  $(i, j)$  is  $r_{i,j} = (n_i - \alpha)/(n_j - \alpha)$ . When  $\alpha \rightarrow 0$ , the probability of sampling species  $i$  is proportional to the absolute frequency  $n_i$ , or, equivalently, to  $z_i$ , which in continuous time is reflected by a constant mutation rate as in (14). However, since  $r_{i,j}$  is increasing in  $\alpha$ , a value of  $\alpha > 0$  reallocates some probability mass from type  $j$  to type  $i$ , so that, for example, for  $n_i = 2$  and  $n_j = 1$ , we have  $r_{i,j} = 2, 3, 5$  for  $\alpha = 0, 0.5, 0.75$ , respectively. Thus,  $\alpha$  has a reinforcement effect on those species that have a higher frequency. On the other hand, the behavior of  $q_{ij}^{(n)}(z)$  for  $z_i$  near the boundary is what ultimately makes the process well defined in a bounded region. For  $z_i \downarrow \varepsilon_n$ , (23) converges to the rate of the one-parameter model (14), so that the associated drift behaves locally as for the one-parameter case and  $z_i$  is kept inside the boundary  $\varepsilon_n > 0$ . Roughly speaking, the decreasing part of the rate function pushes the frequencies towards smaller values, whereas the leftmost plotted part is responsible for keeping the frequencies inside the state space.

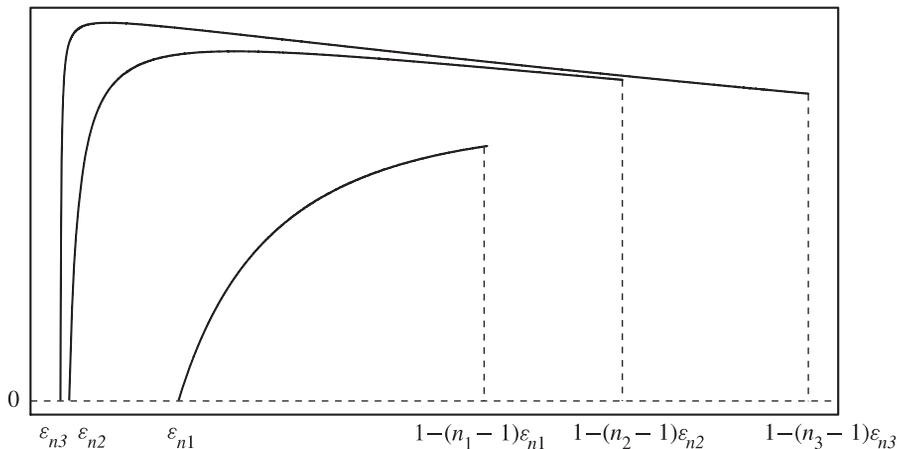


FIGURE 1: Contribution of the (rescaled) state-dependent part of the mutation rate  $q_{ij}^{(n)}(z)$  in (23) with respect to the rate of the one-parameter model. The plot shows a qualitative comparison of (26) as a function of  $z_i$  with  $\epsilon_n = n^{-1.1}$  and  $n_1 < n_2 < n_3$ .

The following result shows that the above-defined operator  $\mathcal{A}_n$  characterizes a Feller diffusion on  $\Delta_{n,\epsilon_n}$ . Let  $\|\cdot\|$  denote the supremum norm.

**Theorem 2.** *Let  $\mathcal{A}_n$  be the operator defined by (20)–(22) and (24). Then the closure of  $\mathcal{A}_n$  in  $C(\Delta_{n,\epsilon_n})$  is single valued and generates a Feller semigroup  $\{\mathcal{T}_n(t)\}$  on  $C(\Delta_{n,\epsilon_n})$ . For each  $\nu_n \in \mathcal{P}(\Delta_{n,\epsilon_n})$ , there exists a strong Markov process  $Z^{(n)}(\cdot) = \{Z^{(n)}(t), t \geq 0\}$  corresponding to  $\{\mathcal{T}_n(t)\}$  with initial distribution  $\nu_n$  and such that*

