On the spectrum and ergodicity of a neutral multi-allelic Moran model

Josué Corujo

Institut de Recherche Mathématique Avancée, UMR 7501 Université de Strasbourg et CNRS, 7 rue René-Descartes, 67000 Strasbourg, France
E-mail address: josue.corujorodriguez@math.unistra.fr
URL: https://josuecorujo.github.io

Abstract. The purpose of this paper is to provide a complete description of the eigenvalues of the generator of a neutral multi-type Moran model, and the applications to the study of the speed of convergence to stationarity. The Moran model we consider is a non-reversible in general, continuous-time Markov chain with an unknown stationary distribution. Specifically, we consider N individuals such that each one of them is of one type among K possible allelic types. The individuals interact in two ways: by an independent irreducible mutation process and by a reproduction process, where a pair of individuals is randomly chosen, one of them dies and the other reproduces. Our main result provides explicit expressions for the eigenvalues of the infinitesimal generator matrix of the Moran process, in terms of the eigenvalues of the jump rate matrix. As consequences of this result, we study the convergence in total variation of the process to stationarity and show a lower bound for the mixing time of the Moran process. Furthermore, we study in detail the spectral decomposition of the neutral multi-allelic Moran model with parent independent mutation scheme, which is the unique mutation scheme that makes the neutral Moran process reversible. Under the parent independent mutation, we also prove the existence of a cutoff phenomenon in the chi-square and the total variation distances when initially all the individuals are of the same type and the number of individuals tends to infinity. Additionally, in the absence of reproduction, we prove that the total variation distance to stationarity of the parent independent mutation process when initially all the individuals are of the same type has a Gaussian profile.

1. Introduction and main results

This paper is devoted to the study of a continuous-time Markov model of N particles on K sites with interaction, which is known as the neutral multi-allelic Moran model in the population genetics literature (Etheridge and Griffiths, 2009): the K sites correspond to K allelic types in a population

of N individuals. The state space of the process is the K-dimensional N-discrete simplex:

$$\mathcal{E}_{K,N} := \left\{ \eta \in [N]_0^K \ : \ |\eta| = N \right\}, \tag{1.1}$$

where $[N]_0 := \{0, 1, \ldots, N\}$ and $|\cdot|$ stands for the sum of elements in a vector. The set $\mathcal{E}_{K,N}$ is a finite set with cardinality $\mathrm{Card}(\mathcal{E}_{K,N}) = \binom{K-1+N}{N}$. The process is in state $\eta \in \mathcal{E}_{K,N}$ if there are $\eta(k) \in [N]_0$ individuals with allelic type $k \in [K] := \{1, 2, \ldots, K\}$. Consider $Q = (\mu_{i,j})_{i,j=1}^K$ the infinitesimal rate matrix of an irreducible Markov chain on $[K]$, which is called the *mutation matrix* of the Moran process. The infinitesimal generator of the neutral multi-allelic Moran process, denoted $\mathcal{Q}_{N,p}$, acts on a real function f on $\mathcal{E}_{K,N}$ as follows:

$$\left(\mathcal{Q}_{N,p} f \right)(\eta) := \sum_{i,j \in [K]} \eta(i) \left(\mu_{i,j} + \frac{p}{N} \eta(j) \right) [f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)], \tag{1.2}$$

for all $\eta \in \mathcal{E}_{K,N}$, where $\mathbf{e}_k$ is the k-th canonical vector of $\mathbb{R}^K$ (cf. Etheridge and Griffiths, 2009). In words, $\mathcal{Q}_{N,p}$ drives a process of N individuals, where each individual has one of K possible types of alleles and where the type of the individual changes following two processes: a mutation process where individuals mutate independently of each other and a Moran type reproduction process, where the individuals interact. The N individuals mutate independently from type $i \in [K]$ to type $j \in [K] \setminus \{i\}$ with rate $\mu_{i,j}$. In addition, with uniform rate $p \geq 0$, one of the N individuals is uniformly chosen to be removed from the population and another one, also randomly chosen, is duplicated. Note that the transitions of an individual due to a reproduction is not independent of the position of the other individuals. As in the original model, introduced by Moran (1958), the same individual removed from the population can be duplicated, in this case the state of the system does not change. In the variation of the model where the removed individual cannot be duplicated, the factor $\frac{p}{N}$ in (1.2) must be replaced by $\frac{p}{N-1}$.

