

Inhomogeneous Wright–Fisher construction of two-parameter Poisson–Dirichlet diffusions

PIERPAOLO DE BLASI

University of Torino and Collegio Carlo Alberto

MATTEO RUGGIERO

University of Torino and Collegio Carlo Alberto

DARIO SPANÒ

University of Warwick

April 15, 2014

The recently introduced two-parameter Poisson–Dirichlet diffusion extends the infinitely-many-neutral-alleles model, related to Kingman’s distribution and to Fleming–Viot processes. The role of the additional parameter has been shown to regulate the clustering structure of the population, but is yet to be fully understood in the way it governs the reproductive process. Here we shed some light on these dynamics by providing a finite-population construction, with finitely-many species, of the two-parameter infinite-dimensional diffusion. The construction is obtained in terms of Wright–Fisher chains that feature a classical symmetric mutation mechanism and a frequency-dependent immigration, whose inhomogeneity is investigated in detail. The local immigration dynamics are built upon an underlying array of Bernoulli trials and can be described by means of a *dartboard* experiment and a rank-dependent type distribution. These involve a delicate balance between reinforcement and redistributive effects, among the current species abundances, for the convergence to hold.

Keywords: infinite-alleles model, Poisson–Dirichlet distribution, reinforcement, immigration, Wright–Fisher model.

MSC: 60J60, 60G57, 92D25.

1 Introduction

The goal of this paper is to provide a discrete-time population construction with finite size and finitely many types of the two-parameter Poisson–Dirichlet diffusion, extending an anal-

Address for correspondence: matteo.ruggiero@unito.it

ogous construction for the celebrated infinitely-many-neutral-alleles model provided in [Ethier and Kurtz \(1981\)](#). Introduced by [Petrov \(2009\)](#), and henceforth called the two-parameter model, this diffusion process is bounded into the infinite-dimensional ordered simplex and describes the temporal evolution of the ranked frequencies of species or types, observed at a single gene locus, in a given population of infinite size. An exhaustive review of these and other models for stochastic population dynamics can be found in [Feng \(2010\)](#). Further investigations of the two-parameter model include [Ruggiero and Walker \(2009\)](#), who provide a particle construction; [Feng and Sun \(2010\)](#), who investigate some path properties using Dirichlet forms; [Feng, Sun, Wang and Xu \(2011\)](#), who find the transition density function; [Ruggiero, Walker and Favaro \(2013\)](#), who show that an instance of the two-parameter model arises as a normalized inverse-Gaussian diffusion conditioned on having a fixed environment; [Ruggiero \(2014\)](#), who shows that the clustering structure in the population is driven by a continuous-state branching process with immigration. The two-parameter model is known to be reversible and ergodic with respect to the Poisson–Dirichlet distribution $\text{PD}(\theta, \alpha)$ ([Pitman, 1995](#); [Pitman and Yor, 1997](#)), where $0 < \alpha < 1$ and $\theta > -\alpha$. When $\alpha = 0$, the model reduces to the infinitely-many-neutral-alleles model, henceforth called the one-parameter model, with Poisson–Dirichlet reversible distribution $\text{PD}(\theta) := \text{PD}(\theta, 0)$ ([Kingman, 1975](#)). The two-parameter Poisson–Dirichlet random measure $\text{PD}(\theta, \alpha)$ has found numerous applications in several fields: see for example the monographs by [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; [Feng \(2010\)](#) for population genetics. However, the two-parameter diffusion model is not as well understood as the one-parameter sub-case, which motivates the need for further investigation.

One of the main differences between the $\text{PD}(\theta)$ and $\text{PD}(\theta, \alpha)$ distributions is the fact that the former arises as the weak limit of ranked Dirichlet frequencies ([Kingman, 1975](#)), whereas a similar construction is not available for the two-parameter case. In the temporal framework, the one-parameter model arises, among other constructions, as the weak limit of a sequence of finite-dimensional diffusion processes of Wright–Fisher type, each with Dirichlet stationary distribution, after a reordering of the frequencies. Each of these Wright–Fisher diffusions can in turn be constructed as weak limit of a sequence of Wright–Fisher Markov chains describing a discrete reproductive model with finitely-many individuals ([Ethier and Kurtz, 1981](#)). In contrast, a construction with both finitely-many types and finitely-many individuals for the case $0 < \alpha < 1$, to the best of our knowledge, is yet to be formulated. The importance of finding instances of processes with these features for the two-parameter

model lies in the possibility of unveiling the reproductive mechanisms acting at the level of individuals, thus providing interpretation for the role played by the parameters θ, α in the dynamics of the population’s type frequencies, partially hidden or difficult to interpret in the infinite-dimensional model. In Section 2 we will provide more comments on this point and on the other existing sequential constructions for the two-parameter model.