$$\mathbb{P}\{Z^{(n)}(\cdot) \in C_{\Delta_{n,\epsilon_n}}([0, \infty))\} = 1.$$

*Proof.* It is easily seen that  $\mathcal{A}_n$  satisfies the positive maximum principle on  $\Delta_{n,\epsilon_n}$ , that is, if  $f \in \mathcal{D}(\mathcal{A}_n)$  and  $z_0 \in \Delta_{n,\epsilon_n}$  are such that  $f(z_0) = \|f\| \geq 0$  then  $\mathcal{A}_n f(z_0) \leq 0$ . This is immediate in the interior of  $\Delta_{n,\epsilon_n}$ , while on the boundaries it follows from (25) and the fact that (22) vanishes at every boundary point. Define  $z^\sigma = z_1^{\sigma_1} \cdots z_n^{\sigma_n}$  and  $\sigma - \delta_i = (\sigma_1, \dots, \sigma_i - 1, \dots, \sigma_n)$  for  $\sigma_1, \dots, \sigma_n \in \mathbb{N}$ . Then

$$\begin{aligned} \mathcal{A}_n z^\sigma &= \frac{1}{2} \sum_i \sigma_i \left[ \frac{\theta}{n-1} z^{\sigma-\delta_i} - \frac{\theta n}{n-1} z^\sigma + \frac{2\alpha i}{n(n+1)} \sum_{j=1}^n (1 - C_1 e^{-C_2 z_j}) z^{\sigma-\delta_i} \right. \\ &\quad - \alpha z^{\sigma-\delta_i} + \alpha C_1 e^{-C_2 z_i} z^{\sigma-\delta_i} \\ &\quad + (\sigma_i - 1) ([1 - (n-1)\epsilon_n] z^{\sigma-\delta_i} - z^\sigma \\ &\quad \quad \quad \left. - \epsilon_n [1 - (n-1)\epsilon_n] z^{\sigma-2\delta_i} + \epsilon_n z^{\sigma-\delta_i}) \right] \\ &\quad - \frac{1}{2} \sum_i \sum_{j \neq i} \sigma_i \sigma_j [z^\sigma - \epsilon_n z^{\sigma-\delta_i} - \epsilon_n z^{\sigma-\delta_j} + \epsilon_n^2 z^{\sigma-\delta_i-\delta_j}]. \end{aligned}$$

for appropriate constants  $C_1, C_2$ . Letting  $L_m$  denote the algebra of polynomials in  $(z_1, \dots, z_n)$  restricted to  $\Delta_{n,\epsilon_n}$  with degree not greater than  $m \in \mathbb{N}$ , the image of  $\mathcal{A}_n$  computed on  $L_m$  contains functions belonging to  $L_m$  and of type  $e^{-z_i} z^c$ . Since, for every  $g(x) \in C(K)$ , with

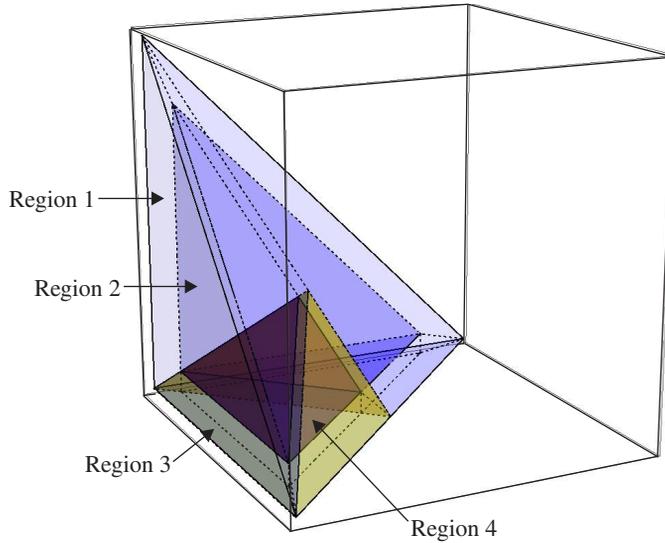


FIGURE 2: Projection onto the first three coordinates of  $\Delta_n$  (region 1),  $\Delta_{n,\varepsilon_n}$  (region 2),  $\nabla_{n,0}$  (region 3), and  $\nabla_{n,\varepsilon_n}$  (region 4) for  $n \geq 3$ . The (ordered)  $n$ th term of the sequence lives in region 4, so that the frequencies are bounded away from 0 for all  $n$ . As  $n$  increases, region 4 converges to region 3, so that the limit process is left free to move in the full ordered simplex.