Note that $\mathcal{Q}_{N,p}$ can be decomposed as $\mathcal{Q}_{N,p} = \mathcal{Q}_N + \frac{p}{N} \mathcal{A}_N$, where $\mathcal{Q}_N$ and $\mathcal{A}_N$ are also infinitesimal generators of Markov chains acting on every $f \in \mathbb{R}^{\mathcal{E}_{K,N}}$ as follows

$$\left(\mathcal{Q}_N f \right)(\eta) := \sum_{i,j \in [K]} \eta(i) \mu_{i,j} [f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)], \tag{1.3}$$

$$\left(\mathcal{A}_N f \right)(\eta) := \sum_{i,j \in [K]} \eta(i) \eta(j) [f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)], \tag{1.4}$$

for every $\eta \in \mathcal{E}_{K,N}$. The processes driven by $\mathcal{Q}_N$ and $\mathcal{A}_N$ are called the *mutation process* and the *reproduction process*, respectively. In words, $\mathcal{Q}_N$ models the dynamic of N indistinguishable particles, where each one moves among K sites according to the process generated by the mutation rate matrix Q. This process is usually called the *compound chain*, cf. Zhou and Lange (2009). On the other hand, $\mathcal{A}_N$ models the dynamic where at uniform rate two individuals are randomly chosen and one of them changes its type to the type of the other one. This paper is devoted to the study of the spectrum of $\mathcal{Q}_N$, $\mathcal{A}_N$ and $\mathcal{Q}_{N,p}$, and of the convergence to stationarity of the generated Markov processes. Before stating our main results in this direction, let us establish some notation.

We recall that if $V_n \in \mathbb{R}^K$, $1 \leq n \leq N$ are N vectors in $\mathbb{R}^K$, their tensor product is the vector $V_1 \otimes V_2 \otimes \cdots \otimes V_N$ defined by $(V_1 \otimes V_2 \otimes \cdots \otimes V_N)(k_1, k_2, \ldots, k_N) := V_1(k_1) V_2(k_2) \ldots V_N(k_N)$, for all $1 \leq k_n \leq K$ and $1 \leq n \leq N$. The tensor $V_1 \otimes V_2 \otimes \cdots \otimes V_N$ can be considered as a function on $[K]^N$. Actually, throughout this paper we completely identify a real function f on $[K]^N$ and the tensor vector V_f such that $V_f(k_1, k_2, \ldots, k_N) = f(k_1, k_2, \ldots, k_N)$, for all $(k_1, k_2, \ldots, k_N) \in [K]^N$.

Let us denote by σ a permutation on $[N]$, i.e. an element of the symmetric group $\mathcal{S}_N$. Then, the permutation of $f \in \mathbb{R}^{[K]^N}$ by σ, denoted by σf, is defined by

$$\sigma f : (k_1, k_2, \ldots, k_N) \mapsto f(k_{\sigma(1)}, k_{\sigma(2)}, \ldots k_{\sigma(N)}),$$

for all $(k_1, k_2, \ldots, k_N) \in [K]^N$. In particular, for $V_1, V_2, \ldots, V_N \in \mathbb{R}^N$ we have

$$\sigma(V_1 \otimes V_2 \otimes \cdots \otimes V_N) = V_{\sigma^{-1}(1)} \otimes V_{\sigma^{-1}(2)} \otimes \cdots \otimes V_{\sigma^{-1}(N)}.$$

A real function f on $[K]^N$ is symmetric if $f = \sigma f$, for all σ in $\mathcal{S}_N$. Moreover, every function f on $[K]^N$ can be symmetrised by the projector Sym, defined as follows:

$$\text{Sym} : f \mapsto \overline{f} = \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} \sigma f. \tag{1.5}$$

Symmetric functions on $[K]^N$ are highly important in the sequel because of their relation to the functions on $\mathcal{E}_{K,N}$. Consider the application $\psi_{K,N} : \mathcal{E}_{K,N} \to [K]^N$ defined by

$$\psi_{K,N} : \eta \mapsto (\underbrace{1, 1, \ldots, 1}_{\eta(1)}, \underbrace{2, 2, \ldots, 2}_{\eta(2)}, \ldots, \underbrace{K, K, \ldots, K}_{\eta(K)}), \tag{1.6}$$

where the number of k in $\overbrace{k, k, \ldots, k}^{\eta(k)}$ is set to 0 if $\eta(k) = 0$. Note that for every symmetric function f on $[K]^N$, the function $\tilde{f} := f \circ \psi_{K,N}$ on $\mathcal{E}_{K,N}$ is well defined. Let U_0 be the all-one vector in $\mathbb{R}^K$ and $U_1, U_2, \ldots, U_{K-1} \in \mathbb{R}^K$ such that $\mathcal{U} := \{U_0, U_1, \ldots, U_{K-1}\}$ is a basis of $\mathbb{R}^K$. Note that this is the type of basis given by the eigenvectors of a diagonalisable rate matrix of dimension K. For every $\eta \in \mathcal{E}_{K-1,L}$, for $1 \leq L \leq N$, let us also denote by $U_\eta \in \mathbb{R}^{[K]^N}$, $V_\eta \in \text{Sym}\left(\mathbb{R}^{[K]^N}\right)$ and $\tilde{V}_\eta \in \mathbb{R}^{\mathcal{E}_{K,N}}$ the vectors:

