WRIGHT–FISHER CONSTRUCTION OF THE TWO-PARAMETER POISSON–DIRICHLET DIFFUSION

By Cristina Costantini[*,1], Pierpaolo De Blasi[†,2,3],
Stewart N. Ethier[‡,4], Matteo Ruggiero[†,2,3] and Dario Spanò[§]

*University of Chieti-Pescara[*], University of Torino[†], University of Utah[‡] and
University of Warwick[§]*

The two-parameter Poisson–Dirichlet diffusion, introduced in 2009 by Petrov, extends the infinitely-many-neutral-alleles diffusion model, related to Kingman's one-parameter Poisson–Dirichlet distribution and to certain Fleming–Viot processes. 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 formulating a K-allele Wright–Fisher model for a population of size N, involving a uniform mutation pattern and a specific state-dependent migration mechanism. Suitably scaled, this process converges in distribution to a K-dimensional diffusion process as $N \to \infty$. Moreover, the descending order statistics of the K-dimensional diffusion converge in distribution to the two-parameter Poisson–Dirichlet diffusion as $K \to \infty$. The choice of the migration mechanism depends on a delicate balance between reinforcement and redistributive effects. The proof of convergence to the infinite-dimensional diffusion is nontrivial because the generators do not converge on a core. Our strategy for overcoming this complication is to prove a priori that in the limit there is no "loss of mass", that is, that, for each limit point of the sequence of finite-dimensional diffusions (after a reordering of components by size), allele frequencies sum to one.

1. Introduction. The goal of this paper is to provide a discrete-time finite-population construction of the two-parameter Poisson–Dirichlet diffusion, extending an analogous construction for the well-known infinitely-many-neutral-alleles diffusion model provided in Ethier and Kurtz (1981). Introduced by Petrov (2009) and henceforth called the two-parameter model, this diffusion process assumes values in the infinite-dimensional ordered simplex (sometimes also called the King-

man simplex)

$$(1.1) \qquad \overline{\nabla}_\infty := \left\{ z = (z_1, z_2, \ldots) \in [0, 1]^\infty : z_1 \geq z_2 \geq \cdots \geq 0, \sum_{i=1}^\infty z_i \leq 1 \right\}$$

and describes the temporal evolution of the ranked frequencies of infinitely many potential alleles, observed at a single gene locus, in a given population of large but finite 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 study some path properties using Dirichlet forms; Feng et al. (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 normalised 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; Ethier (2014), who shows that, with probability one, the diffusion instantly enters the dense subset

$$(1.2) \qquad \nabla_\infty := \left\{ z = (z_1, z_2, \ldots) \in [0, 1]^\infty : z_1 \geq z_2 \geq \cdots \geq 0, \sum_{i=1}^\infty z_i = 1 \right\}$$

and never exits; and Zhou (2015), who simplifies the formula for the transition density and establishes an ergodic inequality.

The two-parameter model is known to be reversible [Petrov (2009)] with respect to the two-parameter Poisson–Dirichlet distribution $PD(\theta, \alpha)$, where $0 \leq \alpha < 1$ and $\theta > -\alpha$ [Perman, Pitman and Yor (1992), Pitman (1995), Pitman and Yor (1997)]. When $\alpha = 0$, the model reduces to the infinitely-many-neutral-alleles diffusion model, henceforth called the one-parameter model, with Poisson–Dirichlet reversible distribution $PD(\theta) := PD(\theta, 0)$ [Kingman (1975)]. The two-parameter Poisson–Dirichlet distribution $PD(\theta, \alpha)$ has found numerous applications in several fields: See, for example, Bertoin (2006) for fragmentation and coalescent theory; Pitman (2006) for excursion theory and combinatorics; Lijoi and Prünster (2009) for Bayesian inference; and Teh and Jordan (2009) for machine learning. However, in the dynamic setting, the two-parameter model is not as well understood as the one-parameter special case, which motivates the need for further investigation.

One of the main differences between the $PD(\theta)$ and $PD(\theta, \alpha)$ 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 dynamical framework, one possible construction of the one-parameter model is as the limit in distribution as $K \to \infty$ of a K-dimensional diffusion process of Wright–Fisher type with components rearranged in descending order. Each of these Wright–Fisher diffusions can in turn be constructed as

the limit in distribution as $N \to \infty$ of a suitably scaled K-allele Wright–Fisher Markov chain model for a randomly mating population of size N with discrete nonoverlapping generations and uniform mutation [Ethier and Kurtz (1981)]. In contrast, an analogous construction in the case $0 < \alpha < 1$ has not, to the best of our knowledge, been published. The importance of finding examples of processes with these features for the two-parameter model lies in the possibility of revealing the reproductive mechanisms acting at the level of individuals, thus providing interpretation for the roles played by the parameters θ and α in the dynamics of the population's allele 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 derived from a Wright–Fisher Markov chain model. As with the one-parameter model, there are two limit operations involved. We start with a K-allele Wright–Fisher model for a randomly mating population of size N with discrete nonoverlapping generations, a uniform mutation pattern, and a specific state-dependent migration mechanism. It is not difficult to see that this process, suitably scaled, converges in distribution to a K-dimensional Wright–Fisher diffusion as $N \to \infty$. The process obtained by applying the descending order statistics to this Wright–Fisher diffusion is itself a diffusion (i.e., the Markov property is retained), which we show converges in distribution to the two-parameter model as $K \to \infty$.

We also show that the two-parameter Poisson–Dirichlet distribution $\mathrm{PD}(\theta, \alpha)$ is the weak limit of the stationary distributions of the Wright–Fisher diffusions we obtain (modified to account for the rearranging of components in descending order), by analogy to what happens in the one-parameter case, where these stationary distributions are symmetric Dirichlet distributions.

Our Wright–Fisher model includes migration and mutation. Mutation is uniform as before but with mutation rate proportional to $\theta + \alpha$ instead of just θ. Migration, which acts first, also depends on α and is governed by a generalisation of the classical island model. In that model, the frequency of allele i on the island after migration (in the gametic pool) is

$$(1.3) \qquad z_i^* = z_i + p_i m - z_i m,$$

with z_i being its frequency on the island prior to migration, m being the migration rate, and p_i being the frequency of allele i in the mainland population. We generalise this in two ways, neither of which is conventional in the population genetics literature. First, we allow the migration rate to be allele-dependent, so that (1.3) is replaced by

$$(1.4) \qquad z_i^* = z_i + p_i m(z) - z_i m_i \qquad \text{where } m(z) = \sum_{j=1}^{K} z_j m_j.$$

Here, m_i is the migration rate for allele i and $m(z)$ is the overall migration rate. The second generalisation allows all parameters to be state-dependent, that is, to depend on the vector z of allele frequencies on the island. Thus, (1.4) is replaced by

$$(1.5) \qquad z_i^* = z_i + p_i(z)m(z) - z_i m_i(z) \qquad \text{where } m(z) = \sum_{j=1}^{K} z_j m_j(z).$$

Here, $m_i(z)$ is the migration rate for allele i, $m(z)$ is the overall migration rate, and $p_i(z)$ is the frequency of allele i in the mainland population. The form of the functions $m_i(z)$ and $p_i(z)$ will be specified later on, but for now we point out only that $m_i(z)$ depends on z_i alone and is a decreasing function of that variable that does not depend in i, and $p_i(z) > p_j(z)$ if $z_i < z_j$. Thus, more frequent alleles on the island are less likely to emigrate (so emigration provides a reinforcement effect), and less frequent alleles on the island are more frequent on the mainland and, therefore, more likely to immigrate (so immigration provides a redistributive effect).

The proof of convergence in distribution of the K-dimensional Wright–Fisher diffusion, with components rearranged in descending order, to the two-parameter model as $K \to \infty$ is nontrivial and requires a new approach. The difficulty arises essentially from the fact that, with $\mathcal{B}_K$ denoting the generator of the reordered K-dimensional diffusion, and $\mathcal{B}$ denoting the generator of the two-parameter model, $\mathcal{B}_K \varphi$ does not converge to $\mathcal{B}\varphi$ on $\overline{\nabla}_\infty$ for certain φ in the domain $\mathscr{D}(\mathcal{B})$ of $\mathcal{B}$ (see Section 2). The simplest such φ is the so-called *homozygosity*, $\varphi_2(z) := \sum_{i=1}^{\infty} z_i^2$. At the same time, it is not possible to eliminate φ_2 from the domain of $\mathcal{B}$, because the resulting space of functions would not be a core for the closure of $\mathcal{B}$. As a consequence, the approach followed in Ethier and Kurtz (1981) to study the one-parameter model fails here, and so do various other similar approaches. A more complete discussion of these issues can be found at the beginning of Section 5.

Here, we take the martingale problem approach, that is, we view the reordered K-dimensional diffusion as the solution of the martingale problem for $\mathcal{B}_K$, and, as is usual in this approach, try to carry out three steps: (i) Show that the sequence of finite-dimensional diffusions is relatively compact; (ii) Show that each of its limit points is a solution to the martingale problem for $\mathcal{B}$; (iii) Show that the martingale problem for $\mathcal{B}$ has a unique solution. As may be expected, the difficulty described above shows up in this approach as well: If the domain of $\mathcal{B}$ includes φ_2, then it is not clear that the limit martingale property will hold for the pair $(\varphi_2, \mathcal{B}\varphi_2)$. On the other hand, if φ_2 is excluded from the domain of $\mathcal{B}$, then the martingale problem for $\mathcal{B}$ may have more than one solution.

However, in the martingale problem framework we are able to overcome the difficulty by proving a priori that, for any limit point Z of the sequence of finite-dimensional diffusions, with probability one, for almost all $t \geq 0$, $Z(t)$ belongs to ∇_∞ [cf. (1.2)]. The argument employed in this proof was inspired by the proof of

Theorem 2.6 of Ethier and Kurtz (1981) and relies on a double limit, taken in the appropriate order. When restricted to ∇_∞, $\mathcal{B}_K\varphi_2$ does converge to $\mathcal{B}\varphi_2$, and this yields that the limit martingale property carries over to $(\varphi_2, \mathcal{B}\varphi_2)$, and thus that the limit martingale problem has a unique solution.

In the one-parameter case, it is possible to also formulate a Wright–Fisher model with infinitely many alleles and obtain the limit process as $N \to \infty$ (K is already ∞); see Ethier and Kurtz (1981), Theorem 3.3. That theorem requires some rather delicate estimates and we were unsuccessful in trying to extend it to the two-parameter setting.

The paper is organised as follows. In Section 2, the two-parameter model is recalled. Section 3 provides the construction of the K-allele Wright–Fisher Markov chain for a population of size N. In Section 4, the Wright–Fisher chain, scaled appropriately, is shown to converge in distribution to a K-dimensional Wright–Fisher diffusion as $N \to \infty$. Then, in Section 5, the K-dimensional diffusion, with coordinates rearranged in descending order, is shown to converge to the two-parameter model as $K \to \infty$. In Section 6, analogous results are proved for the stationary distributions. Section 7 concludes by highlighting a slightly simpler formulation, obtained under the assumption that $\theta \geq 0$, which allows us to separate the roles of θ and α in driving the population dynamics.