In this paper we show that the two-parameter model can be constructed as a suitable scaling limit of Wright–Fisher Markov chains describing a population with discrete, non-overlapping generations subject to a certain type of frequency-dependent transition kernel, which admits interpretation as immigration, and to an additional component of neutral, parent-independent mutation. The mutation component is governed by the parameter θ in exactly the same way as in the traditional neutral Wright–Fisher model. The parameter α regulates how often immigration events occur and the immigration kernel depends on the current population frequencies. The inhomogeneity of the immigration mechanism, which is built upon an underlying scheme of Bernoulli trials, is conveniently described in terms of a *dartboard* experiment and a rank-dependent type distribution as follows. At each generation time, before reproduction, any individual has a probability of becoming susceptible, i.e. targeted by an immigrant. If the targeted individual is of type i , this probability is inversely proportional to the frequency z_i in the home population. Whether or not a susceptible i -type individual is actually killed off and replaced by the immigrant in turn depends on the following outcome. A representative of each type from the immigrating population throws an unlabelled dart at a dartboard associated to type i , with the dartboard size proportional to z_i : if no dart hits the target, no transition occurs; if at least one dart hits the target, then the type- i individual dies and yields to an immigrant. The immigrating type chosen for replacement is j with probability $p_j(z)$, which is increasing in the rank of z_j , so that scarcer species have higher probability of being chosen. It is assumed that, at each generation, there is virtually unlimited availability of immigrants of all possible types. The overall impact of the immigration kernel is a mix of reinforcement effect among the frequencies, provided by the flow of individuals dying out, and redistributive effect, provided by the incoming flow. Assuming, in addition, classical multinomial resampling between generations, the associated measure-valued Markov chain is shown to converge to the two-parameter model, after appropriate rescaling and when both the population size and the number of types go to infinity.

The paper is organised as follows. In Section 2 the two-parameter model is recalled. Section 3 provides the construction of the Wright–Fisher Markov chain with N individuals and K species, together with a discussion of the requirements needed on the immigration kernel for the convergence to the infinite-dimensional diffusion to hold. The convergence

itself is provided in two steps. In Section 4 the Wright–Fisher chain is shown to converge as $N \rightarrow \infty$ to a K -dimensional Feller diffusion on the $(K - 1)$ -dimensional simplex. Then, in Section 5, the K -dimensional diffusion is shown to converge, as $K \rightarrow \infty$, to the two-parameter model.

2 The two-parameter model

The two-parameter model was introduced by Petrov (2009). As its one-parameter counterpart, characterized in Ethier and Kurtz (1981), it describes the temporal evolution of infinitely many frequencies. Define the closure of the infinite-dimensional ordered simplex

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

which is compact under the product topology, and consider, for constants $0 < \alpha < 1$ and $\theta > -\alpha$, the second order differential operator

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

where δ_{ij} is Kronecker delta and \mathcal{B} acts on functions belonging to a dense sub-algebra of the space $C(\bar{\nabla}_\infty)$ of continuous functions on $\bar{\nabla}_\infty$. The closure of \mathcal{B} in $C(\bar{\nabla}_\infty)$ generates a Feller semigroup on $C(\bar{\nabla}_\infty)$, which characterizes the finite-dimensional distributions of the two-parameter model, and the sample paths of the associated process are almost surely continuous functions from $[0, \infty)$ to $C(\bar{\nabla}_\infty)$. The first term in (2) describes the instantaneous covariance, related to the allelic resampling, also called *random genetic drift*, while the term in θ describes a drift behaviour for each component, determined by an underlying *mutation* process. In order to discuss further the interpretation of \mathcal{B} , it is worth noting at this point that the one-parameter model, obtained by setting $\alpha = 0$ in (2), admits the following two interpretations: the first as an *unlabeled* diffusion of Fleming–Viot type with parent-independent mutation and nonatomic mutant offspring distribution on some uncountable type space, obtained after reordering the frequencies by their sizes, see Ethier and Kurtz (1993) (here with labelling we mean appending a mark that indicates the type associated to each frequency); the second as the limit of a sequence of Wright–Fisher processes, obtained by projecting the same Fleming–Viot process onto a partition of the type space, as the number of partition sets goes to infinity, see Dawson (1993). As a result of these correspondences, the interpretation of θ as total mutation rate is unambiguous and independent on the dimension

of the allelic space. Similar interpretations for the two-parameter model, however, are not available: firstly, the existence of a *labeled*, measure-valued version of the process associated to (2) is still an open problem, hence the two-parameter case cannot be seen as a special case of a broader model; secondly, for $0 < \alpha < 1$, the finite-dimensional distributions of $\text{PD}(\theta, \alpha)$ are unknown in closed form, hence it is not clear whether the two-parameter model can be seen as the projective limit of a sequence of finite-dimensional models. Consequently, the interpretation of α cannot be deduced by measure-valued processes and a qualitative difference between settings with finitely-many and infinitely-many types cannot be ruled out. The fact that the role of α has been associated to mutation is mainly due to two existing particle constructions, given in Petrov (2009) and Ruggiero and Walker (2009), where θ and α jointly regulate births from the same distribution. Both constructions fall into the infinitely-many types setting, featuring Moran-type dynamics with overlapping generations and a diffuse mutant type distribution. Another existing construction, given in Ruggiero (2014) and obtained from finite-dimensional diffusions, is not based on a reproductive mechanism and thus is not particularly insightful in terms of the contribution of α to the model dynamics.