$K$  compact, and  $f(x) = e^x g(x) \in C(K)$ , there exists a sequence  $\{p^{(k)}\}$  of polynomials on  $K$  such that  $\|f - p^{(k)}\| \rightarrow 0$ , so that  $\|e^{-z} p^{(k)} - g\| \rightarrow 0$ , it follows that the image of  $\mathcal{A}_n$  is dense in  $C(\Delta_{n,\varepsilon_n})$ , and so is that of  $\lambda - \mathcal{A}_n$  for all but at most countably many  $\lambda > 0$ . Since  $\bigcup_m L_m$  is dense in  $C(\Delta_{n,\varepsilon_n})$ , the Hille–Yosida theorem (see Ethier and Kurtz (1986, Theorem 4.2.2)) now implies that the closure of  $\mathcal{A}_n$  on  $C(\Delta_{n,\varepsilon_n})$  is single valued and generates a strongly continuous, positive, contraction semigroup  $\{\mathcal{T}_n(t)\}$  on  $C(\Delta_{n,\varepsilon_n})$ . The fact that  $(1, 0)$  belongs to the domain of  $\overline{\mathcal{A}}_n$  also implies that  $\{\mathcal{T}_n(t)\}$  is conservative. Note that, for every  $z_0 \in \Delta_{n,\varepsilon_n}$  and  $\delta > 0$ , there exists  $f \in \mathcal{D}(\mathcal{A}_n)$  such that

$$\sup_{z \in B^c(z_0, \delta)} f(z) < f(z_0) = \|f\| \quad \text{and} \quad \mathcal{A}_n f(z_0) = 0,$$

where  $B^c(z_0, \delta)$  is a ball of radius  $\delta$  centered at  $z_0$ . Note that we can take, for example,  $f(z) = -C_\delta \sum_{i=1}^n (z_i - z_0)^4$  for an appropriate constant  $C_\delta$  which depends on  $\delta$ . Then the second assertion follows from Theorem 4.2.7 and Remark 4.2.10 of Ethier and Kurtz (1986). This completes the proof.

Given  $\overline{\nabla}_\infty$  as in (1), define the subspaces

$$\nabla_\infty = \left\{ z \in \overline{\nabla}_\infty : \sum_{i=1}^\infty z_i = 1 \right\} \tag{27}$$

and

$$\nabla_{n,\varepsilon_n} = \{z \in \nabla_\infty : z_n \geq \varepsilon_n > z_{n+1} = 0\}.$$

See Figure 2. Define also the Borel measurable map  $\rho_n : \Delta_n \rightarrow \nabla_\infty$  by

$$\rho_n(z) = (z_{(1)}, \dots, z_{(n)}, 0, 0, \dots), \quad z \in \Delta_n, \tag{28}$$

where  $z_{(i)}$  are the decreasing order statistics of  $z \in \Delta_n$ . It is clear that  $\rho_n$  maps  $\Delta_{n,\varepsilon_n}$  into  $\nabla_{n,\varepsilon_n}$ . If  $Z^{(n)}(\cdot)$  is the Markov process of Theorem 2, our aim is thus to show that

$$\rho_n(Z^{(n)}(\cdot)) \xrightarrow{D} Z(\cdot) \tag{29}$$

in the sense of convergence in distribution in  $C_{\overline{\nabla}_\infty}([0, \infty))$  as  $n \rightarrow \infty$ , where  $Z(\cdot)$  is the diffusion process corresponding to operator (2) with an appropriate domain. To this end, consider the symmetric polynomials

$$\varphi_m(z) = \sum_{i \geq 1} z_i^m, \quad z \in \overline{\nabla}_\infty, \quad m \geq 2, \tag{30}$$

and define

$$\mathcal{D}(\mathcal{B}) = \{\text{subalgebra of } C(\overline{\nabla}_\infty) \text{ generated by } 1, \varphi_3(z), \varphi_4(z), \dots\}. \tag{31}$$