2. The two-parameter model.   The two-parameter model was introduced by Petrov (2009). As with its one-parameter counterpart, characterised in Ethier and Kurtz (1981), it describes the temporal evolution of infinitely many allele frequencies. A natural state space for the process is ∇_∞, defined in (1.2). However, the closure of ∇_∞ (in the product topology on $[0, 1]^\infty$), namely $\overline{\nabla}_\infty$, defined in (1.1), is compact and, therefore, more convenient as a state space. Consider, for parameters $0 \leq \alpha < 1$ and $\theta > -\alpha$, the second-order differential operator $\mathcal{B}$ defined as follows. The domain of $\mathcal{B}$ is

$$(2.1) \qquad \mathscr{D}(\mathcal{B}) := \text{subalgebra of } C(\overline{\nabla}_\infty) \text{ generated by } \varphi_1, \varphi_2, \varphi_3, \ldots,$$

where $\varphi_1 \equiv 1$ and, for $m = 2, 3, \ldots,$ φ_m is defined by

$$(2.2) \qquad \varphi_m(z) := \sum_{i=1}^{\infty} z_i^m.$$

For $\varphi \in \mathscr{D}(\mathcal{B})$, $\mathcal{B}\varphi$ is the continuous extension to $\overline{\nabla}_\infty$ of

$$(2.3) \quad \mathcal{B}\varphi(z) := \frac{1}{2} \sum_{i,j=1}^{\infty} z_i(\delta_{ij} - z_j)\frac{\partial^2\varphi(z)}{\partial z_i \, \partial z_j} - \frac{1}{2} \sum_{i=1}^{\infty} (\theta z_i + \alpha)\frac{\partial\varphi(z)}{\partial z_i}, \qquad z \in \nabla_\infty,$$

with δ_{ij} the Kronecker delta. For example,

$$(2.4) \qquad \mathcal{B}\varphi_2(z) := 1 - \alpha - (1 + \theta)\varphi_2(z), \qquad z \in \overline{\nabla}_\infty.$$

As shown by Petrov (2009), the closure of $\mathcal{B}$ generates a Feller semigroup on $C(\overline{\nabla}_\infty)$, which characterises the finite-dimensional distributions of the two-parameter model Z, and the sample paths of Z belong to $C_{\overline{\nabla}_\infty}[0, \infty)$ with probability one. Recently, Ethier (2014) proved that, for an arbitrary initial distribution $\nu \in \mathscr{P}(\overline{\nabla}_\infty)$, we have

$$\mathbb{P}\big(Z(t) \in \nabla_\infty \text{ for every } t > 0\big) = 1,$$

that is, $\overline{\nabla}_\infty - \nabla_\infty$ acts as an entrance boundary (note however that technically it is not a boundary because ∇_∞ has no interior). In particular, if $\nu(\nabla_\infty) = 1$, then the sample paths of Z belong to $C_{\nabla_\infty}[0, \infty)$ with probability one.

The diffusion coefficients in the first term of (2.3) describe the instantaneous covariance, related to the allelic sampling, also called *random genetic drift*. The interpretation of the drift coefficients in the second term of (2.3) is not as clear, and is the object of primary interest in this paper. It is worth noting that the one-parameter model, obtained by setting $\alpha = 0$ in (2.3), admits the following two interpretations.

First, the one-parameter model, also known as the *unlabelled* infinitely-many-neutral-alleles diffusion model, has a more informative *labelled* version, namely the Fleming–Viot process in $\mathscr{P}(S)$ (the set of Borel probability measures on the compact metric space S with the topology of weak convergence) with mutation operator

$$Ag(x) := \frac{1}{2}\theta \int_S \big(g(\xi) - g(x)\big)\nu_0(d\xi),$$

where $\nu_0 \in \mathscr{P}(S)$ is nonatomic. The unlabelled model is a transformation of the labelled one. The transformation takes $\mu \in \mathscr{P}(S)$ to $z \in \overline{\nabla}_\infty$, where z is the vector of descending order statistics of the sizes of the atoms of μ. See Ethier and Kurtz (1993).

The second interpretation is as the limit in distribution of a K-allele Wright–Fisher diffusion, with components rearranged in descending order, as $K \to \infty$, where the rate of a mutation from allele i to allele j is proportional to θ. See Ethier and Kurtz (1981).

As a result of these correspondences, θ is usually interpreted as the rate at which mutations occur. Similar interpretations for the two-parameter model, however, are not available: First, the existence of a Fleming–Viot process whose unlabelled version is the two-parameter model is an open problem [posed in Feng (2010)]. Second, for $0 < \alpha < 1$, a Kingman-type result expressing $PD(\theta, \alpha)$ as the limit in distribution of a sequence of finite-dimensional random vectors is not available; hence it does not offer a guide for a Wright–Fisher construction, as in the one-parameter case. Consequently, the interpretation of α cannot be deduced from existing work. The role of α has been associated rather indirectly to mutation in a particle construction of the two-parameter model, given in Ruggiero and Walker (2009), where

θ and α jointly regulate births from the same distribution. They propose a Moran-type process for the evolution of N individuals, whereby at exponential times a randomly chosen individual is removed from the population after either giving a simple birth, with the offspring inheriting the parent's type, or giving a birth with mutation, with the offspring being of a type not previously observed. The probabilities of these events are regulated by the weights of Pitman's generalisation of the Blackwell–MacQueen Pólya urn scheme Pitman (1995, 1996). In particular, with probability proportional to $\theta + \alpha k$, where k is the current number of distinct types present in the population, a birth with mutation occurs, whereas with probability proportional to $n_j - \alpha k$, where n_j is the current number of type-j individuals, a simple birth of type j occurs. The original sequential construction in Petrov (2009) instead relates to a discrete Markov chain on the space of partitions of $\{1, \ldots, N\}$ and offers no insight into the role of α at the reproduction level. Both these constructions feature overlapping generations and fall into the infinitely-many-types setting, in the sense that they both allow the possibility of new types appearing in the population chosen from an uncountable genetic pool.

Here, instead, we are interested in a construction of the two-parameter model by means of a classical Wright–Fisher Markov chain, with nonoverlapping generations and finitely many types, since this would reveal details about how the reproduction acts at the individual level, which an inspection of $\mathcal{B}$ does not. As an illustration of this aspect, consider the construction of the one-parameter model via a Wright–Fisher Markov chain with K alleles in a population of size N. If $z = (z_1, \ldots, z_K)$ is the vector of allele frequencies prior to mutation, the frequency of allele i individuals after mutation is $z_i(1 - \sum_{j: j \neq i} u_{ij}) + \sum_{j: j \neq i} z_j u_{ji}$, where

$$(2.5) \qquad u_{ij} := \frac{\theta}{2N(K-1)}, \qquad j \neq i,$$

is the proportion of individuals of allele i that mutate to allele j, for sufficiently large N. It can be easily seen that the expected change of z_i, multiplied by N, is given by the drift coefficient

$$(2.6) \qquad \frac{1}{2}\left[\frac{\theta}{K-1}(1 - z_i) - \theta z_i\right],$$

which converges to $-(1/2)\theta z_i$ when $K \to \infty$. 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 (2.3) with $\alpha = 0$; it is indeed by inspection of (2.5) 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 $K - 1$ alleles; the rate θ determines how often the mutation events occur.

Here, we seek a similar insight, at the same level of magnification, on the action of α in the two-parameter model. In this case, the drift coefficients in (2.3) are

$-\frac{1}{2}(\theta z_i + \alpha)$. The key observation for the following development is to think of them as

$$(2.7) \qquad -\frac{1}{2}(\theta + \alpha)z_i - \frac{1}{2}\alpha(1 - z_i),$$

the first term corresponding to mutation and the second term to migration. The first term is the limit as $K \to \infty$ of the analogue of (2.6), namely

$$\frac{1}{2}\left[\frac{\theta + \alpha}{K - 1}(1 - z_i) - (\theta + \alpha)z_i\right],$$

while the second term should be the limit of the migration terms in the K-allele drift coefficients.

3. A W–F model with state-dependent migration.
Consider a population of N individuals, and let the maximum number of alleles in the population be $K \geq 2$. The population size is assumed to be constant and generations are nonoverlapping. Denote by z_i the relative frequency of allele i in the current generation at the selected locus. We assume the presence of migration and mutation, as discussed in Section 1. The state space is

$$(3.1) \qquad \Delta_K := \left\{z = (z_1, \ldots, z_K) \in [0, 1]^K : z_1 \geq 0, \ldots, z_K \geq 0, \sum_{i=1}^{K} z_i = 1\right\}$$

or, more precisely,

$$(3.2) \qquad \Delta_K^N := \left\{z = (z_1, \ldots, z_K) \in \Delta_K : Nz \in \mathbb{Z}^K\right\}.$$

The frequency of allele i after migration (in the gametic pool) is

$$(3.3) \qquad z_i^* = z_i + p_i(z)m(z) - z_i m_i(z) \qquad \text{where } m(z) = \sum_{j=1}^{K} z_j m_j(z),$$

as discussed in the Introduction [see (1.5)]. With u_{ij} denoting the proportion of individuals of allele i that mutate to allele j, the frequency of allele i after mutation (in the gametic pool) is

$$(3.4) \qquad z_i^{**} := z_i^* + \sum_{j:j\neq i} z_j^* u_{ji} - z_i^* \sum_{j:j\neq i} u_{ij}.$$

Finally, random genetic drift is modelled by multinomial sampling, which amounts to assuming that each individual of the next generation chooses its parent at random from the current generation. Then the next generation's allele frequencies $z_1', \ldots, z_K'$ are formed according to the rule

$$(3.5) \qquad z' \mid z \sim N^{-1} \text{ multinomial}(N, z_1^{**}, \ldots, z_K^{**}),$$

that is, Nz' has a multinomial distribution with sample size N and cell probabilities $(z_1^{**}, \ldots, z_K^{**})$. This is the classic Wright–Fisher model with migration and mutation in the state space Δ_K^N, and without migration it corresponds to equation (2.2) in Ethier and Kurtz (1981). For a more complete description of the Wright–Fisher model and its underlying assumptions, we refer the reader to Section 9.9 of Nagylaki (1992), but with selection replaced by migration. [In Nagylaki's notation, (9.158), (9.155) and (9.146) are replaced respectively by $p_i^* = p_i + \hat{p}_i(\mathbf{p})m(\mathbf{p}) - p_i m_i(\mathbf{p})$ with $m(\mathbf{p}) = \sum_j p_j m_j(\mathbf{p})$, $\widetilde{P}_{ij}^* = p_i^* p_j^*$, and $P_{ij}^* = (2 - \delta_{ij}) p_i^* p_j^*$. Finally, our N is Nagylaki's $2N$.]