Here instead we are interested in a construction of the two-parameter model by means of a finite population with finitely-many types, closer in the spirit to a classical Wright–Fisher process with non-overlapping generations, since this would disclose details about how the reproduction acts at the individual level, which an inspection of \mathcal{B} does not reveal. As an illustration of this aspect, consider the construction of the one-parameter model via a classical Wright–Fisher chain with K alleles in a population of size N . The frequency of type i individuals after mutation is $z_i(1 - \sum_{j \neq i} u_{ij}) + \sum_{j \neq i} u_{ji}z_j$, where

$$(3) \quad u_{ij} = \frac{\theta}{N(K-1)} \wedge 1, \quad j \neq i,$$

is the probability of a mutation from type i to type j . It can be easily seen that the expected change of z_i , up to a factor of N , is given by the drift

$$(4) \quad \frac{\theta}{K-1}(1 - z_i) - \theta z_i,$$

which converges to $-\theta z_i$ when $K \rightarrow \infty$. See Ethier and Kurtz (1981) for more details. This construction provides insight into the role of θ in the mutation process, only partially readable from (2); it is indeed by inspection of u_{ij} that one can see that the probability of an individual mutation is inversely proportional to the population size and the mutant type distribution is uniform on the other $K - 1$ species, while the rate θ regulates how often the mutation events occur. Here we are after a similar insight, at the same level of magnification, on the action of α in the two-parameter model.

3 A Wright–Fisher model with frequency-dependent immigration

Consider a population of N individuals and fix, throughout the section, the maximum number of species in the population to be $K \geq 2$. The population size is assumed to be constant and individuals live only one period so that generations do not overlap. Denote by z_i the relative frequency of individuals in the current generation with a type- i allele at the selected locus. We assume the presence of an immigration mechanism, whereby, in order to keep the population size constant, immigrants replace current inhabitants. The immigration is governed by the state-dependent transition kernel $u_{ij}^{(\alpha)}(z)$, $j \neq i$, which gives the probability that a type- i individual is replaced by a type- j individual, so that the proportion of type i after immigration is

$$(5) \quad z_i^* = \sum_{j=1}^K z_j u_{ji}^{(\alpha)}(z), \quad u_{ii}^{(\alpha)}(z) = 1 - \sum_{j \neq i} u_{ij}^{(\alpha)}(z).$$

We also allow for mutation, which occurs from type i to type j with state-independent probability $u_{ij}^{(\theta)}$; hence the frequencies after mutations are

$$(6) \quad z_i^{**} = \sum_{j=1}^K z_j^* u_{ji}^{(\theta)}, \quad u_{ii}^{(\theta)} = 1 - \sum_{j \neq i} u_{ij}^{(\theta)}.$$

Finally, random genetic drift is modeled by multinomial resampling, which amounts to assume that each individual of the next generation chooses her parent at random from the current. Let

$$(7) \quad \Delta_K = \left\{ z \in [0, 1]^K : z_i \geq 0, \sum_{i=1}^K z_i = 1 \right\}$$

and let the current generation be given by the vector of proportions $z = (z_1, \dots, z_K) \in \Delta_K$. Then the next generation z' is formed according to the rule

$$(8) \quad z' \mid z \sim N^{-1} \text{MN}(N, z_1^{**}, \dots, z_K^{**}),$$

that is Nz' has a multinomial distribution with sample size N and mean vector $(Nz_1^{**}, \dots, Nz_K^{**})$.

Here we assume the mutation probabilities $u_{ji}^{(\theta)}$ are as in (3) with $\theta > 0$ (in Lemma 3.1 below we will be able to relax this constraint in order to meet the usual condition for PD(θ, α) distributions), and the immigration kernel is defined, for $0 \leq \alpha < 1$, by

$$(9) \quad u_{ij}^{(\alpha)}(z) = \frac{\alpha r_i(z) p_j(z)}{N z_i}, \quad j \neq i.$$

transition kernel	transition occurrence	type distribution
mutation	$\frac{\theta}{N}$	$\frac{1}{K-1}$
immigration	$\frac{\alpha r_i(z)}{n_i}$	$p_j(z)$

Table 1: Decomposition of the transition kernels.

Before specifying the conditions we require for $r_i(z)$ and $p_j(z)$, it is instructive to compare the mutation and immigration kernels by decomposing them into two simpler conditional events, in order to evaluate under the magnifying glass the effect of the inhomogeneity in (9). These are described in Table 1 as follows:

(i) *transition occurrence*:

- at speed controlled by θ , individuals undergo a mutation with probability inversely proportional to the population size;
- at speed controlled by α individuals become susceptible to a replacement via immigration with probability inversely proportional to the current size n_i of their type sub-population; a susceptible type- i individual is then removed with state-dependent probability $r_i(z)$;

(ii) *sampling of the incoming type*:

- the mutant type is chosen uniformly among the other $K - 1$ species;
- the immigrant is of type j with state-dependent probability $p_j(z)$.