$$U_\eta := U_{k_1} \otimes U_{k_2} \otimes \cdots \otimes U_{k_L} \otimes \underbrace{U_0 \otimes \cdots \otimes U_0}_{N-L \text{ times}}, \tag{1.7}$$

$$V_\eta := \text{Sym}(U_\eta), \tag{1.8}$$

$$\tilde{V}_\eta := V_\eta \circ \psi_{K,N}, \tag{1.9}$$

where $(k_1, k_2, \ldots, k_L) = \psi_{K-1,L}(\eta)$, for every $\eta \in \mathcal{E}_{K-1,L}$ and $L \in [N]$. In Section 2 we analyse the link between the spaces $\text{Sym}\left(\mathbb{R}^{[K]^N}\right)$ and $\mathbb{R}^{\mathcal{E}_{K,N}}$, and we clarify the nature of the definitions previously introduced. The next theorem explains the connection between the eigenstructures of Q and $\mathcal{Q}_N$

Theorem 1.1 (Eigenstructure of $\mathcal{Q}_N$). *Assume $K \geq 2$, $N \geq 1$. Let $\mathcal{U} = \{U_0, U_1, \ldots, U_{r-1}\}$ be a set of r linearly independent right eigenvectors of Q such that U_0 is the all-one vector. Let $\lambda_0 = 0, \lambda_1, \ldots, \lambda_{K-1}$ be the K complex roots of the characteristic polynomial of Q, counting algebraic multiplicities, such that $QU_k = \lambda_k U_k$, for $k \in \{0, 1, \ldots, r - 1\}$. Consider λ_η defined as follows*

$$\lambda_\eta := \sum_{k=1}^{K-1} \eta(k)\lambda_k, \text{ for every } \eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}. \tag{1.10}$$

Then,

(a) *The non-zero eigenvalues of $\mathcal{Q}_N$ are given by λ_η, for all $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$.*

(b) *Every function $\tilde{V}_\eta$, as defined in (1.9), for any $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$ satisfying $\eta(r) = \cdots = \eta(K-1) = 0$ is a right eigenfunction of $\mathcal{Q}_N$ such that $\mathcal{Q}_N \tilde{V}_\eta = \lambda_\eta \tilde{V}_\eta$.*

(c) *In particular, if Q is diagonalisable, then $\mathcal{Q}_N$ is diagonalisable.*

Remark 1.2 (Monotonicity of the spectrum of $\mathcal{Q}_N$ in N). Theorem 1.4 implies that the spectrum of $\mathcal{Q}_N$, for fixed values of K, is an increasing function of N in the sense of the inclusion of sets.

The proof of Theorem 1.1 can be found in Section 3.1. It can be seen as a continuous-time generalisation of the results provided by Zhou and Lange (2009) for the discrete-time analogue of the mutation process driven by $\mathcal{Q}_N$. We emphasize that our hypotheses do not require the mutation rate matrix Q to be reversible, or even diagonalisable.

The next result deals with the spectrum of $\mathcal{A}_N$.

Theorem 1.3 (Spectrum of $\mathcal{A}_N$). *Assume $K \geq 2$ and $N \geq 2$. The eigenvalues of $\mathcal{A}_N$ are*

$$
\begin{aligned}
0 \quad &\text{with multiplicity } K \text{ and} \\
-L(L-1) \quad &\text{with multiplicity } \binom{K+L-2}{L}, \text{ for } 2 \leq L \leq N.
\end{aligned}
$$

Additionally, the infinitesimal rate matrix $\mathcal{A}_N$ is diagonalisable.

The proof of Theorem 1.3 is deferred to Section 3.2. Theorem 1.3 can be seen as a generalisation, for $K \geq 3$, of the results by Zhou (2008, §4.2.2) for the discrete analogous of the reproduction process driven by $\mathcal{A}_N$, for $K = 2$.

Unlike the case of independent mutation process, the dynamics of the neutral multi-allelic Moran process driven by $\mathcal{Q}_{N,p}$, for $p > 0$, is that of an interacting particle system, which makes the study of its spectrum harder. Our main result is precisely a characterisation of the eigenvalues of $\mathcal{Q}_{N,p}$ in terms of those of Q.