We turn now to specifying the migration and mutation in sufficient detail to derive a K-allele diffusion approximation. Consider parameter values of $0 \le \alpha < 1$ and $\theta > -\alpha$ (the case $\theta \ge 0$, which allows a simplification, is treated separately in Section 7). We assume that the migration rates are given by

$$(3.6) \qquad m_i(z) := \frac{\alpha r_i(z)}{2N}, \qquad i = 1, \ldots, K,$$

and that the mutation rates u_{ij} are given by

$$(3.7) \qquad u_{ij} := \frac{\theta + \alpha}{2N(K-1)}, \qquad j \ne i,$$

for sufficiently large N [cf. (2.5)]. The functions p_i in (3.3) and r_i in (3.6), defined on Δ_K for $i = 1, \ldots, K$, are assumed to satisfy the following properties: $(p_1, \ldots, p_K)$ is a C^4 map of Δ_K into Δ_K and is symmetric in the sense that, for every permutation σ of $\{1, 2, \ldots, K\}$,

$$p_i(z_{\sigma(1)}, \ldots, z_{\sigma(K)}) = p_{\sigma(i)}(z), \qquad i = 1, \ldots, K, z \in \Delta_K;$$

$p_i(z) > p_j(z)$ if $z_i < z_j$ for all $z \in \Delta_K$ and $i \ne j$; $r_i(z) = r(z_i)$ for $i = 1, \ldots, K$ and $z \in \Delta_K$, where $r : [0, 1] \mapsto [0, \infty)$ is C^4 and is decreasing.

In Section 5, we will be more specific as to the form of $p_i(z)$ and $r_i(z)$ [see (5.2) and (5.3) below]. In Section 7, we will give a simpler formulation of $p_i(z)$ and $r_i(z)$ in the special case $\theta \ge 0$. Here and later, for notational simplicity, we suppress the dependence on K of the defined quantities whenever this does not create confusion.

To summarise, our Markov chain $Z_K^N(\cdot) = \{Z_K^N(\tau), \tau = 0, 1, \ldots\}$ has state space Δ_K^N [see (3.2)] and its transition probabilities are specified by (3.3)–(3.7).

From (3.3) and (3.4), we can write the frequency of allele i (in the gametic pool) at reproductive age in terms of the allele frequencies before the action of migration and mutation as

$$(3.8) \qquad z_i^{**} = z_i + N^{-1} b_i(z) + o(N^{-1}),$$

uniformly in $z \in \Delta_K^N$, where, in view of the rescaling, we have isolated the relevant drift term for the ith component, namely

$$(3.9) \quad b_i(z) := \frac{1}{2} \left[\frac{\theta + \alpha}{K - 1}(1 - z_i) - (\theta + \alpha)z_i + \alpha p_i(z) \sum_{j=1}^{K} z_j r_j(z) - \alpha z_i r_i(z) \right].$$

For later use, note that

$$(3.10) \qquad b_i(z) \geq 0 \qquad \text{if } z_i = 0, i = 1, 2, \ldots, K,$$

and

$$(3.11) \qquad \sum_{i=1}^{K} b_i(z) = 0, \qquad z \in \Delta_K.$$

4. Diffusion approximation with K alleles. Recall (3.1), and define the second-order differential operator

$$(4.1) \qquad \mathcal{A}_K := \frac{1}{2} \sum_{i,j=1}^{K} a_{ij}(z) \frac{\partial^2}{\partial z_i \, \partial z_j} + \sum_{i=1}^{K} b_i(z) \frac{\partial}{\partial z_i}, \qquad \mathscr{D}(\mathcal{A}_K) = C^2(\Delta_K),$$

with

$$(4.2) \qquad a_{ij}(z) := z_i (\delta_{ij} - z_j)$$

and $b_i(z)$, which of course depends on K, as in (3.9). Here,

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

and the choice of the extension $\tilde{f}$ to which the partial derivatives are applied does not matter. Let $C(\Delta_K)$ be endowed with the supremum norm. The following result states that $\mathcal{A}_K$ characterises a Feller diffusion on Δ_K.

PROPOSITION 4.1. *Let $\mathcal{A}_K$ be as in (4.1)–(4.2) and (3.9). Then 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 $\nu_K \in \mathscr{P}(\Delta_K)$, there exists a strong Markov process $Z_K(\cdot) = \{Z_K(t), t \geq 0\}$, with initial distribution ν_K, such that*

$$\mathbb{E}(f(Z_K(t+s)) \mid Z_K(u), u \leq s) = \mathcal{T}_K(t) f(Z_K(s)), \qquad f \in C(\Delta_K), s, t \geq 0.$$

Furthermore,

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

PROOF. Noting that $b_1, \ldots, b_K \in C^4(\Delta_K)$, the first assertion follows from Ethier (1976) and Sato (1978), using (3.10) and (3.11). The second assertion follows from Theorem 4.2.7 in Ethier and Kurtz (1986). Note that for every $z^0 \in \Delta_K$ and $\varepsilon > 0$ there exists $f \in \mathscr{D}(\mathcal{A}_K)$ such that

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

where $B(z^0, \varepsilon)$ is the ball of radius ε centred at z^0. Take, for example, $f(z) := 2 - \sum_{i=1}^{K}(z_i - z_i^0)^4$. Then the third assertion follows from Proposition 4.2.9 and Remark 4.2.10 in Ethier and Kurtz (1986). $\square$

The diffusion of Proposition 4.1 is a good approximation, in the sense of the limit in distribution as the population size tends to infinity, of a suitably rescaled version of the Wright–Fisher Markov chain described in Section 3. This is formalised by the next theorem. Here and later $\Rightarrow$ denotes convergence in distribution (or weak convergence) and $D_{\Delta_K}[0, \infty)$ denotes the space of *càdlàg* sample paths in Δ_K with the Skorokhod topology.

THEOREM 4.2.   *Let $\{Z_K^N(\tau), \tau = 0, 1, \ldots\}$ be the Δ_K^N-valued Markov chain with one-step transitions as in* (3.3)–(3.7), *let Z_K be the Feller diffusion of Proposition* 4.1. *If $Z_K^N(0) \Rightarrow Z_K(0)$, then*

$$Z_K^N(\lfloor N \cdot \rfloor) \Rightarrow Z_K(\cdot) \qquad \text{in } D_{\Delta_K}[0, \infty)$$

as $N \to \infty$.

PROOF.   From (3.5) and (3.8), letting $\mathbb{E}_z(\cdot) := \mathbb{E}(\cdot \mid z)$ and similarly for $\mathbb{P}_z$, we have that

$$\mathbb{E}_z[z_i' - z_i] = \mathbb{E}_z[z_i' - z_i^{**}] + z_i^{**} - z_i = N^{-1} b_i(z) + o(N^{-1})$$

and

$$\mathbb{E}_z[(z_i' - z_i)(z_j' - z_j)] = N^{-1} z_i^{**}(\delta_{ij} - z_j^{**}) + o(N^{-1}) = N^{-1} a_{ij}(z) + o(N^{-1}),$$

uniformly in z. Furthermore, it can be easily seen that $\mathbb{E}_z[(z_i' - z_i)^4] = o(N^{-1})$, so that Chebyshev's inequality implies Dynkin's condition for the continuity of paths of the limit process, that is, $\mathbb{P}_z(|z_i' - z_i| > \delta) = o(N^{-1})$ for every $\delta > 0$. Again, these estimates are uniform in z. Denote by $\mathcal{T}_K^N$ the semigroup operator associated to the Markov chain $Z_K^N(\cdot)$ and by I the identity operator. Then a Taylor expansion, together with the above expressions, yields, for every $f \in C^2(\Delta_K)$,

$$\begin{aligned}
(\mathcal{T}_K^N - I) f(z) &= \mathbb{E}_z[f(z') - f(z)] \\
&= \mathbb{E}_z\Bigg[\sum_{i=1}^K (z_i' - z_i) f_{z_i}(z) + \frac{1}{2} \sum_{i,j=1}^K (z_i' - z_i)(z_j' - z_j) f_{z_i z_j}(z) \\
&\qquad + \int_0^1 (1-t) \sum_{i,j=1}^K (z_i' - z_i)(z_j' - z_j) \\
&\qquad \times \big[f_{z_i z_j}(z + t(z' - z)) - f_{z_i z_j}(z) \big] dt \Bigg] \\
&= \frac{1}{N} \sum_{i=1}^K b_i(z) f_{z_i}(z) + \frac{1}{2N} \sum_{i,j=1}^K a_{ij}(z) f_{z_i z_j}(z) + o\Big(\frac{1}{N}\Big),
\end{aligned}$$

uniformly in z, where the $o(N^{-1})$ term above is due to

$$\left| \mathbb{E}_z\left[\int_0^1 (1-t) \sum_{i,j=1}^K (z_i' - z_i)(z_j' - z_j)[f_{z_i z_j}(z + t(z' - z)) - f_{z_i z_j}(z)]\, dt \right] \right|$$

$$\leq \mathbb{E}_z\left[\frac{1}{2} \sum_{i,j=1}^K |z_i' - z_i| |z_j' - z_j| 2\|f_{z_i z_j}\|; |z' - z| > \delta \right]$$

$$+ \mathbb{E}_z\left[\frac{1}{2} \sum_{i,j=1}^K |z_i' - z_i| |z_j' - z_j| \omega(f_{z_i z_j}, \delta); |z' - z| \leq \delta \right]$$

$$\leq \sum_{i,j=1}^K \|f_{z_i z_j}\| \mathbb{P}_z(|z' - z| > \delta)$$

$$+ \frac{1}{2} \sum_{i,j=1}^K \mathbb{E}_z[(z_i' - z_i)^2]^{1/2} \mathbb{E}_z[(z_j' - z_j)^2]^{1/2} \omega(f_{z_i z_j}, \delta)$$

$$= o(N^{-1}) + O(N^{-1}) \max_{i,j} \omega(f_{z_i z_j}, \delta),$$

$\omega(g, \delta) := \sup_{|z' - z| \leq \delta} |g(z') - g(z)|$ being the modulus of continuity of the function g. It follows that, for every $f \in C^2(\Delta_K)$,

$$(4.3) \qquad \|N(\mathcal{T}_K^N - I)f - \mathcal{A}_K f\| \to 0$$

as $N \to \infty$, where $\mathcal{A}_K$ is as in (4.1). An application of Theorems 1.6.5 and 4.2.6 in Ethier and Kurtz (1986) implies the statement of the theorem. $\quad\square$

Having justified our first limit operation, we now apply the descending order statistics to our limit Wright–Fisher diffusion $Z_K(\cdot)$. First, we define the continuous map $\rho_K : \Delta_K \mapsto \nabla_\infty$ by

$$\rho_K(z) := (z_{(1)}, \dots, z_{(K)}, 0, 0, \dots),$$

where $z_{(1)} \geq z_{(2)} \geq \cdots \geq z_{(K)}$ are the descending order statistics of the coordinates of $z \in \Delta_K$. We will show in the next section that, with suitable definitions of $p_i(z)$ and $r_i(z)$ and assuming convergence of the initial distributions, $\rho_K(Z_K(\cdot)) \Rightarrow Z(\cdot)$ as $K \to \infty$, with $Z(\cdot)$ denoting the two-parameter model in $\overline{\nabla}_\infty$.