Hence the mutation kernel acts uniformly on the individuals for choosing the starting state and uniformly on the species for choosing the arrival state. The inhomogeneity of the immigration kernel instead clearly depends on the choice of $r_i(z)$ and $p_j(z)$, whose definition is based on the following system of Bernoulli trials parametrized by z . Consider an array $\zeta = (\zeta_{ij})_{i,j=1,\dots,K}$ such that

$$(10) \quad \zeta_{ij} \stackrel{iid}{\sim} \text{Bernoulli}(z_i), \quad j = 1, \dots, K,$$

and define

$$(11) \quad r_i(z) = \mathbb{P} \left\{ \sum_{j=1}^K \zeta_{ij} > 0 \right\} = 1 - (1 - z_i)^K,$$

that is the probability of observing at least a success along the i th row. This corresponds to the dartboard experiment outlined in the Introduction. Furthermore, define

$$(12) \quad p_j(z) \propto \sum_{\ell: z_\ell \geq z_j} \mathbb{P}\{\zeta_{\ell j} = 0\} = \sum_{\ell: z_\ell \geq z_j} (1 - z_\ell),$$

that is $p_j(z)$ is proportional to the sum of failure probabilities along the j -th column, where the sum extends over rows ℓ such that $z_\ell \geq z_j$. Note that $p_j(z)$ can be seen as a continuous version of the normalised rank of z_j in z , namely $2j/(K(K+1))$, when the z_j 's are decreasingly ordered. The overall effect of the immigration kernel (9) is a mix of redistribution and reinforcement among the frequencies. Indeed, scarcer species have higher chances of being chosen for replacement, which occurs with probability α/n_i , but, once chosen, have lower probability of being replaced, since $r_i(z)$ is increasing in z_i ; the combined effect of $r_i(z)/n_i = r_i(z)/(Nz_i)$ favours the replacement of scarcer species, thus yielding a reinforcement, as it can be easily checked that $r_i(z)/z_i$ is decreasing in z_i . Finally, scarcer species have higher probability of being the immigrant types, since $p_j(z)$ is decreasing in the rank of z_j and favours a redistribution. See Remark 5.3 in Section 5 for possible extensions in the definition of $r_i(z)$ and $p_j(z)$.

From (5) and (6), and exploiting the linearity of (4), we can write the frequency of type- i individuals at reproductive age in terms of that before the action of immigration and mutation as

$$(13) \quad z_i^{**} = z_i + b_i(z)/N + o(N^{-1}),$$

where, in view of the rescaling, we have isolated the relevant drift term for the i th component

$$(14) \quad b_i(z) = \frac{1}{2} \left[\frac{\theta}{K-1} (1 - z_i) - \theta z_i + \alpha p_i(z) \sum_{j=1}^K r_j(z) - \alpha r_i(z) \right].$$

For later use, we record the following Lemma, stating the boundary conditions on the drift which are needed in the Section 4.

Lemma 3.1. *Let $r_i(z)$ and $p_i(z)$ satisfy (11)-(12), and let*

$$(15) \quad 0 \leq \alpha < 1 \quad \text{and} \quad \theta > -\alpha.$$

Then (14) satisfies the boundary conditions

$$b_i(z) \geq 0 \quad \text{if } z_i = 0 \quad \text{and} \quad b_i(z) \leq 0 \quad \text{if } z_i = 1,$$

for all parameter values.

Proof. When $z_i = 0$, $r_i(z) = 0$, cf. (11), so that (14) reduces to

$$b_i(z) = \frac{1}{2} \left[\frac{\theta}{K-1} + \alpha p_i(z) \sum_{j \neq i} r_j(z) \right],$$

which is clearly nonnegative since $\alpha \geq 0$, cf. (15). When $z_i = 1$, $z_j = 0$ for all $j \neq i$ so that $r_i(z) = 1$ and $r_j(z) = 0$ for all $j \neq i$. Moreover, $z_i = 1$ implies $p_i(z) = 0$ by (12), so that $b_i(z) = -(\theta + \alpha)/2$, which is nonpositive under (15). \square

Note that the constraint (15) reflects the usual parameter range for PD(θ, α) distributions with positive α .

4 Diffusion approximation with K types

Recall (7), and define the second-order differential operator

$$(16) \quad \mathcal{A}_K = \frac{1}{2} \sum_{i,j=1}^K z_i(\delta_{ij} - z_j) \frac{\partial^2}{\partial z_i \partial z_j} + \sum_{i=1}^K b_i(z) \frac{\partial}{\partial z_i},$$

with $b_i(z)$ as in (14). The domain of \mathcal{A}_K is taken to be

$$\mathcal{D}(\mathcal{A}_K) = \{f : f \in C^2(\Delta_K)\}$$

where

$$C^2(\Delta_K) = \{f \in C(\Delta_K) : \exists \tilde{f} \in C^2(\mathbb{R}^K), \tilde{f}|_{\Delta_K} = f\},$$

and $C(\Delta_K)$ is endowed with the supremum norm $\|f\| = \sup_{z \in \Delta_K} |f(z)|$. Denote also by $C_E([0, \infty))$ the space of continuous functions from $[0, \infty)$ to E , and let $\mathcal{P}(E)$ be the space of Borel probability measures on E . The following result states that \mathcal{A}_K characterizes a Feller diffusion on Δ_K .