**Lemma 1.**  $\mathcal{D}(\mathcal{B})$  is dense in  $C(\overline{\nabla}_\infty)$ .

*Proof.* In Ethier and Kurtz (1981) (see proof of Theorem 2.5) it was proved that the closure of  $\mathcal{D}_0(\mathcal{B})$ , defined as

$$\mathcal{D}_0(\mathcal{B}) = \{\text{subalgebra of } C(\overline{\nabla}_\infty) \text{ generated by } 1, \varphi_2(z), \varphi_3(z), \dots\}, \tag{32}$$

equals  $C(\overline{\nabla}_\infty)$ . Note that

$$z_1 = \lim_{m \rightarrow \infty} \varphi_m(z)^{1/m}, \quad z_2 = \lim_{m \rightarrow \infty} (\varphi_m(z) - z_1^m)^{1/m},$$

etc., from which

$$\varphi_2(z) = \lim_{m \rightarrow \infty} \varphi_m(z)^{2/m} + (\varphi_m(z) - z_1^m)^{2/m} + \dots$$

so that  $\varphi_2 \in \overline{\mathcal{D}(\mathcal{B})}$ . It follows that  $\overline{\mathcal{D}_0(\mathcal{B})} \equiv \overline{\mathcal{D}(\mathcal{B})}$ , from which the result follows.

Before providing the convergence argument, we recall the relevant theorems about the formal existence and the sample path properties of the process  $Z(\cdot)$  appearing in (29).

**Theorem 3.** (Petrov (2009).) *Let  $\mathcal{B}$  be operator (2) with domain (32). The closure of  $\mathcal{B}$  in  $C(\overline{\nabla}_\infty)$  generates a Feller semigroup  $\{\mathcal{T}(t)\}$  on  $C(\overline{\nabla}_\infty)$ , and, for each  $\nu \in \mathcal{P}(\overline{\nabla}_\infty)$ , there exists a strong Markov process  $Z(\cdot) = \{Z(t), t \geq 0\}$  corresponding to  $\{\mathcal{T}(t)\}$  with initial distribution  $\nu$  and such that*

$$\mathbb{P}\{Z(\cdot) \in C_{\overline{\nabla}_\infty}([0, \infty))\} = 1.$$

Let  $\nabla_\infty$  be as in (27). The following result shows that if the initial distribution of the Markov process  $Z(\cdot)$  of Theorem 3 is  $\text{PD}(\theta, \alpha)$  then the law of the process is concentrated on  $C_{\nabla_\infty}([0, \infty))$ .

**Theorem 4.** (Feng and Sun (2010).) *Let  $Z(\cdot) = \{Z(t), t \geq 0\}$  be the Markov process of Theorem 3, and assume that  $Z(0) \sim \text{PD}(\theta, \alpha)$ . Then*

$$\mathbb{P}\{Z(t) \in \nabla_\infty \text{ for all } t \geq 0\} = 1.$$

Denote by  $\mathcal{B}_n$  the right-hand side of (20), with coefficients as in (22) and (24), and domain

$$\mathcal{D}(\mathcal{B}_n) = \{f \in C(\nabla_{n,\varepsilon_n}) : f \circ \rho_n \in C^2(\Delta_{n,\varepsilon_n})\}.$$

We are now ready to state the main result of the section.