Here, we simply observe that $\rho_K(Z_K(\cdot))$ is Markovian despite the fact that ρ_K is not one-to-one. The state space of $\rho_K(Z_K(\cdot))$ is

$$\nabla_K := \{z \in \nabla_\infty : z_{K+1} = 0\} \subset \overline{\nabla}_\infty$$

and its generator $\mathcal{B}_K$ is given by

$$(4.4) \qquad \mathcal{B}_K := \frac{1}{2} \sum_{i,j=1}^K a_{ij}(z) \frac{\partial^2}{\partial z_i\, \partial z_j} + \sum_{i=1}^K b_i(z) \frac{\partial}{\partial z_i},$$

ostensibly the same as $\mathcal{A}_K$ in (4.1)–(4.2) and (3.9), except that now $z \in \nabla_K$ instead of $z \in \Delta_K$. In addition,

$$(4.5) \qquad \mathscr{D}(\mathcal{B}_K) := \{ f \in C^2(\nabla_K) : f \circ \rho_K \in C^2(\Delta_K) \}.$$

Hidden in this definition are certain implicit boundary conditions needed to preserve the inequalities $z_1 \geq z_2 \geq \cdots \geq z_K$ [see Ethier and Kurtz (1981) for more details]. The following result generalises Proposition 2.4 of Ethier and Kurtz (1981).

PROPOSITION 4.3. *The closure in $C(\nabla_K)$ of the operator $\mathcal{B}_K$ defined by (4.4)–(4.5), (4.2) and (3.9) is single-valued and generates a Feller semigroup $\{\mathcal{U}_K(t)\}$ on $C(\nabla_K)$. Given $v \in \mathscr{P}(\Delta_K)$, let $Z_K(\cdot)$ be as in Proposition 4.1. Then $\rho_K(Z_K(\cdot))$ is a strong Markov process corresponding to $\{\mathcal{U}_K(t)\}$ with initial distribution $v_K \circ \rho_K^{-1}$ and almost all sample paths in $C_{\nabla_K}[0, \infty)$.*

PROOF.   The proof is exactly as in the cited paper, the key observation being that, for every permutation σ of $\{1, 2, \ldots, K\}$,

$$b_i(z_{\sigma(1)}, \ldots, z_{\sigma(K)}) = b_{\sigma(i)}(z), \qquad z \in \Delta_K, i = 1, 2, \ldots, K.$$

As a byproduct of this, we find that, if $f \in \mathscr{D}(\mathcal{B}_K)$,

$$(4.6) \qquad (\mathcal{B}_K f) \circ \rho_K = \mathcal{A}_K(f \circ \rho_K) \qquad \text{on } \Delta_K. \qquad \square$$

5. Convergence to the infinite-dimensional diffusion.   We now turn to our second limit operation, namely the convergence of the reordered Wright–Fisher diffusion $\rho_K(Z_K(\cdot))$ to the two-parameter model, that is, the $\overline{\nabla}_\infty$-valued diffusion process with generator $\mathcal{B}$ introduced in Section 2. To this end we will specify explicitly the functions p_i and r_i that determine the migration mechanism and provide some probabilistic interpretation of our choice, but the results of this section hold more generally (see Remark 5.7).

The drift coefficients of $\mathcal{B}$ are $-\frac{1}{2}(\theta z_i + \alpha)$, which we rewrite as in (2.7), while those of $\mathcal{B}_K$ are given by (3.9). In view of the comments at the end of Section 2, the functions p_i and r_i should satisfy

$$(5.1) \qquad -\frac{1}{2}\alpha(1 - z_i) = \lim_{K \to \infty} \frac{1}{2}\left[\alpha p_i(z) \sum_{j=1}^{K} z_j r_j(z) - \alpha z_i r_i(z) \right].$$

One way to achieve this is to take $r_i(z) = (1 - z_i)/z_i$ and $p_i(z) = o(1/K)$. However, this is problematic for two reasons. First, r_i is unbounded; second, requiring $p_i(z) = o(1/K)$ uniformly in i and z is inconsistent with $\sum_{i=1}^{K} p_i(z) = 1$. We can address both issues by instead defining

$$(5.2) \qquad r_i(z) := \begin{cases} (1 - z_i)\big[1 - (1 - z_i)^K\big]/z_i, & \text{if } z_i > 0, \\ K, & \text{if } z_i = 0, \end{cases}$$

and

$$(5.3) \qquad p_i(z) := \frac{(1 - z_i)^K}{\sum_{l=1}^{K}(1 - z_l)^K}.$$

An alternative formulation is in terms of the following system of Bernoulli trials parameterised by the current state z. Let the array $\zeta = (\zeta_{ij})_{i,j=1,\ldots,K}$ be such that, along row i, $\zeta_{i1}, \ldots, \zeta_{iK}$ are i.i.d. Bernoulli(z_i), for $i = 1, 2, \ldots, K$. With G_i being the number of failures in row i before the first success,

$$r_i(z) = \sum_{k=1}^{K}(1 - z_i)^k = \sum_{k=1}^{K} \mathbb{P}(G_i \geq k) = \mathbb{E}[G_i].$$

Furthermore, $p_i(z)$ is proportional to the probability of observing no successes in row i. Incidentally, $p_i(z)$ has also a direct probabilistic interpretation via Bayes's theorem. Let I be a row of the array chosen uniformly at random. Then $p_i(z)$ is the probability of choosing row i given that we observe all failures along the row, that is,

$$p_i(z) = \mathbb{P}\left\{ I = i \ \Big| \ \sum_{j=1}^{K} \zeta_{Ij} = 0 \right\}.$$

Let Z be the two-parameter model. In order to prove that $\rho_K(Z_K(\cdot)) \Rightarrow Z(\cdot)$ as $K \to \infty$, the usual argument is to show that

$$(5.4) \qquad \|\mathcal{B}_K \eta_K \varphi - \eta_K \mathcal{B}\varphi\| \to 0 \qquad \text{as } K \to \infty,$$

where $\eta_K : C(\overline{\nabla}_\infty) \mapsto C(\nabla_K)$ is given by the restriction $\eta_K \varphi = \varphi|_{\nabla_K}$, and $\varphi \in \mathscr{D}(\mathcal{B})$ is given by

$$\varphi = \varphi_{m_1} \cdots \varphi_{m_l}, \qquad m_1, \ldots, m_l \in \{2, 3, \ldots\}, l \in \mathbb{N}.$$

[Notice that η_K maps $\mathscr{D}(\mathcal{B})$ into $\mathscr{D}(\mathcal{B}_K)$.]

Unfortunately, despite the fact that (5.1) holds with this choice, (5.4) fails if one or more of the subscripts $m_1, \ldots, m_l$ is equal to 2. Similar to what was done in Ethier and Kurtz (1981) in the proof of Theorem 2.6, we can enlarge the domain of $\mathcal{B}$ to the algebra generated by 1 and the functions φ_m defined by (2.2) for all real $m \geq 2$ (not just integers). Then (5.4) holds for

$$\varphi = \varphi_{m_1} \cdots \varphi_{m_l}, \qquad m_1, \ldots, m_l > 2, l \in \mathbb{N}.$$

For example, if $\varphi = \varphi_{2+\varepsilon}$ for $0 < \varepsilon < 1$, then $\|\mathcal{B}_K \eta_K \varphi - \eta_K \mathcal{B}\varphi\| = O(K^{-\varepsilon})$. This would suffice if we could show that

$$\mathscr{D}_0(\mathcal{B}) := \text{sub-algebra of } C(\overline{\nabla}_\infty) \text{ generated by 1 and } \varphi_m, \qquad m \in (2, \infty),$$

is a core for the closure of $\mathcal{B}$ [cf. Ethier and Kurtz (1986), Section 1.3]. This also appears to fail. In fact, this algebra is not even a core in the bounded-pointwise sense, as

$$\underset{\varepsilon \to 0+}{\text{bp-lim}} \, \mathcal{B}\varphi_{2+\varepsilon}(z) = (1-\alpha)\sum_{i=1}^{\infty} z_i - (1+\theta)\varphi_2(z),$$

which is not equal to (2.4) except on ∇_{∞}.

As mentioned in Section 2, recently Ethier (2014) proved that, for any initial distribution ν concentrated on ∇_{∞}, the paths of Z belong to $C_{\nabla_{\infty}}[0, \infty)$ with probability one. In view of this result and of the above discussion, one might think of taking ∇_{∞} as state space, rather than $\overline{\nabla}_{\infty}$. But ∇_{∞} is not compact, therefore, the usual sufficient conditions for convergence in distribution include, besides (5.4), the following compact containment condition: For every $\varepsilon, T > 0$, there exists a compact set $\Gamma_{\varepsilon,T}$ such that

$$(5.5) \qquad \inf_K \mathbb{P}\big(\rho_K\big(Z_K(t)\big) \in \Gamma_{\varepsilon,T}, \forall t \leq T\big) \geq 1 - \varepsilon.$$

Notice that, since ∇_{∞} is not a complete metric space, convergence might hold without the compact containment condition [see, e.g., Billingsley (1968), Theorems 6.1 and 6.2]. In any case, (5.5) is not easy to prove and we have not pursued this approach.

A further alternative strategy would be to show that, for every $\psi \in \mathscr{D}(\mathcal{B})$, there exists a sequence $\{\psi_K\} \subset \mathscr{D}_0(\mathcal{B})$ such that

$$\|\eta_K \psi_K - \eta_K \psi\| \to 0 \quad \text{and} \quad \|\mathcal{B}_K \eta_K \psi_K - \eta_K \mathcal{B}\psi\| \to 0,$$

as $K \to \infty$, so that $\{(\psi, \mathcal{B}\psi) : \psi \in \mathscr{D}(\mathcal{B})\}$ belongs to the extended limit of $\mathcal{B}_K$ [cf. Definition 1.4.3 of Ethier and Kurtz (1986)]. Then Theorem 1.6.1 of Ethier and Kurtz (1986) would yield

$$\big\|\mathcal{T}_K(t)\eta_K \varphi - \eta_K \mathcal{T}(t)\varphi\big\| \to 0, \qquad \varphi \in C(\overline{\nabla}_{\infty}), t \geq 0,$$

where $\{\mathcal{T}(t)\}$ is the Feller semigroup on $C(\overline{\nabla}_{\infty})$ whose generator is the closure of $\mathcal{B}$. Even this strategy seems not to be viable.