Theorem 4.1. *Let \mathcal{A}_K be as in (16). The closure in $C(\Delta_K)$ of \mathcal{A}_K is single-valued and generates a Feller semigroup $\{\mathcal{T}_K(t)\}$ on $C(\Delta_K)$. For each $\nu_K \in \mathcal{P}(\Delta_K)$, there exists a strong Markov process $Z^{(K)}(\cdot) = \{Z^{(K)}(t), t \geq 0\}$, with initial distribution ν_K , such that*

$$\mathbb{E}(f(Z^{(K)}(t+s)) | Z^{(K)}(u), u \leq s) = \mathcal{T}_K(t)f(Z^{(K)}(s)), \quad f \in C(\Delta_K).$$

Furthermore,

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

Proof. It can be easily seen that \mathcal{A}_K satisfies the positive maximum principle on Δ_K , that is if $f \in \mathcal{D}(\mathcal{A}_K)$, $z_0 \in \Delta_K$ are such that $f(z_0) = \|f\| \geq 0$, then $\mathcal{A}_K f(z_0) \leq 0$. This is immediate in the interior of Δ_K , while on the boundaries it follows from Lemma 3.1. Denote now $z^m = z_1^{m_1} \cdots z_K^{m_K}$ and $m - \delta_i = (m_1, \dots, m_i - 1, \dots, m_K)$ for $m_1, \dots, m_K \in \mathbb{N}$. From (14) we have

$$\begin{aligned} \mathcal{A}_K z^m &= \frac{1}{2} \sum_{i=1}^K m_i(m_i - 1)(z^{m-\delta_i} - z^m) - \frac{1}{2} \sum_{1 \leq i \neq j \leq K} m_i m_j z^m \\ &\quad + \frac{1}{2} \sum_{i=1}^K m_i \left[\frac{\theta}{K-1} z^{m-\delta_i} - \frac{\theta K}{K-1} z^m \right. \\ &\quad \left. + \alpha \left(p_i(z) \sum_{j=1}^K r_j(z) - r_i(z) \right) z^{m-\delta_i} \right]. \end{aligned}$$

Since $r_i(z)$ and $p_i(z)$ are continuous in z , cf. (11) and (12), the image of \mathcal{A}_K is dense in $C(\Delta_K)$, and so is that of $\lambda - \mathcal{A}_K$ for all but at most countably many $\lambda > 0$. Since the algebras L_m of polynomials with degree at most m are such that $\cup_{m \geq 1} L_m$ is dense in $C(\Delta_K)$, the Hille-Yosida Theorem now implies that the closure of \mathcal{A}_K on $C(\Delta_K)$ is single-valued and generates a strongly continuous, positive, contraction semigroup $\{\mathcal{T}_K(t)\}$ on $C(\Delta_K)$. See for example Theorem 4.2.2 in Ethier and Kurtz, 1986. The Feller property is completed by noting that $(1, 0)$ belongs to the domain of $\bar{\mathcal{A}}_K$, so that $\{\mathcal{T}_K(t)\}$ is also conservative. The second assertion follows from Theorem 4.2.7 in Ethier and Kurtz (1986). Note now that for every $z_0 \in \Delta_K$ and $\epsilon > 0$ there exists $f \in \mathcal{D}(\mathcal{A}_K)$ such that

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

where $B^c(z_0, \epsilon)$ is a ball of radius ϵ centered at z_0 . Take for example $f(z) = -C_\epsilon \sum_{i=1}^K (z_i - z_0)^4$ for an appropriate constant C_ϵ which depends on ϵ . Then the third assertion follows from Theorem 4.2.7 and Remark 4.2.10 in Ethier and Kurtz (1986). \square

The diffusion of Theorem 4.1 is a good approximation, in the sense of the limit in distribution as the population size grows to infinity, of a suitably rescaled version of the Wright–Fisher Markov chain described in Section 3. This is formalized by the next Theorem. Here and later “ \xrightarrow{D} ” denotes convergence in distribution.

Theorem 4.2. Let $Z_N^{(K)}(\cdot) = \{Z_N^{(K)}(h), h \in \mathbb{N}\}$ be the Δ_K -valued Markov chain with one-step transitions as in (8) and (13), let $Z^{(K)}$ be the Feller diffusion of Theorem 4.1, and define $\tilde{Z}_N^{(K)}(t) := Z_N^{(K)}(\lfloor Nt \rfloor)$ for $t \geq 0$. If $Z_N^{(K)}(0) \xrightarrow{D} Z^{(K)}(0)$, then

$$\tilde{Z}_N^{(K)}(\cdot) \xrightarrow{D} Z^{(K)}(\cdot), \quad \text{in } C_{\Delta_K}([0, \infty)),$$

as $N \rightarrow \infty$.