Theorem 1.4 (Spectrum of $\mathcal{Q}_{N,p}$). *Assume $K \geq 2$, $N \geq 1$ and $p \in [0, \infty)$. Let us denote by λ_k, $k \in [K-1]$, the nonzero $K-1$ roots, counting algebraic multiplicities, of the characteristic polynomial of Q. For any $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$, let us define*

$$
\lambda_{\eta,p} := \sum_{k=1}^{K-1} \eta(k)\lambda_k - \frac{p}{N}|\eta|(|\eta| - 1).
$$

Then, the eigenvalues of $\mathcal{Q}_{N,p}$, counting algebraic multiplicities, are 0 and $\lambda_{\eta,p}$, for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$.

The proof of Theorem 1.4 is given in Section 3.3.

Remark 1.5 (Relation to the spectrum of the Wright–Fisher diffusion). The eigenstructure of the Wright–Fisher diffusion is a special case of the eigenstructure in a Lambda–Fleming–Viot process studied by Griffiths (2014, Thm. 5). Indeed, taking $W = 0$ in (Griffiths, 2014, Thm. 5), one obtains the spectrum of the neutral Wright–Fisher diffusion, which coincides with the spectrum provided by Theorem 1.4. This is not surprising since the Wright–Fisher diffusion is the limit process for the Moran model (cf. Etheridge, 2011, Lemma 2.39).

Applications to the ergodicity of neutral multi-allelic Moran process. The relation between the spectral properties of $\mathcal{Q}_{N,p}$ and Q can be used to estimate the speed of convergence to stationarity of the Moran process.

Let us first recall the *total variation distance*. For two probability measures ν_1 and ν_2 defined on the same discrete space Ω, the total variation distance is defined as follows:

$$
d^{\mathrm{TV}}(\nu_1, \nu_2) := \sup_{A \subset \Omega} |\nu_1(A) - \nu_2(A)| = \frac{1}{2} \sup_{f:\Omega \to [-1,1]} \left| \int f d\nu_1 - \int f d\nu_2 \right| = \frac{1}{2} \|\nu_1 - \nu_2\|_1,
$$

where $\|\cdot\|_1$ denotes the 1-norm in $\mathbb{R}^\Omega$.

The total variation distance to stationarity at time t of an ergodic process driven by a generator L on Ω, with initial distribution ν, is given by $d^{\mathrm{TV}}(\nu e^{tL}, \pi)$, where ν is the initial distribution on Ω and π is the stationary distribution of the process driven by L, i.e. the unique probability vector such that $\pi L = 0$. We are interested in the relationship between the spectrum of an infinitesimal

rate matrix and the convergence to stationarity of the Markov process it drives. Let us define the *maximum total variation distance* to stationarity of the process driven by L, denoted $\mathrm{D}_L^{\mathrm{TV}}$, as follows:

$$\mathrm{D}_L^{\mathrm{TV}}(t) := \max_\nu \mathrm{d}^{\mathrm{TV}}(\nu\, \mathrm{e}^{tL}, \pi),$$

where the maximum runs over all possible initial distributions on Ω. Using the convexity of d^{TV}, we can prove that $\mathrm{D}_L^{\mathrm{TV}}(t) = \frac{1}{2}\|\mathrm{e}^{tL} - \Pi\|_\infty$, where Π stands for the matrix with every row equal to π, and $\|\cdot\|_\infty$ denotes the infinite norm of matrices (cf. Levin and Peres, 2017, Ch. 4). Recall that the infinite norm of a n-dimensional matrix $(A_{i,j})_{i,j=1}^n$ is simply the maximum absolute row sum, i.e. $\|A\| = \max\limits_{1 \le i \le n} \sum_{j=1}^n |a_{i,j}|$.

As a consequence of Theorem 1.4, the *second largest eigenvalue in modulus* (SLEM) of $\mathrm{e}^{t\mathcal{Q}_{N,p}}$ is equal to that of e^{tQ}, for every $t \ge 0$. The SLEM of the generator of the process is useful to study the asymptotic convergence of the process in total variation. Hence, in Section 4 we study the ergodicity of the process driven by $\mathcal{Q}_{N,p}$ in total variation using the spectral properties of Q. We also analyse several examples of neutral multi-allelic Moran processes with diagonalisable and non-diagonalisable mutation rate matrices.

For a real positive function f, we denote by $\Theta(f)$ another real positive function such that $C_1 f(t) \le \Theta(f)(t) \le C_2 f(t)$, for two constants $0 < C_1 \le C_2 < \infty$ and for all $t \ge T$, with $T > 0$ large enough.