**Theorem 5.** *Let  $Z^{(n)}(\cdot)$  and  $Z(\cdot)$  be the Markov processes of Theorem 2 and Theorem 3, with initial distributions  $\nu_n \in \mathcal{P}(\Delta_{n,\varepsilon_n})$  and  $\nu \in \mathcal{P}(\bar{\nabla}_\infty)$ , respectively. If  $\nu_n \circ \rho_n^{-1} \xrightarrow{D} \nu$  then (29) holds in  $C_{\bar{\nabla}_\infty}([0, \infty))$ . If, in addition,  $\nu$  is PD( $\theta, \alpha$ ) then (29) holds in  $C_{\nabla_\infty}([0, \infty))$ .*

*Proof.* For  $\rho_n$  as in (28), define  $\pi_n : C(\bar{\nabla}_\infty) \rightarrow C(\Delta_n)$  by  $\pi_n f = f \circ \rho_n$  and note that  $\pi_n : \mathcal{D}(\mathcal{B}) \rightarrow \mathcal{D}(\mathcal{A}_n)$ . Since, for every  $f \in \mathcal{D}(\mathcal{B})$ , we have  $\pi_n \mathcal{B}f = \mathcal{B}(f \circ \rho_n)$ , for all such functions and  $z \in \Delta_{n,\varepsilon_n}$ , we have

$$\begin{aligned} &\mathcal{A}_n \pi_n f(z) - \pi_n \mathcal{B}f(z) \\ &= \frac{1}{2} \sum_{i,j=1}^n [a_{ij}^{(n)}(z) - z_i(\delta_{ij} - z_j)] \frac{\partial^2 f(\rho_n(z))}{\partial z_i \partial z_j} + \sum_{i=1}^n [b_i^{(n)}(z) + \theta z_i + \alpha] \frac{\partial f(\rho_n(z))}{\partial z_i}. \end{aligned}$$

It can be easily verified that the absolute value of the first term in square brackets is bounded above by a term of order  $O(n\varepsilon_n)$ , and that

$$\begin{aligned} &|\mathcal{A}_n \pi_n f(z) - \pi_n \mathcal{B}f(z)| \\ &\leq O(n\varepsilon_n) \sum_{i,j=1}^n \left| \frac{\partial^2 f(\rho_n(z))}{\partial z_i \partial z_j} \right| + O(n^{-1}) \sum_{i=1}^n \left| \frac{\partial f(\rho_n(z))}{\partial z_i} \right| \\ &\quad + O(n^{-1}) \sum_{i=1}^n i \left| \frac{\partial f(\rho_n(z))}{\partial z_i} \right| + \alpha \sum_{i=1}^n \exp\left\{ \frac{-2(z_i - \varepsilon_n)}{\varepsilon_n} \right\} \left| \frac{\partial f(\rho_n(z))}{\partial z_i} \right|. \end{aligned} \tag{33}$$

Observe now that, for  $f \in \mathcal{D}(\mathcal{B})$  of type  $\varphi_{m_1} \times \dots \times \varphi_{m_k}$ , we have  $f(\rho_n(z)) = f(z)$  and

$$\sum_{i=1}^n \left| \frac{\partial f(z)}{\partial z_i} \right| = \sum_{i=1}^n \sum_{j=1}^k m_j z_i^{m_j-1} \prod_{h \neq j} \varphi_{m_h} \leq \sum_{j=1}^k m_j \sum_{i=1}^n z_i^{m_j-1}, \tag{34}$$

which is bounded above by  $\sum_{j=1}^k m_j < \infty$ . Let  $m_j = 3$  (cf. (31)). Then (34) implies that

$$\sum_{i=1}^n \exp\left\{ \frac{-2(z_i - \varepsilon_n)}{\varepsilon_n} \right\} \left| \frac{\partial f(z)}{\partial z_i} \right| \leq n\varepsilon_n^2 \sum_{j=1}^k m_j \rightarrow 0$$

uniformly as  $n \rightarrow \infty$ , where we have used (19) and the fact that, for  $f$  as above, the left-hand side is maximized when  $z_i = \varepsilon_n$ . Furthermore,