Proof. From (8) and (13), and denoting $\Delta z_i = z'_i - z_i$, we have that conditionally on z

$$\mathbb{E}(\Delta z_i) = \mathbb{E}(z'_i - z_i^{**}) + \mathbb{E}(z_i^{**} - z_i) = N^{-1}b_i(z) + o(N^{-1}),$$

and, similarly,

$$\mathbb{E}(\Delta z_i \Delta z_j) = -N^{-1}z_i^{**}z_j^{**} + o(N^{-1}) = -N^{-1}z_i z_j + o(N^{-1}).$$

Furthermore, it can be easily seen that $\mathbb{E}((\Delta z_i)^4) = o(N^{-1})$, so that Chebyshev’s inequality implies Dynkin’s condition for the continuity of paths of the limit process, that is $\mathbb{P}(|\Delta z_i| > \delta \mid z) = o(N^{-1})$ for all $\delta > 0$. For $f \in C^2(\Delta_K)$, denote by T_N the semigroup operator associated to the Markov chain $Z_N^{(K)}(\cdot)$. Then a Taylor expansion, together with the above expressions, yields

$$\begin{aligned} (T_N - I)f(z) &= \mathbb{E} \left[\sum_{i=1}^K \Delta z_i \frac{\partial f}{\partial z_i} + \frac{1}{2} \sum_{i,j=1}^K \Delta z_i \Delta z_j \frac{\partial^2 f}{\partial z_i \partial z_j} + o \left(\sum_{i=1}^K (\Delta z_i)^2 \right) \middle| z \right] \\ &= \frac{1}{N} \sum_{i=1}^K b_i(z) \frac{\partial f}{\partial z_i} + \frac{1}{2N} \sum_{i,j=1}^K z_i (\delta_{ij} - z_j) \frac{\partial^2 f}{\partial z_i \partial z_j} + o(1), \end{aligned}$$

It follows that

$$\sup_{z \in \Delta_K} |N(T_N - I)f(z) - \mathcal{A}_K f(z)| \rightarrow 0, \quad f \in C^2(\Delta_K),$$

as $n \rightarrow \infty$, for \mathcal{A}_K as in (16). An application of Theorems 1.6.5 and 4.2.6 in Ethier and Kurtz (1986) implies the statement of the theorem, with $C_{\Delta_K}([0, \infty))$ replaced by the space $D_{\Delta_K}([0, \infty))$ of càdlàg functions. Using the fact that the limit process has continuous sample paths, the full statement now follows from a relativization of the Skorohod topology to $C_{\Delta_K}([0, \infty))$, see Section 18 of Billingsley (1968). \square

5 Convergence to the infinite-dimensional diffusion

Recall now (1), and define ρ_K to be the Borel measurable map $\rho_K : \Delta_K \rightarrow \bar{\nabla}_\infty$, such that

$$(17) \quad \rho_K(z) = (z_{(1)}, \dots, z_{(K)}, 0, 0, \dots), \quad z \in \Delta_K,$$

where $z_{(1)} \geq z_{(2)} \geq \dots$ are the ranked values of $z \in \Delta_K$. Let also

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

and

$$\nabla_K = \left\{ z \in \nabla_\infty : z_{K+1} = 0 \right\}.$$

Denote by \mathcal{B}_K the operator \mathcal{A}_K in (16) when taken to have domain

$$\mathcal{D}(\mathcal{B}_K) = \{ f \in C(\nabla_K) : f \circ \rho_K \in C^2(\Delta_K) \}.$$

Finally, recalling operator (2), consider the symmetric polynomials

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

sometimes termed *homozigosity*, and let

$$(18) \quad \mathcal{D}(\mathcal{B}) = \left\{ \text{subalgebra of } C(\bar{\nabla}_\infty) \text{ generated by } 1, \varphi_3(z), \varphi_4(z), \dots \right\}.$$

Note also that usually the domain for the infinite-dimensional operator is taken to be the algebra generated by $1, \varphi_2(z), \varphi_3(z), \dots$. The fact that also $\varphi_2(z)$ is excluded here is a peculiarity of the current construction, that is the convergence to the two-parameter model holds, under our assumptions, only on this restricted domain. Fortunately, this restriction is irrelevant, as implied by the following lemma, which is the same as Lemma 1 in [Ruggiero \(2014\)](#).

Lemma 5.1. $\mathcal{D}(\mathcal{B})$ is dense in $C(\bar{\nabla}_\infty)$.

The following result formalizes the convergence in distribution of the K -dimensional diffusion with operator \mathcal{B}_K to the two-parameter model.

Theorem 5.2. Let $Z^{(K)}(\cdot)$ be the Markov process of Theorem 4.1 with initial distribution $\nu_K \in \mathcal{P}(\Delta_K)$. Let also \mathcal{B} be as in (2) and denote $Z(\cdot)$ the Markov processes corresponding to the Feller semigroup generated by the closure in $C(\bar{\nabla}_\infty)$ of \mathcal{B} , with initial distribution $\nu \in \mathcal{P}(\bar{\nabla}_\infty)$. If $\nu_K \circ \rho_K^{-1} \xrightarrow{D} \nu$, then

$$\rho_K(Z^{(K)}(\cdot)) \xrightarrow{D} Z(\cdot), \quad \text{in } C_{\bar{\nabla}_\infty}([0, \infty)),$$

If in addition $\nu \in \mathcal{P}(\nabla_\infty)$, then the convergence holds in $C_{\nabla_\infty}([0, \infty))$.