Having considered each of these routes, we have turned to the martingale problem approach. In this approach, $\rho_K(Z_K(\cdot))$ is viewed as a solution to the martingale problem for $\mathcal{B}_K$ (in fact the unique solution). The usual procedure consists of three steps: (i) Show that $\{\rho_K(Z_K(\cdot))\}$ is relatively compact; (ii) Show that each of its limit points is a solution to the martingale problem for $\mathcal{B}$; (iii) Show that the martingale problem for $\mathcal{B}$ has a unique solution.

However, in the present setup, it is not clear how to carry out the second and third steps. In fact, if $\mathscr{D}(\mathcal{B})$ is taken as the domain of $\mathcal{B}$, then it is not clear that the limit martingale relation will hold for φ_2 (and any product in which φ_2 is a factor) because $\|\mathcal{B}_K \eta_K \varphi_2 - \eta_K \mathcal{B}\varphi_2\|$ does not converge to zero, as outlined above. On the other hand, if $\mathscr{D}_0(\mathcal{B})$ is taken as the domain of $\mathcal{B}$, then the martingale problem for

$\mathcal{B}$ may have more than one solution. For instance, if the initial distribution is the unit mass at $z = 0$, then the identically zero stochastic process is a solution.

We solve these problems by proving a priori that, for any limit point Z of $\{\rho_K(Z_K(\cdot))\}$, with probability one, $Z(t) \in \nabla_\infty$ for almost all $t \geq 0$. This is done in Lemma 5.3 below. On ∇_∞, $\mathcal{B}\varphi_2$ can be approximated by $\mathcal{B}\varphi_{2+\varepsilon}$, for $\varepsilon \to 0+$, and this yields that the limit martingale relation, which holds for functions in $\mathcal{D}_0(\mathcal{B})$, carries over to all functions in $\mathcal{D}(\mathcal{B})$, and thus that the limit martingale problem has a unique solution (Theorem 5.6).

LEMMA 5.1. $\{\rho_K(Z_K(\cdot))\}$ is relatively compact in $D_{\overline{\nabla}_\infty}[0, \infty)$.

PROOF.   By Proposition 4.3, $\rho_K(Z_K(\cdot))$ is a strong Markov process with generator the closure of $\mathcal{B}_K$ and sample paths in $C_{\nabla_K}[0, \infty)$. Therefore, $\rho_K(Z_K(\cdot))$ is a solution of the martingale problem for $\mathcal{B}_K$ [see, e.g., Proposition 4.1.7 in Ethier and Kurtz (1986)].

We have, for $m, K \in \{2, 3, \ldots\}$ and $z \in \nabla_K$,

$$\mathcal{B}_K \eta_K \varphi_m(z)$$

$$= \binom{m}{2}(\varphi_{m-1} - \varphi_m)(z)$$

$$+ \frac{m}{2}\left\{\frac{\theta + \alpha}{K-1}(\varphi_{m-1} - \varphi_m)(z) - \theta\varphi_m(z)\right.$$

$$+ \alpha\sum_{j=1}^{K} z_j r_j(z) \sum_{i=1}^{K} p_i(z)z_i^{m-1} - \alpha\sum_{i=1}^{K}[z_i + z_i r_i(z)]z_i^{m-1}\right\}$$

$$= \binom{m}{2}(\varphi_{m-1} - \varphi_m)(z)$$

$$+ \frac{m}{2}\frac{\theta + \alpha}{K-1}(\varphi_{m-1} - \varphi_m)(z) - \frac{m}{2}(\theta\varphi_m + \alpha\varphi_{m-1})(z)$$

$$(5.6) \qquad + \frac{m}{2}\alpha\left\{\sum_{i=1}^{K}[1 - z_i - z_i r_i(z)]z_i^{m-1} + \sum_{j=1}^{K} z_j r_j(z) \sum_{i=1}^{K} p_i(z)z_i^{m-1}\right\}$$

$$= \binom{m}{2}(\varphi_{m-1} - \varphi_m)(z)$$

$$+ \frac{m}{2}\frac{\theta + \alpha}{K-1}(\varphi_{m-1} - \varphi_m)(z) - \frac{m}{2}(\theta\varphi_m + \alpha\varphi_{m-1})(z)$$

$$+ \frac{m}{2}\alpha\sum_{i=1}^{K}(1 - z_i)^{K+1}z_i^{m-1} + \frac{m}{2}\alpha\sum_{j=1}^{K} z_j r_j(z) \sum_{i=1}^{K} p_i(z)z_i^{m-1}$$

$$= \binom{m}{2}(\varphi_{m-1} - \varphi_m)(z)$$

$$+ \frac{m}{2}\frac{\theta + \alpha}{K-1}(\varphi_{m-1} - \varphi_m)(z) - \frac{m}{2}(\theta\varphi_m + \alpha\varphi_{m-1})(z)$$

$$+ \frac{m}{2}\alpha \sum_{i=1}^{K}(1 - z_i)^K z_i^{m-1}\left(1 - z_i + \frac{\sum_{j=1}^{K} z_j r_j(z)}{\sum_{l=1}^{K}(1 - z_l)^K}\right),$$

where the third equality uses (5.2) and the fourth uses (5.3).

Now, since $z_i \leq 1/i$ for $i = 1, \ldots, K$, we have

$$\sum_{l=1}^{K}(1 - z_l)^K \geq \sum_{l=\lfloor K/2 \rfloor + 1}^{K}(1 - z_l)^K$$

$$\geq \lceil K/2 \rceil \left(1 - \frac{1}{\lfloor K/2 \rfloor + 1}\right)^K$$

$$\geq (K/2)e^{-2},$$

so that

$$(5.7) \qquad \frac{\sum_{j=1}^{K} z_j r_j(z)}{\sum_{l=1}^{K}(1 - z_l)^K} \leq \frac{K}{(K/2)e^{-2}} = 2e^2.$$

In addition,

$$(5.8) \qquad \sum_{i=1}^{K}(1 - z_i)^K z_i^{m-1} \leq K \sup_{0 \leq u \leq 1}(1 - u)^K u^{m-1} \leq K\left(\frac{m-1}{K+m-1}\right)^{m-1}.$$

Therefore, for each integer $m \geq 2$,

$$(5.9) \qquad \sup_{K}\|\mathcal{B}_K \eta_K \varphi_m\| \leq C(\alpha, \theta, \varphi_m).$$

For $\varphi, \psi \in \mathscr{D}(\mathcal{B})$, we can use the analogue of the first equation in (2.13) of Ethier and Kurtz (1981), namely

$$(5.10) \qquad \begin{aligned} \mathcal{B}_K \eta_K(\varphi\psi) &= (\eta_K\psi)\mathcal{B}_K\eta_K\varphi + (\eta_K\varphi)\mathcal{B}_K\eta_K\psi \\ &\quad + \langle \mathrm{grad}(\eta_K\varphi), a\,\mathrm{grad}(\eta_K\psi)\rangle, \end{aligned}$$

where a is given by (4.2), to obtain

$$\|B_K\eta_K(\varphi\psi)\| \leq \|\psi\|\|B_K\eta_K\varphi\| + \|\varphi\|\|B_K\eta_K\psi\| + 2\sup_{i\geq 1}\|\varphi_{z_i}\|\sup_{j\geq 1}\|\psi_{z_j}\|.$$

Then we can see, by induction on l, that (5.9) holds with φ_m replaced by φ of the form $\varphi = \varphi_{m_1}\varphi_{m_2}\cdots\varphi_{m_l}$, $m_1, \ldots, m_l \in \{2, 3, \ldots\}$, $l \in \mathbb{N}$, hence for every $\varphi \in \mathscr{D}(\mathcal{B})$.

Since $\mathscr{D}(\mathcal{B})$ is dense in $C(\overline{\nabla}_\infty)$, the lemma follows from Theorems 3.9.1 and 3.9.4 of Ethier and Kurtz (1986). □

LEMMA 5.2.   *For $2 < m < 3$ and $K \geq 2$,*

$$\mathcal{B}_K \eta_K (\varphi_2 - \varphi_m) \geq 1 - \alpha - \frac{m(m-1-\alpha)}{2}\varphi_{m-1}$$

$$- \left[(1+\theta)\varphi_2 - \frac{m(m-1+\theta)}{2}\varphi_m \right]$$

$$- \left[\frac{3(\theta+\alpha)}{2(K-1)} + \frac{\alpha(1+2e^2)}{2(K+1)} \right] \qquad on \, \nabla_K.$$

PROOF.   Let $2 < m < 3$ and $K \geq 2$. We have, on ∇_K,

$$\mathcal{B}_K \eta_K (\varphi_2 - \varphi_m)$$

$$= 1 - \alpha - \frac{m(m-1-\alpha)}{2}\varphi_{m-1}$$

$$- \left[(1+\theta)\varphi_2 - \frac{m(m-1+\theta)}{2}\varphi_m \right]$$

$$+ \frac{\theta+\alpha}{K-1}\left[1 - \frac{m}{2}\varphi_{m-1} \right] - \frac{\theta+\alpha}{K-1}\left[\varphi_2 - \frac{m}{2}\varphi_m \right] + \alpha R_{K,m},$$

where

$$R_{K,m}(z) := \sum_{i=1}^{K}[1 - z_i - z_i r_i(z)]z_i \left(1 - \frac{m}{2}z_i^{m-2} \right)$$

$$+ \sum_{j=1}^{K} z_j r_j(z) \sum_{i=1}^{K} p_i(z)z_i \left(1 - \frac{m}{2}z_i^{m-2} \right).$$

Since $z_i \leq 1/i$ for $i = 1, \ldots, K$, we obtain the inequalities

$$1 - \frac{m}{2}z_i^{m-2} \geq 0, \qquad i \geq 2, \qquad 1 - \frac{m}{2}z_1^{m-2} \geq -\frac{1}{2},$$

and hence

$$R_{K,m}(z) \geq -\frac{1}{2}\left[(1-z_1)^{K+1}z_1 + \sum_{j=1}^{K} z_j r_j(z) p_1(z)z_1 \right].$$

In addition, by (5.3) and (5.7),

$$\sum_{j=1}^{K} z_j r_j(z) p_1(z) = \frac{\sum_{j=1}^{K} z_j r_j(z)}{\sum_{l=1}^{K}(1-z_l)^K}(1-z_1)^K \leq 2e^2(1-z_1)^K.$$

Then, by the second inequality in (5.8), we get $R_{K,m}(z) \geq -(1+2e^2)/(2(K+1))$. Notice also that $1 - \frac{m}{2}\varphi_{m-1}(z) \geq -\frac{1}{2}$ and that $\varphi_2(z) - \frac{m}{2}\varphi_m(z) \leq 1$. The conclusion follows.   □