Corollary 1.6 (Asymptotic exponential ergodicity in total variation). *Let us denote by $\mathrm{e}^{-\rho t}$ the SLEM of e^{tQ} and by $s \in \mathbb{N}$ the largest multiplicity in the minimal polynomial of e^{tQ} of all the eigenvalues with modulus $\mathrm{e}^{-\rho t}$, for some $t \ge 0$. Then,*

$$\mathrm{D}_{\mathcal{Q}_{N,p}}^{\mathrm{TV}}(t) = \Theta\big(\mathrm{D}_Q^{\mathrm{TV}}(t)\big) = \Theta\big(t^{s-1}\mathrm{e}^{-\rho t}\big).$$

Corollary 1.6 is proved in Section 4.

Note that the constant in the asymptotic expression in Corollary 1.6 may depend on N. In addition, if Q has a real negative eigenvalue, we can show a lower bound for the mixing time of the process driven by $\mathcal{Q}_{N,p}$.

Theorem 1.7 (Lower bound for convergence in total variation). *Assume $K \ge 2$, $N \ge 2$ and $p \in [0, \infty)$ and let $-\lambda < 0$ be an eigenvalue of Q with associated right-eigenvector $V = [v_1, \ldots, v_K]$. Let $\nu_{N,p}$ be the stationary distribution of the process driven by $\mathcal{Q}_{N,p}$ and let us denote*

$$t_{N,c} := \frac{\ln N - c}{2\lambda} \quad \text{and} \quad \kappa := 8(2 + \|Q\|_\infty/\lambda).$$

Then,

$$\mathrm{d}^{\mathrm{TV}}(\delta_{N\mathbf{e}_k}\mathrm{e}^{t_{N,c}\mathcal{Q}_{N,p}}, \nu_{N,p}) \ge 1 - \kappa\frac{\|V\|_\infty}{|v_k|}\mathrm{e}^{-c},$$

for all $c \ge 0$ and for any $k \in [K]$ such that $v_k \ne 0$. In particular,

$$\mathrm{D}_{\mathcal{Q}_{N,p}}^{\mathrm{TV}}\left(\frac{\ln N - c}{2\lambda}\right) \ge 1 - \kappa\mathrm{e}^{-c}.$$

The proof of Theorem 1.7 is deferred to Section 4.1.

The lower bound provided by Theorem 1.7 ensures that the mixing time of the neutral multi-allelic Moran model is at least of order of $\ln N/2\lambda$. The characterisation of the mixing times of Markov chains is a hard task in general. Lower bounds such that in Theorem 1.7 can be very difficult to prove, even more taking in account that the Markov chain generated by $\mathcal{Q}_{N,p}$ is not necessarily reversible. Our proof relies on spectral arguments and make use of a detailed understanding of the eigenvalues and eigenvectors of $\mathcal{Q}_{N,p}$. This result can be seen as an important application of Theorem 1.4.

Our results do not allow us to prove an upper bound ensuring the existence of a *cutoff phenomenon*. A further study needs to be done in this direction. However, for the parent independent

mutation scheme, a deeper analysis can be done to prove the existence of a cutoff phenomenon in both the chi-square and the total variation distances, as we explain next.

Study of the neutral multi-allelic Moran model with parent independent mutation. Consider the following mutation rate matrix:

$$Q_{\boldsymbol{\mu}} := \begin{pmatrix} -|\boldsymbol{\mu}| + \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_K \\ \mu_1 & -|\boldsymbol{\mu}| + \mu_2 & \mu_3 & \cdots & \mu_K \\ \mu_1 & \mu_2 & -|\boldsymbol{\mu}| + \mu_3 & \cdots & \mu_K \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mu_1 & \mu_2 & \mu_3 & \cdots & -|\boldsymbol{\mu}| + \mu_K \end{pmatrix}, \tag{1.11}$$

where $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K) \in (0, \infty)^K$ and $|\boldsymbol{\mu}|$ stands for the sum of the entries of $\boldsymbol{\mu}$. Let us define

$$(\mathcal{L}_{N,p} f)(\eta) := \sum_{i,j=1}^{K} \eta(i) \left(\mu_j + \frac{p}{N} \eta(j) \right) \left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta) \right],$$

for every f on $\mathcal{E}_{K,N}$ and all $\eta \in \mathcal{E}_{K,N}$, the infinitesimal generator of the neutral multi-allelic Moran process with mutation rate matrix $Q_{\boldsymbol{\mu}}$. The process driven by $\mathcal{L}_{N,p}$ is a special case of the neutral multi-allelic Moran process considered before, but with the difference that the mutation rate only depends on the type of the new individual, i.e. mutation changes each type i individual to type j at rate μ_j, for all $i, j \in [K]$. This is the neutral multi-allelic Moran process with *parent independent* mutation (cf. Etheridge, 2011). Note that $\mathcal{L}_{N,p} = \mathcal{L}_N + \frac{p}{N} \mathcal{A}_N$, where $\mathcal{L}_N := \mathcal{L}_{N,0}$, satisfies

$$(\mathcal{L}_N f)(\eta) := \sum_{i,j=1}^{K} \eta(i) \mu_j \left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta) \right],$$

for every f on $\mathcal{E}_{K,N}$ and all $\eta \in \mathcal{E}_{K,N}$.