$$O(n^{-1}) \sum_{i=1}^n i \left| \frac{\partial f(z)}{\partial z_i} \right| \leq O(n^{-1}) \sum_{j=1}^k m_j \sum_{i=1}^n i z_i^{m_j-1} \rightarrow 0$$

for  $m_j \geq 3$  since  $z_i \leq i^{-1}$ . Finally,

$$\begin{aligned} \sum_{i,j=1}^n \left| \frac{\partial^2 f(z)}{\partial z_i \partial z_j} \right| &\leq \sum_{i,j=1}^{\infty} \left[ \partial_{ij} \varphi_{m_h} \prod_{\ell \neq h} \varphi_{m_\ell} + \sum_{q \neq h} \partial_i \varphi_{m_h} \partial_j \varphi_{m_q} \prod_{\ell \neq h,q} \varphi_{m_\ell} \right] \\ &= \left[ m_h(m_h - 1) \varphi_{m_h-2} \prod_{\ell \neq h} \varphi_{m_\ell} + \sum_{q \neq h} m_h m_q \varphi_{m_h+m_q-2} \prod_{\ell \neq h,q} \varphi_{m_\ell} \right. \\ &\quad \left. + \sum_{q \neq h} m_h m_q \varphi_{m_h-1} \varphi_{m_q-1} \prod_{\ell \neq h,q} \varphi_{m_\ell} \right] \\ &\leq \left[ m_h(m_h - 1) + \sum_{q \neq h} m_h m_q + \sum_{q \neq h} m_h m_q \right], \end{aligned}$$

whose right-hand side is bounded. From (33), the above arguments and (19) imply that

$$\| \mathcal{A}_n \pi_n f - \pi_n \mathcal{B} f \| \rightarrow 0, \quad f \in \mathcal{D}(\mathcal{B}).$$

Given that  $\rho_n(Z^{(n)})$  clearly satisfies a compact containment condition (cf. Ethier and Kurtz (1986, Remark 3.7.3)), and that Theorem 3 implies that the closure of  $\mathcal{B}$  in  $C_{\bar{\nabla}_\infty}([0, \infty))$  generates a strongly continuous contraction semigroup, the hypotheses of Corollary 4.8.7 of Ethier and Kurtz (1986) are satisfied and the first assertion of the theorem follows with  $C_{\bar{\nabla}_\infty}([0, \infty))$  replaced by  $D_{\bar{\nabla}_\infty}([0, \infty))$ , the space of càdlàg functions from  $[0, \infty)$  to  $\bar{\nabla}_\infty$ . Furthermore, the convergence also holds in  $C_{\bar{\nabla}_\infty}([0, \infty)) \subset D_{\bar{\nabla}_\infty}([0, \infty))$ , since the limit probability measure is concentrated on  $C_{\bar{\nabla}_\infty}([0, \infty))$  and the Skorokhod topology relative to  $C_{\bar{\nabla}_\infty}([0, \infty))$  coincides with the topology on  $C_{\bar{\nabla}_\infty}([0, \infty))$ . See, for example, Billingsley (1986, Section 18). Finally, the second assertion follows from Theorem 4 and from further relativization to  $C_{\nabla_\infty}([0, \infty))$  of the topology on  $C_{\bar{\nabla}_\infty}([0, \infty))$ . This completes the proof.

### Acknowledgements

This research was supported by the European Research Council (ERC) through grant StG ‘N-BNP’ 306406. The author is grateful to an anonymous referee and to Pierpaolo De Blasi for carefully reading the manuscript and for several useful suggestions which lead to an improvement of the presentation.

### References

AOKI, M. (2008). Thermodynamic limit of macroeconomic or financial models: one- and two-parameter Poisson–Dirichlet models. *J. Econom. Dynam. Control* **32**, 66–84.

ARRATIA, R., BARBOUR, A. D. AND TAVARÈ, S. (1992). Poisson process approximations of the Ewens sampling formula. *Ann. Appl. Prob.* **2**, 519–535.

BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes*. Cambridge University Press.

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. John Wiley, New York.