LEMMA 5.3.   *For every limit point Z of $\{\rho_K(Z_K(\cdot))\}$ in $D_{\overline{\nabla}_\infty}[0, \infty)$, we have*

$$\mathbb{E}\left[\int_0^\infty \left(1 - \sum_{i=1}^\infty Z_i(t)\right) dt\right] = 0.$$

PROOF.   The proof is inspired by the first part of the proof of Theorem 2.6 in Ethier and Kurtz (1981). As $\rho_K(Z_K(\cdot))$ is a solution of the martingale problem for $\mathcal{B}_K$, Lemma 5.2 implies that, for $2 < m < 3$ and $K \geq 2$,

$$
\begin{aligned}
\mathbb{E}&[(\varphi_2 - \varphi_m)(\rho_K(Z_K(T)))] \\
&\geq \mathbb{E}[(\varphi_2 - \varphi_m)(\rho_K(Z_K(0)))] \\
&\quad + \mathbb{E}\left[\int_0^T \left(1 - \alpha - \frac{m(m-1-\alpha)}{2}\varphi_{m-1}(\rho_K(Z_K(t)))\right) dt\right] \\
&\quad - \mathbb{E}\left[\int_0^T \left((1+\theta)\varphi_2 - \frac{m(m-1+\theta)}{2}\varphi_m\right)(\rho_K(Z_K(t))) \, dt\right] \\
&\quad - \left[\frac{3(\theta+\alpha)}{2(K-1)} + \frac{\alpha(1+2e^2)}{2(K+1)}\right]T.
\end{aligned}
$$
(5.11)

Let Z be the limit in distribution of some subsequence $\{\rho_{K_h}(Z_{K_h})\}$. Since $\varphi_2, \varphi_m, \varphi_{m-1}$ are continuous and all integrands are bounded, by taking the limit as $h \to \infty$ along the subsequence $\{K_h\}$ in (5.11), we obtain

$$
\begin{aligned}
(1-\alpha)&\mathbb{E}\left[\int_0^T \left(1 - \frac{m(m-1-\alpha)}{2(1-\alpha)}\varphi_{m-1}(Z(t))\right) dt\right] \\
&\leq \mathbb{E}[(\varphi_2 - \varphi_m)(Z(T))] - \mathbb{E}[(\varphi_2 - \varphi_m)(Z(0))] \\
&\quad + (1+\theta)\mathbb{E}\left[\int_0^T \left(\varphi_2 - \frac{m(m-1+\theta)}{2(1+\theta)}\varphi_m\right)(Z(t)) \, dt\right].
\end{aligned}
$$
(5.12)

Since $\varphi_{m-1}(z)$ converges to $\sum_{i=1}^\infty z_i$ boundedly and pointwise on $\overline{\nabla}_\infty$, we obtain the assertion by taking the limit as $m \to 2+$ in (5.12). $\square$

LEMMA 5.4.   *For $0 < \varepsilon < 1$, let $\varphi = \varphi_{m_1} \cdots \varphi_{m_l}$, where $m_1, \ldots, m_l \in [2 + \varepsilon, \infty)$. Then*

$$\|\mathcal{B}_K \eta_K \varphi - \eta_K \mathcal{B}\varphi\| = O(K^{-\varepsilon}) \qquad \text{as } K \to \infty.$$

PROOF.   Consider first $\varphi = \varphi_m$ with $m \geq 2 + \varepsilon$. Then

$$\mathcal{B}\varphi_m = \binom{m}{2}(\varphi_{m-1} - \varphi_m) - \frac{m}{2}(\theta\varphi_m + \alpha\varphi_{m-1}).$$

Recalling (5.6)–(5.8), we have

$$\|\mathcal{B}_K \eta_K \varphi_m - \eta_K \mathcal{B} \varphi_m\|$$

$$\leq \frac{m}{2} \frac{\theta + \alpha}{K - 1} \sup_{z \in \nabla_K} |(\varphi_{m-1} - \varphi_m)(z)|$$

$$+ \sup_{z \in \nabla_K} \frac{m}{2} \alpha \sum_{i=1}^{K} (1 - z_i)^K z_i^{m-1} \left(1 - z_i + \frac{\sum_{j=1}^{K} z_j r_j(z)}{\sum_{l=1}^{K} (1 - z_l)^K} \right)$$

$$\leq \frac{m}{2} \frac{\theta + \alpha}{K - 1} + \frac{m}{2} \alpha K \left(\frac{m - 1}{K + m - 1} \right)^{m-1} (1 + 2e^2) = O(K^{-\varepsilon}),$$

as required.

By an analogue of the first equation in (2.13) of Ethier and Kurtz (1981), namely

$$(5.13) \qquad \mathcal{B}(\varphi \psi) = \psi \mathcal{B} \varphi + \varphi \mathcal{B} \psi + \langle \operatorname{grad} \varphi, a \operatorname{grad} \psi \rangle,$$

we get, by (5.10),

$$\|\mathcal{B}_K \eta_K (\varphi \psi) - \eta_K \mathcal{B}(\varphi \psi)\|$$

$$\leq \|\psi\| \|\mathcal{B}_K \eta_K \varphi - \eta_K \mathcal{B} \varphi\| + \|\varphi\| \|\mathcal{B}_K \eta_K \psi - \eta_K \mathcal{B} \psi\|.$$

Thus, the statement of the lemma follows by induction on l. $\quad\square$

LEMMA 5.5. *Let $\varphi \in \mathcal{D}_0(\mathcal{B})$ and $p \in \mathbb{N}$. Then $\mathcal{B}(\varphi_{2+\varepsilon}^p \varphi) \to \mathcal{B}(\varphi_2^p \varphi)$ boundedly and pointwise on ∇_∞.*

PROOF. By (5.13),

$$\mathcal{B}(\varphi_{2+\varepsilon}^p \varphi) = \varphi \mathcal{B} \varphi_{2+\varepsilon}^p + \varphi_{2+\varepsilon}^p \mathcal{B} \varphi + \langle \operatorname{grad} \varphi_{2+\varepsilon}^p, a \operatorname{grad} \varphi \rangle$$

$$= \varphi \left[p \varphi_{2+\varepsilon}^{p-1} \mathcal{B} \varphi_{2+\varepsilon} + \binom{p}{2} \varphi_{2+\varepsilon}^{p-2} \langle \operatorname{grad} \varphi_{2+\varepsilon}, a \operatorname{grad} \varphi_{2+\varepsilon} \rangle \right]$$

$$+ \varphi_{2+\varepsilon}^p \mathcal{B} \varphi + p \varphi_{2+\varepsilon}^{p-1} \langle \operatorname{grad} \varphi_{2+\varepsilon}, a \operatorname{grad} \varphi \rangle$$

$$\to \varphi \left[p \varphi_2^{p-1} \mathcal{B} \varphi_{2+} + \binom{p}{2} \varphi_2^{p-2} \langle \operatorname{grad} \varphi_2, a \operatorname{grad} \varphi_2 \rangle \right]$$

$$+ \varphi_2^p \mathcal{B} \varphi + p \varphi_2^{p-1} \langle \operatorname{grad} \varphi_2, a \operatorname{grad} \varphi \rangle$$

boundedly and pointwise on $\overline{\nabla}_\infty$ as ε goes to zero, where

$$\mathcal{B} \varphi_{2+}(z) := \lim_{\varepsilon \to 0} \mathcal{B} \varphi_{2+\varepsilon}(z) = (1 - \alpha) \sum_{i=1}^{\infty} z_i - (1 + \theta) \varphi_2(z), \qquad z \in \overline{\nabla}_\infty.$$

We are also using

$$\langle \operatorname{grad} \varphi_{2+\varepsilon}, a \operatorname{grad} \varphi_{2+\varepsilon} \rangle = (2+\varepsilon)^2 \big(\varphi_{3+2\varepsilon} - \varphi_{2+\varepsilon}^2\big)$$

$$\rightarrow 4\big(\varphi_3 - \varphi_2^2\big) = \langle \operatorname{grad} \varphi_2, a \operatorname{grad} \varphi_2 \rangle$$

and similarly $\langle \operatorname{grad} \varphi_{2+\varepsilon}, a \operatorname{grad} \varphi \rangle \rightarrow \langle \operatorname{grad} \varphi_2, a \operatorname{grad} \varphi \rangle$, both boundedly and pointwise on $\overline{\nabla}_\infty$. Of course,

$$\mathcal{B}\varphi_2(z) = 1 - \alpha - (1+\theta)\varphi_2(z), \qquad z \in \overline{\nabla}_\infty,$$

so $\mathcal{B}\varphi_{2+} = \mathcal{B}\varphi_2$ on ∇_∞. We conclude that

$$\mathcal{B}\big(\varphi_{2+\varepsilon}^p \varphi\big) \rightarrow \varphi \mathcal{B}\varphi_2^p + \varphi_2^p \mathcal{B}\varphi + \langle \operatorname{grad} \varphi_2^p, a \operatorname{grad} \varphi \rangle = \mathcal{B}\big(\varphi_2^p \varphi\big)$$

boundedly and pointwise on ∇_∞ (but not on $\overline{\nabla}_\infty$). $\square$

We are now ready to state our main result.

THEOREM 5.6.   *Let Z_K be the diffusion process of Proposition* 4.1 *with initial distribution $\nu_K \in \mathscr{P}(\Delta_K)$. Let $\mathcal{B}$ be given by* (2.1)–(2.3) *and let Z be the diffusion process corresponding to the Feller semigroup generated by the closure in $C(\overline{\nabla}_\infty)$ of $\mathcal{B}$, with initial distribution $\nu \in \mathscr{P}(\overline{\nabla}_\infty)$. If $\nu_K \circ \rho_K^{-1} \Rightarrow \nu$, then*

$$\rho_K\big(Z_K(\cdot)\big) \Rightarrow Z(\cdot) \qquad in\ C_{\overline{\nabla}_\infty}[0, \infty).$$

If in addition $\nu(\nabla_\infty) = 1$, then the convergence holds in $C_{\nabla_\infty}[0, \infty)$.

PROOF.   First, we prove convergence in $D_{\overline{\nabla}_\infty}[0, \infty)$. The proof of this claim is in three steps:

 (i) Every limit point of $\{\rho_K(Z_K(\cdot))\}$ is a solution of the martingale problem for $\mathcal{B}$ as an operator on $\mathscr{D}_0(\mathcal{B})$.
 (ii) Every limit point of $\{\rho_K(Z_K(\cdot))\}$ is a solution of the martingale problem for $\mathcal{B}$ as an operator on $\mathscr{D}(\mathcal{B})$.
 (iii) The martingale problem for $\mathcal{B}$ as an operator on $\mathscr{D}(\mathcal{B})$ has a unique solution for every initial distribution ν.