Proof. Define $\pi_K : C(\bar{\nabla}_\infty) \rightarrow C(\Delta_K)$ by $\pi_K f = f \circ \rho_K$ and note that $\pi_K : \mathcal{D}(\mathcal{B}) \rightarrow \mathcal{D}(\mathcal{A}_K)$. Since for every $f \in \mathcal{D}(\mathcal{B})$ we have $\pi_K \mathcal{B}f = \mathcal{B}(f \circ \rho_K)$, for all such functions and $z \in \Delta_K$, from (14) we have

$$\begin{aligned} \mathcal{A}_K \pi_K f(z) - \pi_K \mathcal{B}f(z) &= \\ &= \frac{1}{2} \sum_{i=1}^K \left[\frac{\theta}{K-1} (1 - z_i) + \alpha p_i(z) \sum_{j=1}^K r_j(z) + \alpha [1 - r_i(z)] \right] \frac{\partial f(\rho_K(z))}{\partial z_i}, \end{aligned}$$

from which

$$\begin{aligned} (19) \quad |\mathcal{A}_K \pi_K f(z) - \pi_K \mathcal{B}f(z)| &\leq \frac{\theta}{K-1} \sum_{i=1}^K \left| \frac{\partial f(\rho_K(z))}{\partial z_i} \right| \\ &+ \alpha \sum_{i=1}^K p_i(z) \sum_{j=1}^K r_j(z) \left| \frac{\partial f(\rho_K(z))}{\partial z_i} \right| + \alpha \sum_{i=1}^K [1 - r_i(z)] \left| \frac{\partial f(\rho_K(z))}{\partial z_i} \right|. \end{aligned}$$

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

$$\sum_{i=1}^K \left| \frac{\partial f(z)}{\partial z_i} \right| = \sum_{i=1}^K \sum_{l=1}^k m_l z_i^{m_l-1} \prod_{h \neq l} \varphi_{m_h} \leq \sum_{l=1}^k m_l \sum_{i=1}^K z_i^{m_l-1},$$

which is bounded above by $\sum_{l=1}^k m_l < \infty$. Furthermore, using the fact that $m_l \geq 3$, see (18), the third term on the right hand side of (19) equals

$$\alpha \sum_{i=1}^K [1 - r_i(z)] \sum_{l=1}^k m_l z_i^{m_l-1} \leq \alpha \sum_{l=1}^k m_l \sum_{i=1}^K [1 - r_i(z)] z_i^2$$

Recall the definition of $r_i(z)$ in (11) and note that $\sup_{u \in [0,1]} (1-u)^K u^2 = o(K^{-1})$, which implies that $\sum_{i=1}^K [1 - r_i(z)] z_i^2 \rightarrow 0$ uniformly in $z \in \nabla_K$, as $K \rightarrow \infty$. As for the second term on the right hand side of (19), we have

$$\alpha \sum_{i=1}^K p_i(z) \sum_{j=1}^K r_j(z) \sum_{l=1}^k m_l z_i^{m_l-1} \leq \alpha K \sum_{l=1}^k m_l \sum_{i=1}^K p_i(z) z_i^2,$$

where we have used again $m_l \geq 3$ and the fact that $r_j(z) \leq 1$. Recall now the definition of $p_i(z)$ in (12) and define $\text{rank}_i(z) = \sum_{j: z_j \geq z_i} 1$ to be the maximum rank among the frequencies with the same value as z_i . Note that $\sum_{j: z_j \geq z_i} (1 - z_j) \leq \text{rank}_i(z)$ and that $\text{rank}_i(z) \geq i$ for $z \in \nabla_K$. Moreover $z_i \leq 1/\text{rank}_i(z)$ and the normalizing constant in $p_i(z)$ is larger than

$$\begin{aligned} \sum_{i=1}^K \sum_{j=1}^i (1 - z_j) &= \frac{K(K+1)}{2} - \sum_{j=1}^K (K-j+1) z_j \\ &\geq \frac{K(K+1)}{2} - K = \frac{K(K-1)}{2}. \end{aligned}$$

Hence we have

$$\sum_{i=1}^K p_i(z) z_i^2 \leq \sum_{i=1}^K \frac{2 \operatorname{rank}_i(z)}{K(K-1)} z_i^2 \leq \sum_{i=1}^K \frac{2}{K(K-1) \operatorname{rank}_i(z)} \leq \frac{2}{K(K-1)} \sum_{i=1}^K \frac{1}{i}$$

from which we conclude that

$$\sup_{z \in \nabla_K} \alpha K \sum_{l=1}^k m_l \sum_{i=1}^K p_i(z) z_i^2 \rightarrow 0.$$

From (19), the above arguments imply that

$$(20) \quad \|\mathcal{A}_K \pi_K f - \pi_K \mathcal{B} f\| \rightarrow 0, \quad f \in \mathcal{D}(\mathcal{B}).$$