BORODIN, A. AND OLSHANSKI, G. (2009). Infinite-dimensional diffusions as limits of random walks on partitions. *Prob. Theory Relat. Fields* **144**, 281–318.

ETHIER, S. N. AND KURTZ, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Prob.* **13**, 429–452.

ETHIER, S. N. AND KURTZ, T. G. (1986). *Markov Processes. Characterization and Convergence*. John Wiley, New York.

FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics*. Springer, Heidelberg.

FENG, S. AND SUN, W. (2010). Some diffusion processes associated with two parameter Poisson–Dirichlet distribution and Dirichlet process. *Prob. Theory Relat. Fields* **148**, 501–525.

- FENG, S. AND WANG, F.-Y. (2007). A class of infinite-dimensional diffusion processes with connection to population genetics. *J. Appl. Prob.* **44**, 938–949.
- FENG, S., SUN, W., WANG, F.-Y. AND XU, F. (2011). Functional inequalities for the two-parameter extension of the infinitely-many-neutral-alleles diffusion. *J. Funct. Anal.* **260**, 399–413.
- FERGUSON, T. S. (1973). A Bayesian analysis of some nonparametric problems. *Ann. Statist.* **1**, 209–230.
- GÖING-JAESCHKE, A. AND YOR, M. (2003). A survey and some generalizations of Bessel processes. *Bernoulli* **9**, 313–349.
- IKEDA, N. AND WATANABE, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd edn. North-Holland, Amsterdam.
- KARLIN, S. AND TAYLOR, H. M. (1981). *A Second Course in Stochastic Processes*. Academic Press, New York.
- KAWAZU, K. AND WATANABE, S. (1971). Branching processes with immigration and related limit theorems. *Theory Prob. Appl.* **16**, 36–54.
- KINGMAN, J. F. C. *et al.* (1975). Random discrete distribution. *J. R. Statist. Soc. B* **37**, 1–22.
- KORWAR, R. M. AND HOLLANDER, M. (1973). Contributions to the theory of Dirichlet processes. *Ann. Prob.* **1**, 705–711.
- LI, Z.-H. (2006). Branching processes with immigration and related topics. *Front. Math. China* **1**, 73–97.
- LIJOL, A. AND PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics*, Cambridge University Press, pp. 80–136.
- LIJOL, A., MENA, R. H. AND PRÜNSTER, I. (2007). Controlling the reinforcement in Bayesian non-parametric mixture models. *J. R. Statist. Soc. B* **69**, 715–740.
- PERMAN, M., PITMAN, J. AND YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Prob. Theory Relat. Fields* **92**, 21–39.
- PETROV, L. A. (2009). A two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funct. Anal. Appl.* **43**, 279–296.
- PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Prob. Theory Relat. Fields* **102**, 145–158.
- PITMAN, J. (2003). Poisson–Kingman partitions. In *Statistics and Science: A Festschrift for Terry Speed* (IMS Lecture Notes Monogr. Ser. **40**), Institute of Mathematical Statistics, Beachwood, OH, pp. 1–34.
- PITMAN, J. (2006). *Combinatorial Stochastic Processes* (Lecture Notes Math. **1875**). Springer, Berlin.
- PITMAN, J. AND YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Prob.* **25**, 855–900.
- RUGGIERO, M. AND WALKER, S. G. (2009). Countable representation for infinite dimensional diffusions derived from the two-parameter Poisson–Dirichlet process. *Electron. Commun. Prob.* **14**, 501–517.
- RUGGIERO, M., WALKER, S. G. AND FAVARO, S. (2013). Alpha-diversity processes and normalized inverse-Gaussian diffusions. *Ann. Appl. Prob.* **23**, 386–425.
- TEH, Y. W. AND JORDAN, M. I. (2010). Hierarchical Bayesian nonparametric models with application. In *Bayesian Nonparametrics*, Cambridge University Press, pp. 158–207.
- WATTERSON, G. A. (1976). The stationary distribution of the infinitely-many neutral alleles diffusion model. *J. Appl. Prob.* **13**, 639–651. (Correction: **14** (1977), 897.)