Remark 1.8 (Spectrum of $\mathcal{L}_{N,p}$). Using Theorem 1.4 we can easily recover the spectrum of $\mathcal{L}_{N,p}$, which is a well-known result in the population genetics literature. For $K \geq 2$, $N \geq 2$ and $p \geq 0$, the infinitesimal generator $\mathcal{L}_{N,p}$ is diagonalisable with eigenvalues λ_L with multiplicity $\binom{K+L-2}{L}$, where

$$\lambda_{L,p} := -|\boldsymbol{\mu}|L - \frac{p}{N} L(L-1), \tag{1.12}$$

for $L \in [N]_0$. In particular, the spectral gap of $\mathcal{L}_{N,p}$ is $\rho = |\boldsymbol{\mu}|$.

The *complete graph model* studied by Cloez and Thai (2016a) in the context of the Fleming–Viot particle processes is a particular case of the reversible process driven by $Q_{\boldsymbol{\mu}}$ above when $\mu_j = 1/K$, for all $j \in [K]$. In this case, the eigenvalues of the mutation rate are $\beta_0 = 0$ and $\beta_1 = -1$, this last one with multiplicity $K - 1$. In particular, the explicit expression (1.12) improves Lemma 2.14 in (Cloez and Thai, 2016a).

For a real x and $n \in \mathbb{N}_0$ and $\eta \in \mathcal{E}_{K,N}$, we denote by $x_{(n)}$, $x_{[n]}$ and $\binom{N}{\eta}$ the *increasing factorial coefficient*, the *decreasing factorial coefficient* and the *multinomial coefficient*, defined by

$$x_{(n)} := \prod_{k=0}^{n-1} (x+k), \quad x_{[n]} := \prod_{k=0}^{n-1} (x-k) \quad \text{and} \quad \binom{N}{\eta} := \frac{N!}{\prod\limits_{j=1}^{K} \eta(j)!}.$$

We set by convention $x_{(0)} := 1$ and $x_{[0]} := 1$, even for $x = 0$.

The *multinomial distribution* on $\mathcal{E}_{K,N}$ with parameters N and $\mathbf{q} = (q_1, \ldots, q_K) \in (0,1)^K$ such that $|\mathbf{q}| = 1$, denoted $\mathcal{M}(\cdot \mid N, \mathbf{q})$, satisfies

$$\mathcal{M}(\eta \mid N, \mathbf{q}) = \binom{N}{\eta} \prod_{i=1}^{K} q_i^{\eta(i)},$$

for all $\eta \in \mathcal{E}_{K,N}$. Furthermore, the *Dirichlet multinomial distribution* on $\mathcal{E}_{K,N}$ with parameters N and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots, \alpha_K) \in (0, \infty)^K$, denoted $\mathcal{DM}(\cdot \mid N, \boldsymbol{\alpha})$, satisfies

$$\mathcal{DM}(\eta \mid N, \boldsymbol{\alpha}) = \frac{1}{|\boldsymbol{\alpha}|_{(N)}} \binom{N}{\eta} \prod_{k=1}^{K} (\alpha_k)_{(\eta(k))},$$

for all $\eta \in \mathcal{E}_{K,N}$. The distribution $\mathcal{DM}(\cdot \mid N, \boldsymbol{\alpha})$ is a mixture of $\mathcal{M}(\cdot \mid N, \mathbf{q})$ by a *Dirichlet distribution*. See (Mosimann, 1962) for the original reference to the Dirichlet multinomial distribution and (Johnson et al., 2005, §13.1), for more details.