Proof of (i). Let Z be the limit in distribution of an arbitrary subsequence $\{\rho_{K_h}(Z_{K_h}(\cdot))\}$; see Lemma 5.1. Since $\rho_{K_h}(Z_{K_h}(\cdot))$ is a solution of the martingale problem for $\mathcal{B}_{K_h}$,

$$\varphi\big(\rho_{K_h}(Z_{K_h}(t))\big) - \int_0^t \mathcal{B}_{K_h} \eta_{K_h} \varphi\big(\rho_{K_h}(Z_{K_h}(s))\big)\, ds =: M_\varphi^{(K_h)}(t)$$

is a continuous martingale for every $\varphi \in \mathscr{D}_0(\mathcal{B})$.

By Lemma 5.4, $M_\varphi^{(K_h)}$ converges in distribution to

$$\varphi\big(Z(\cdot)\big) - \int_0^{\cdot} \mathcal{B}\varphi\big(Z(s)\big)\, ds.$$

On the other hand, $M_\varphi^{(K_h)}(t)$ is uniformly bounded for every t, hence the limit is a martingale.

Proof of (ii). It is enough to prove that

$$\big(\varphi_2^p\varphi\big)(Z(\cdot)) - \int_0^\cdot \mathcal{B}\big(\varphi_2^p\varphi\big)(Z(s))\, ds =: M(\cdot)$$

is a martingale for every $\varphi \in \mathscr{D}_0(\mathcal{B})$ and every $p \in \mathbb{N}$.

By Lemma 5.3, almost surely we have

(5.14) $$\int_0^\infty \big(1 - \mathbb{I}_{\nabla_\infty}(Z(s))\big)\, ds = 0.$$

Therefore, by Step 1, for every $\varepsilon > 0$,

$$\big(\varphi_{2+\varepsilon}^p\varphi\big)(Z(t)) - \int_0^t \mathbb{I}_{\nabla_\infty}(Z(s))\mathcal{B}\big(\varphi_{2+\varepsilon}^p\varphi\big)(Z(s))\, ds = M_\varepsilon(t)$$

is a martingale.

It follows from Lemma 5.5 that, almost surely, for all $t \geq 0$, $M_\varepsilon(t)$ converges to

$$\big(\varphi_2^p\varphi\big)(Z(t)) - \int_0^t \mathbb{I}_{\nabla_\infty}(Z(s))\mathcal{B}\big(\varphi_2^p\varphi\big)(Z(s))\, ds,$$

which in turn, by (5.14), almost surely, for all $t \geq 0$, equals $M(t)$. On the other hand, for every $t \geq 0$, $M_\varepsilon(t)$ is uniformly bounded, hence M is a martingale.

Proof of (iii). A sufficient condition for uniqueness of the solution to the martingale problem for $\mathcal{B}$ is that, for each $\lambda > 0$, $\mathcal{R}(\lambda I - \mathcal{B})$, where $\mathcal{R}$ denotes the range and I is the identity operator, is separating, that is, such that, for any pair of probability measures $\mu, \nu \in \mathscr{P}(\overline{\nabla}_\infty)$, $\int_{\overline{\nabla}_\infty} f(z)\mu(dz) = \int_{\overline{\nabla}_\infty} f(z)\nu(dz)$ for every $f \in \mathcal{R}(\lambda I - \mathcal{B})$ implies $\mu = \nu$ [see, e.g., Costantini and Kurtz (2015), Corollary 2.14]. In the present setup, since the closure of $\mathcal{B}$ generates a strongly continuous contraction semigroup on $C(\overline{\nabla}_\infty)$ by Petrov (2009), then, for each $\lambda > 0$, $\mathcal{R}(\lambda I - \mathcal{B})$ is dense in $C(\overline{\nabla}_\infty)$ [see, e.g., Proposition 1.2.1 in Ethier and Kurtz (1986)], therefore, the condition is satisfied.

Finally, the convergence holds in $C_{\overline{\nabla}_\infty}[0, \infty) \subset D_{\overline{\nabla}_\infty}[0, \infty)$ because the distributions of the processes $\rho_K(Z_K(\cdot))$ and $Z(\cdot)$ are concentrated on $C_{\overline{\nabla}_\infty}[0, \infty)$ and the Skorokhod topology relativised to $C_{\overline{\nabla}_\infty}[0, \infty)$ coincides with the uniform-on-compact-sets topology on $C_{\overline{\nabla}_\infty}[0, \infty)$ [see, e.g., Billingsley (1968), Section 18]. The last assertion of the theorem follows from Ethier (2014) by the same argument. $\square$

REMARK 5.7.   A more careful inspection of the proofs shows that, if the mutation rates are given by (3.7), all the results of this section hold for functions p_i and r_i satisfying the conditions of Section 3 [in particular, $r_i(z) = r(z_i)$] and the following set of conditions: For Lemma 5.1, we need only assume

$$\sup_{z \in \nabla_K} \sum_{j=1}^K z_j r_j(z) \sum_{i=1}^K p_i(z)z_i = O(1) \qquad \text{as } K \to \infty.$$

For Lemma 5.2 with a possibly weaker but still adequate lower bound, it suffices that

$$1 - u - ur(u) \geq 0, \qquad u \in [0, 1],$$

$$\sup_{u \in [0,1]} [1 - u - ur(u)]u = o(1) \qquad \text{as } K \to \infty$$

and

$$\sup_{z \in \nabla_K} \sum_{j=1}^{K} z_j r_j(z) p_1(z) z_1 = o(1) \qquad \text{as } K \to \infty.$$

For Lemma 5.4 with a possibly slower but still adequate rate of convergence, it is enough that

$$\sup_{z \in \nabla_K} \sum_{i=1}^{K} [1 - z_i - z_i r(z_i)] z_i^{1+\varepsilon} = o(1) \qquad \text{as } K \to \infty, 0 < \varepsilon < 1$$

and

$$\sup_{z \in \nabla_K} \sum_{j=1}^{K} z_j r_j(z) \sum_{i=1}^{K} p_i(z) z_i^{1+\varepsilon} = o(1) \qquad \text{as } K \to \infty, 0 < \varepsilon < 1.$$

6. Convergence of stationary distributions. We have seen that, for each $K \geq 2$, $Z_K^N(\lfloor N \cdot \rfloor) \Rightarrow Z_K(\cdot)$ as $N \to \infty$ (Theorem 4.2) and $\rho_K(Z_K(\cdot)) \Rightarrow Z(\cdot)$ as $K \to \infty$ (Theorem 5.6). Now we want to obtain the analogous results for the stationary distributions. Our Wright–Fisher Markov chain model is irreducible and aperiodic and, therefore, has a unique stationary distribution $\mu_K^N \in \mathscr{P}(\Delta_K^N)$, which we regard as belonging to $\mathscr{P}(\Delta_K)$. Our K-dimensional diffusion process Z_K in Δ_K is ergodic by Theorem 3.2 of Shiga (1981) and, therefore, has a unique stationary distribution $\mu_K \in \mathscr{P}(\Delta_K)$. Technically, Shiga's theorem does not apply to our model because, although our drift coefficients due to mutation meet his Condition II, our drift coefficients due to migration,

$$b_i(z) := \alpha p_i(z) \sum_{j=1}^{K} z_j r_j(z) - \alpha z_i r_i(z),$$

are not of the form of his drift coefficients due to selection,

$$b_i(z) := z_i \left(\gamma_i(z) - \sum_{j=1}^{K} z_j \gamma_j(z) \right).$$

Nevertheless, our drift coefficients due to migration do satisfy (3.10) and (3.11), which together with smoothness is all that is needed for Shiga's proof. Finally, we denote by $\mathrm{PD}(\theta, \alpha) \in \mathscr{P}(\overline{\nabla}_\infty)$ the two-parameter Poisson–Dirichlet distribution,

which is the unique stationary distribution of Z in $\overline{\nabla}_\infty$. We will prove that, for each $K \geq 2$, $\mu_K^N \Rightarrow \mu_K$ on Δ_K as $N \to \infty$, and that $\mu_K \circ \rho_K^{-1} \Rightarrow \mathrm{PD}(\theta, \alpha)$ on $\overline{\nabla}_\infty$ as $K \to \infty$. This is the two-parameter analogue of Kingman's result showing that the one-parameter Poisson–Dirichlet distribution $\mathrm{PD}(\theta)$ is the weak limit of the descending order statistics of the symmetric Dirichlet distribution with parameter $\theta/(K-1)$. It is not entirely analogous in that the symmetric Dirichlet distribution with parameter $\theta/(K-1)$ is much more explicit than μ_K. Nevertheless, it does allow us to give an interpretation to $\mathrm{PD}(\theta, \alpha)$ in the context of population genetics.

THEOREM 6.1.   *For each $K \geq 2$, $\mu_K^N \Rightarrow \mu_K$ on Δ_K as $N \to \infty$.*

PROOF.   For fixed $K \geq 2$, $\{\mu_K^N\}$ is relatively compact because Δ_K is compact. It is enough to show that, if $\{N_m\}$ is a subsequence such that $\mu_K^{N_m} \Rightarrow \mu$ as $m \to \infty$, then $\mu = \mu_K$. Given $f \in C^2(\Delta_K)$,

$$\int_{\Delta_K} \mathcal{A}_K f \, d\mu = \lim_{m \to \infty} \int_{\Delta_K} \mathcal{A}_K f \, d\mu_K^{N_m}$$

$$= \lim_{m \to \infty} \int_{\Delta_K} N_m (\mathcal{T}_K^{N_m} - I) f \, d\mu_K^{N_m} = 0,$$

where the second equality uses (4.3). This shows that μ is the unique stationary distribution of Z_K, which we have denoted by μ_K. $\square$

THEOREM 6.2. $\mu_K \circ \rho_K^{-1} \Rightarrow \mathrm{PD}(\theta, \alpha)$ *on $\overline{\nabla}_\infty$ as $K \to \infty$.*