Consider also that for all t , $\rho_K(Z^{(K)}(t)) \in \bar{\nabla}_\infty$, where $\bar{\nabla}_\infty$ is compact, hence $\rho_K(Z^{(K)}(\cdot))$ clearly satisfies a compact containment condition (cf. Remark 3.7.3 in Ethier and Kurtz, 1986), and that $\overline{\mathcal{D}(\mathcal{B})}$ contains the algebra generated by $1, \varphi_2(z), \varphi_3(z), \dots$, that separates the points of $\bar{\nabla}_\infty$ (see for example the proof of Theorem 4.2.5 in Ethier and Kurtz, 1981). Using now Lemma 5.1 for extending (20) to $f \in C(\bar{\nabla}_\infty)$, together with the fact that the closure of \mathcal{B} in $C(\bar{\nabla}_\infty)$ generates a strongly continuous contraction semigroup, all the hypotheses of Corollary 4.8.7 in Ethier and Kurtz (1986) are satisfied and the first assertion of the Theorem follows with $C_{\bar{\nabla}_\infty}([0, \infty))$ replaced by the space $D_{\bar{\nabla}_\infty}([0, \infty))$ of càdlàg functions. Additionally, the convergence 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 Skorohod topology relativized to $C_{\bar{\nabla}_\infty}([0, \infty))$ coincides with the uniform topology on $C_{\bar{\nabla}_\infty}([0, \infty))$. See for example Billingsley (1968), Section 18. Finally, the second assertion follows from Proposition 2.2 in Feng and Sun (2010) and from further relativization to $C_{\nabla_\infty}([0, \infty))$ of the topology on $C_{\bar{\nabla}_\infty}([0, \infty))$. \square

Remark 5.3. We conclude by noting that all the arguments which lead from the definition of the Wright–Fisher chain to the above convergence result are still valid, upon minor adaptations, if one replaces the Bernoulli(z_i) distribution with a Poisson(z_i) in (10). In fact, the key requirements for $r_i(z)$ and $p_j(z)$ for the convergence to the two-parameter model to hold are that $r_i(z) = r(z_i)$ for a continuous function $r : [0, 1] \rightarrow [0, 1]$ such that

$$\lim_{u \rightarrow 0} r(u) = 0, \quad \sup_{u \in [0, 1]} [1 - r(u)] u^2 = o(K^{-1}) \quad \text{as } K \rightarrow \infty,$$

together with the fact that $p(z) = (p_j(z), 1 \leq j \leq K)$ defines a continuous mapping $p : \Delta_K \rightarrow \Delta_K$ such that

$$\sup_{z \in \Delta_K} \sum_{j=1}^K p_j(z) z_j^2 = o(K^{-1}), \quad \text{as } K \rightarrow \infty.$$

\square

6 Acknowledgments

The first two authors are supported by the European Research Council (ERC) through StG “N-BNP” 306406.

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.
- BERTOIN, J. (2006). *Random fragmentation and coagulation processes*. Cambridge University Press, Cambridge.
- BILLINGSLEY, P. (1968). *Convergence of probability measures*. Wiley, New York.
- DAWSON, D.A. (1993). *Measure-valued Markov processes*. Ecole d’Eté de Probabilités de Saint Flour XXI. Lecture Notes in Mathematics **1541**. Springer, Berlin.
- ETHIER, S.N. and KURTZ, T.G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. Appl. Probab.* **13**, 429–452.
- ETHIER, S.N. and KURTZ, T.G. (1986). *Markov processes: characterization and convergence*. Wiley.
- ETHIER, S.N. and KURTZ, T.G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31**, 345–386.
- FENG, S. (2010). *The Poisson–Dirichlet distribution and related topics*. Springer, Heidelberg.
- FENG, S. and SUN, W. (2010). Some diffusion processes associated two-parameter Poisson–Dirichlet distribution and Dirichlet process. *Probab. Theory Relat. Fields* **148**, 501–525.
- 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.
- KINGMAN, J.F.C. (1975). Random discrete distributions. *J. Roy. Statist. Soc. Ser. B* **37**, 1–22.
- LIJOI, A. and PRÜNSTER, I. (2009). Models beyond the Dirichlet process. In Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. (Eds.), *Bayesian Nonparametrics*, Cambridge University Press.
- PETROV, L. (2009). Two-parameter family of diffusion processes in the Kingman simplex. *Funct. Anal. Appl.* **43**, 279–296.
- PITMAN, J. (1995). Exchangeable and partially exchangeable random partitions. *Probab. Theory and Relat. Fields* **102**, 145–158.

- PITMAN, J. (2006). *Combinatorial stochastic processes*. École d'été de Probabilités de Saint-Flour XXXII. Lecture Notes in Math. **1875**. Springer-Verlag.
- PITMAN, J. and YOR, M. (1997). The two-parameter Poisson–Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900.
- PÓLYA, G. and SZEGÖ, G. (1978). *Problems and theorems in analysis. I. Series, integral calculus, theory of functions*. Springer-Verlag, Berlin-New York.
- RUGGIERO, M. (2014). Species dynamics in the two-parameter Poisson–Dirichlet diffusion model. *J. Appl. Probab.* **51**, 174–190.
- RUGGIERO, M. and WALKER, S.G. (2009). Countable representation for infinite-dimensional diffusions derived from the two-parameter Poisson–Dirichlet process. *Electron. Comm. Probab.* **14**, 501–517.
- RUGGIERO, M., WALKER, S.G. and FAVARO, S. (2013). Alpha-diversity processes and normalized inverse-Gaussian diffusions. *Ann. Appl. Probab.* **23**, 386–425.
- TEH, Y.W. and JORDAN, M.I. (2009). Bayesian nonparametrics in machine learning. In Hjort, N.L., Holmes, C.C. Müller, P., Walker, S.G. (Eds.), *Bayesian Nonparametrics*, Cambridge University Press.