It is known in the population genetics literature that the process driven by $\mathcal{L}_{N,p}$, for $p > 0$, is reversible with stationary distribution $\mathcal{DM}(\cdot \mid N, N\boldsymbol{\mu}/p)$, see e.g. (Etheridge and Griffiths, 2009). Besides, the stationary distribution of the process driven by $\mathcal{L}_N$ is $\mathcal{M}(\cdot \mid N, \boldsymbol{\mu}/|\boldsymbol{\mu}|)$, see e.g. (Zhou and Lange, 2009). Let us define the distribution $\nu_{N,p}$ on $\mathcal{E}_{K,N}$, for all $p \geq 0$, as follows

$$\nu_{N,p}(\eta) := \begin{cases} \mathcal{DM}(\eta \mid N, N\boldsymbol{\mu}/p) & \text{if} \quad p > 0 \\ \mathcal{M}(\eta \mid \boldsymbol{\mu}/|\boldsymbol{\mu}|) & \text{if} \quad p = 0, \end{cases} \tag{1.13}$$

for all $\eta \in \mathcal{E}_{K,N}$. Then, $\nu_{N,p}$ is the stationary distribution of $\mathcal{L}_{N,p}$, for all $p \geq 0$. Besides, the stationary distribution is continuous when $p \to 0$, in the sense that

$$\lim_{p \to 0} \nu_{N,p}(\eta) = \nu_{N,0}(\eta) =: \nu_N(\eta),$$

for every $\eta \in \mathcal{E}_{K,N}$.

In their study of the spectral properties of the discrete-time analogous of $\mathcal{Q}_N$, Zhou and Lange (2009) mainly focus on the case where the process driven by Q is reversible, which is proved to be a necessary and sufficient condition for the reversibility of $\mathcal{Q}_N$. However, the reversibility of Q is not sufficient to ensure the reversibility of the neutral multi-allelic Moran model driven by $\mathcal{Q}_{N,p}$, for $p > 0$, as we discuss in Section 5.1. In fact, the reversibility of neutral multi-allelic Moran processes is completely characterised via the next result.

Lemma 1.9 (Reversible neutral Moran process and parent independent mutation)**.** *Assume $K \geq 2$, $N \geq 2$ and $p > 0$. The process driven by $\mathcal{Q}_{N,p}$ is reversible if and only if the mutation rate matrix has the form $Q_{\boldsymbol{\mu}}$ as in (1.11), for some vector $\boldsymbol{\mu}$, i.e. if and only if $\mathcal{Q}_{N,p}$ can be written as $\mathcal{L}_{N,p}$. Furthermore, the stationary distribution of the process driven by $\mathcal{L}_{N,p}$ is $\nu_{N,p}$ as defined by (1.13).*

The previous result is not surprising because of its analogy with the theory on the measure-valued Fleming–Viot process studied by Ethier and Kurtz (1993). Indeed, the measure-valued Fleming–Viot process is reversible if and only if its mutation factor is parent independent, see e.g. (Ethier and Kurtz, 1993, Thm. 8.2) and (Li et al., 1999, Thm. 1.1). Although the *"if part"* in Lemma 1.9 is well known in the literature on Moran processes, we have not found an explicit statement, or a proof, of this equivalence on the level of prelimit Markov chains considered here. Thus, for the sake of completeness, we provide a proof of Lemma 1.9 in Appendix C.

Section 5 is devoted to the study of the spectral properties of $\mathcal{L}_{N,p}$, for $p \geq 0$, and its applications to the study of the convergence to stationarity. Our results in this section include a complete description of the set of eigenvalues and eigenfunctions of $\mathcal{L}_{N,p}$ and an explicit expression for its transition function. The eigenfunctions of $\mathcal{L}_{N,p}$, $p > 0$, are explicitly given in terms of *multivariate Hahn polynomials*, which are orthogonal with respect to the compound Dirichlet multinomial distribution (cf. Karlin and McGregor, 1975; Khare and Zhou, 2009). The eigenfunctions of $\mathcal{L}_N$, i.e. for

$p = 0$, are explicitly given in terms of *multivariate Krawtchouk polynomials*, which are orthogonal with respect to the multinomial distribution (cf. Diaconis and Griffiths, 2014; Karlin and McGregor, 1965; Zhou and Lange, 2009).

Remark 1.10 (Relation to Ewens' sampling distribution). The Moran model, where mutation rates are constant across all types, i.e., $\mu_{i,j} = \theta$ for all i, j, has a connection to the Ewens' sampling distribution. This link was explored, for instance, by Watterson (1984, § 3.2) and Zhou (2008, § 4.1). Notably, in this scenario, we can treat mutation types as indistinguishable and ignore individual allelic types, resulting in a Markov process on the space of integer partitions. For example, in a 3-type Moran process, states $(3, 2, 2)$, $(2, 3, 2)$, and $(2, 2, 3)$ correspond to the same integer partition of $N = 7$. It would be worthwhile to explore if the techniques introduced in this paper could be used to compute the spectral elements of the lumped Markov chain in this scenario. This research direction falls beyond the scope of this paper, it was suggested by one of the anonymous reviewers and merits further investigation.