PROOF. $\{\mu_K \circ \rho_K^{-1}\}$ is relatively compact because $\overline{\nabla}_\infty$ is compact. It is enough to show that, if $\{K_h\}$ is a subsequence such that $\mu_{K_h} \circ \rho_{K_h}^{-1} \Rightarrow \mu$ on $\overline{\nabla}_\infty$ as $h \to \infty$, then $\mu = \mathrm{PD}(\theta, \alpha)$. First, we show that μ is concentrated on ∇_∞. It is intuitively clear and easy to prove that $\mu_K \circ \rho_K^{-1}$, which belongs to $\mathscr{P}(\nabla_K)$ but can also be regarded as belonging to $\mathscr{P}(\overline{\nabla}_\infty)$, is the unique stationary distribution of $\rho_K(Z_K(\cdot))$; indeed,

$$\int_{\nabla_K} \mathcal{B}_K f \, d(\mu_K \circ \rho_K^{-1}) = \int_{\Delta_K} (\mathcal{B}_K f) \circ \rho_K \, d\mu_K$$

$$= \int_{\Delta_K} \mathcal{A}_K (f \circ \rho_K) \, d\mu_K = 0,$$

provided $f \in \mathscr{D}(\mathcal{B}_K)$. Here, we have used (4.6). Lemma 5.2 therefore implies that, for $2 < m < 3$,

$$0 = \int_{\nabla_{K_h}} \mathcal{B}_{K_h} \eta_{K_h} (\varphi_2 - \varphi_m) \, d(\mu_{K_h} \circ \rho_{K_h}^{-1})$$

$$\geq \int_{\overline{\nabla}_\infty} \left[\left(1 - \alpha - \frac{m(m-1-\alpha)}{2} \varphi_{m-1} \right) \right.$$

$$-\left((1+\theta)\varphi_2 - \frac{m(m-1+\theta)}{2}\varphi_m\right)\right]d\left(\mu_{K_h}\circ\rho_{K_h}^{-1}\right)$$

$$-\left[\frac{3(\theta+\alpha)}{2(K_h-1)}+\frac{\alpha(1+2e^2)}{2(K_h+1)}\right]$$

$$\rightarrow\int_{\overline{\nabla}_\infty}\left[\left(1-\alpha-\frac{m(m-1-\alpha)}{2}\varphi_{m-1}\right)\right.$$

$$\left.-\left((1+\theta)\varphi_2 - \frac{m(m-1+\theta)}{2}\varphi_m\right)\right]d\mu.$$

In particular, this last integral is nonpositive. Now let $m\rightarrow 2+$ to conclude that

$$(1-\alpha)\int_{\overline{\nabla}_\infty}\left(1-\sum_{i=1}^\infty z_i\right)\mu(dz)\leq 0,$$

or that $\mu(\nabla_\infty)=1$.

Next, from Lemma 5.4 and $\mu_{K_h}\circ\rho_{K_h}^{-1}\Rightarrow\mu$ we get

$$0=\lim_{h\rightarrow\infty}\int_{\nabla_{K_h}}\mathcal{B}_{K_h}\eta_{K_h}\varphi\,d\left(\mu_{K_h}\circ\rho_{K_h}^{-1}\right)$$

$$=\lim_{h\rightarrow\infty}\int_{\overline{\nabla}_\infty}\mathcal{B}\varphi\,d\left(\mu_{K_h}\circ\rho_{K_h}^{-1}\right)=\int_{\overline{\nabla}_\infty}\mathcal{B}\varphi\,d\mu$$

for all $\varphi\in\mathscr{D}_0(\mathcal{B})$. Let $\varphi\in\mathscr{D}_0(\mathcal{B})$ and $p\in\mathbb{N}$. Then from Lemma 5.5, we have

$$0=\lim_{\varepsilon\rightarrow 0+}\int_{\overline{\nabla}_\infty}\mathcal{B}(\varphi_{2+\varepsilon}^p\varphi)\,d\mu=\lim_{\varepsilon\rightarrow 0+}\int_{\nabla_\infty}\mathcal{B}(\varphi_{2+\varepsilon}^p\varphi)\,d\mu$$

$$=\int_{\nabla_\infty}\mathcal{B}(\varphi_2^p\varphi)\,d\mu=\int_{\overline{\nabla}_\infty}\mathcal{B}(\varphi_2^p\varphi)\,d\mu,$$

implying that $\int_{\overline{\nabla}_\infty}\mathcal{B}\varphi\,d\mu=0$ for every $\varphi\in\mathscr{D}(\mathcal{B})$. This tells us that $\mu=\mathrm{PD}(\theta,\alpha)$, completing the proof. □

7. The special case $\theta\geq 0$. The arguments of the previous sections assumed $0\leq\alpha<1$ and $\theta>-\alpha$, which are the usual parameter constraints for $\mathrm{PD}(\theta,\alpha)$ distributions with nonnegative α and ensure that the mutation rate $\theta+\alpha$ in (3.7) is positive. It is interesting to note that if one imposes the stronger requirement that θ be nonnegative instead of $\theta>-\alpha$, then a modification of the construction allows us to separate the roles of θ and α, which account for different mechanisms rather than jointly contributing to the mutation events. To this end, assume $\theta\geq 0$ and modify (3.7) to

$$(7.1)\qquad\qquad\qquad\overline{u}_{ij}:=\frac{\theta}{2N(K-1)}$$

and (5.2) to

$$\overline{r}_i(z) := \begin{cases} [1 - (1 - z_i)^K]/z_i & \text{if } z_i > 0, \\ K & \text{if } z_i = 0. \end{cases}$$

Accordingly, (3.9) becomes

$$\overline{b}_i(z) := \frac{1}{2}\left[\frac{\theta}{K-1}(1 - z_i) - \theta z_i + \alpha p_i(z) \sum_{j=1}^{K} z_j \overline{r}_j(z) - \alpha z_i \overline{r}_i(z) \right].$$

Similar arguments to those in the proof of Theorem 5.6 still hold in this setting, and there is an analogue of Remark 5.7. Now, however, θ alone is responsible for mutation through (7.1), while α acts only through the migration mechanism. In contrast, the combined action of θ and α in (3.7) is partially remindful of the action of the same parameters in Pitman's urn scheme construction of the PD(θ, α) distribution [see, e.g., Pitman (1995), equation (15)], where θ and α jointly determine the probability of observing a new type in the sequence.

REFERENCES

BERTOIN, J. (2006). *Random Fragmentation and Coagulation Processes. Cambridge Studies in Advanced Mathematics* **102**. Cambridge Univ. Press, Cambridge. MR2253162

BILLINGSLEY, P. (1968). *Convergence of Probability Measures*. Wiley, New York. MR0233396

COSTANTINI, C. and KURTZ, T. G. (2015). Viscosity methods giving uniqueness for martingale problems. *Electron. J. Probab.* **20** no. 67, 1–27. MR3361255

ETHIER, S. N. (1976). A class of degenerate diffusion processes occurring in population genetics. *Comm. Pure Appl. Math.* **29** 483–493. MR0428488

ETHIER, S. N. (2014). A property of Petrov's diffusion. *Electron. Commun. Probab.* **19** no. 65, 1–4. MR3262071

ETHIER, S. N. and KURTZ, T. G. (1981). The infinitely-many-neutral-alleles diffusion model. *Adv. in Appl. Probab.* **13** 429–452. MR0615945

ETHIER, S. N. and KURTZ, T. G. (1986). *Markov Processes*: *Characterization and Convergence*. Wiley, New York. MR0838085

ETHIER, S. N. and KURTZ, T. G. (1993). Fleming–Viot processes in population genetics. *SIAM J. Control Optim.* **31** 345–386. MR1205982

FENG, S. (2010). *The Poisson–Dirichlet Distribution and Related Topics*. Springer, Heidelberg. MR2663265

FENG, S. and SUN, W. (2010). Some diffusion processes associated with two parameter Poisson–Dirichlet distribution and Dirichlet process. *Probab. Theory Related Fields* **148** 501–525. MR2678897

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. MR2737405

KINGMAN, J. F. C. (1975). Random discrete distributions. *J. R. Stat. Soc. Ser. B* **37** 1–22. MR0368264

LIJOI, A. and PRÜNSTER, I. (2010). Models beyond the Dirichlet process. In *Bayesian Nonparametrics* (N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker, eds.) 80–136. Cambridge Univ. Press, Cambridge. MR2730661

NAGYLAKI, T. (1992). *Introduction to Theoretical Population Genetics*. Springer, Berlin. MR1224677

PERMAN, M., PITMAN, J. and YOR, M. (1992). Size-biased sampling of Poisson point processes and excursions. *Probab. Theory Related Fields* **92** 21–39. MR1156448

PETROV, L. A. (2009). A two-parameter family of infinite-dimensional diffusions on the Kingman simplex. *Funct. Anal. Appl.* **43** 279–296. MR2596654

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

PITMAN, J. (1996). Some developments of the Blackwell–MacQueen urn scheme. In *Statistics, Probability and Game Theory* (T. S. Ferguson, L. S. Shapley and J. B. MacQueen, eds.). *Institute of Mathematical Statistics Lecture Notes—Monograph Series* **30**. Inst. Math. Statist., Hayward, CA. MR1481784

PITMAN, J. (2006). *Combinatorial Stochastic Processes. École D'été de Probabilités de Saint-Fleur XXXII. Lecture Notes in Math.* **1875**. Springer, Berlin. MR2245368

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

RUGGIERO, M. (2014). Species dynamics in the two-parameter Poisson–Dirichlet diffusion model. *J. Appl. Probab.* **51** 174–190. MR3189450

RUGGIERO, M. and WALKER, S. G. (2009). Countable representation for infinite-dimensional diffusions derived from the two-parameter Poisson–Dirichlet process. *Electron. Commun. Probab.* **14** 501–517. MR2564485

RUGGIERO, M., WALKER, S. G. and FAVARO, S. (2013). Alpha-diversity processes and normalized inverse-Gaussian diffusions. *Ann. Appl. Probab.* **23** 386–425. MR3059239

SATO, K. (1978). Diffusion operators in population genetics and convergence of Markov chains. In *Measure Theory Applications to Stochastic Analysis* (*Proc. Conf., Res. Inst. Math., Oberwolfach*, 1977). *Lecture Notes in Math.* **695** 127–137. Springer, Berlin. MR0527079

SHIGA, T. (1981). Diffusion processes in population genetics. *J. Math. Kyoto Univ.* **21** 133–151. MR0606316

TEH, Y. W. and JORDAN, M. I. (2010). Hierarchical Bayesian nonparametric models with applications. In *Bayesian Nonparametrics* (N. L. Hjort, C. C. Holmes, P. Müller and S. G. Walker, eds.) 158–207 Cambridge Univ. Press, Cambridge. MR2730663

ZHOU, Y. (2015). Ergodic inequality of a two-parameter infinitely-many-alleles diffusion model. *J. Appl. Probab.* **52** 238–246. MR3336858

C. COSTANTINI
DEPARTMENT OF ECONOMIC STUDIES
UNIVERSITY OF CHIETI-PESCARA
VIALE PINDARO, 42
65127 PESCARA
ITALY
E-MAIL: c.costantini@unich.it

P. DE BLASI
M. RUGGIERO
ESOMAS DEPARTMENT
UNIVERSITY OF TORINO
CORSO UNIONE SOVIETICA 218/BIS
10134 TORINO
ITALY
E-MAIL: pierpaolo.deblasi@unito.it
matteo.ruggiero@unito.it

S. N. ETHIER
DEPARTMENT OF MATHEMATICS
UNIVERSITY OF UTAH
155 S. 1400 E., JWB 223
SALT LAKE CITY, UTAH 84112
USA
E-MAIL: ethier@math.utah.edu

D. SPANÒ
DEPARTMENT OF STATISTICS
UNIVERSITY OF WARWICK
COVENTRY, CV4 7AL
UNITED KINGDOM
E-MAIL: D.Spano@warwick.ac.uk