Cutoff phenomenon. The *cutoff phenomenon* has been a rich topic of research on Markov chains since its introduction by the works of Aldous, Diaconis and Shahshahani in the 1980s (cf. Aldous, 1983; Aldous and Diaconis, 1986; Diaconis and Shahshahani, 1981). A Markov chain presents a cutoff if it exhibits an abrupt transition in its convergence to stationarity. Some of the most used notions of convergence are, as we consider here, the total variation and the chi-square distances. A good introduction to this subject can be found in the classic book of Levin and Peres (2017, Ch. 18) and in the exhaustive work of Chen, Saloff-Coste et al. See for instance (Chen, 2006; Chen et al., 2017; Chen and Saloff-Coste, 2008, 2010; Saloff-Coste, 1997).

A typical scenario for the existence of a cutoff is a Markov chain with a high degree of symmetry. Hence, the cutoff phenomenon has been deeply studied for the movement on N independent particles on K sites, a model which is usually known as the *product chain.* Ycart (1999) studied the cutoff in total variation for N independent particles driven by a diagonalisable rate matrix. Later, Barrera et al. (2006) and Connor (2010) studied the cutoff on this model according to other notions of distance. See also (Chen et al., 2017; Chen and Kumagai, 2018; Lacoin, 2015) and (Levin and Peres, 2017, Ch. 20) for more recent studies about the cutoff on product chains. The Moran model we consider here preserves the high level of symmetry of the product chain, but the movements of the particles are not independent. Indeed, the particles interact according to a reproduction process that favours the jumps to the sites with greater proportions of individuals.

Before formally defining the cutoff phenomenon, let us recall the *chi-square divergence* (sometimes called "distance"), which naturally arises in the context of reversible Markov chains. The chi-square divergence of ν_2 with respect to the target distribution ν_1 is defined by

$$\chi^2(\nu_2 \mid \nu_1) := \sum_{\omega \in \Omega} \frac{[\nu_2(\omega) - \nu_1(\omega)]^2}{\nu_1(\omega)} = \|\nu_2 - \nu_1\|_{\frac{1}{\nu_1}}^2,$$

where $\|\cdot\|_{\frac{1}{\nu_1}}$ stands for the norm in $l^2(\mathbb{R}^\Omega, \frac{1}{\nu_1})$, and $\frac{1}{\nu_1}$ is the measure $\omega \mapsto 1/\mu_1(\omega)$.

The chi-square divergence is not a metric, but a measure of the difference between two probability distributions. Note that the chi-square divergence, as well as the total variation distance, are special cases of the so called $f-$divergence functions, which measure the "difference" between two probability distributions, see (Nielsen and Nock, 2014). In this context, $\chi^2(\mu_2 \mid \mu_1)$ is also known as *Pearson chi-square divergence.*

Abusing notation, let us define the functions χ_η^2 and d_η^{TV}, as follows

$$d_\eta^{\mathrm{TV}}(t) := d^{\mathrm{TV}}(\delta_\eta e^{t\mathcal{L}_{N,p}}, \nu_{N,p}) = \frac{1}{2} \sum_{\xi \in \mathcal{E}_{K,N}} \left| \left(e^{t\mathcal{L}_{N,p}} \delta_\xi \right)(\eta) - \nu_{N,p}(\xi) \right|,$$

$$\chi_\eta^2(t) := \chi^2(\delta_\eta e^{t\mathcal{L}_{N,p}} \mid \nu_{N,p}) = \sum_{\xi \in \mathcal{E}_{K,N}} \frac{\left[\left(e^{t\mathcal{L}_{N,p}} \delta_\xi \right)(\eta) - \nu_{N,p}(\xi) \right]^2}{\nu_{N,p}(\xi)}.$$

The functions d_η^{TV} and χ_η^2 are thus measures of the convergence to stationary of the process driven by $\mathcal{L}_{N,p}$ at time t and with initial configuration $\eta \in \mathcal{E}_{K,N}$. In agreement with Khare and Zhou (2009); Zhou (2008) we call χ_η^2 and d_η^{TV} the *total variation* and the *chi-square distances* to stationarity, respectively.

As the number of individuals varies we obtain an infinite family of continuous-time finite Markov chains $\{(\mathcal{E}_{K,N}, \mathcal{L}_{N,p}, \nu_{N,p}), N \geq 2\}$. For each $N \geq 2$ let us denote by $\chi_{N\mathbf{e}_k}^2(t)$ (resp. $d_{N\mathbf{e}_k}^{\mathrm{TV}}(t)$) the chi-square distance (resp. total variation distance) to stationarity of the process driven by $\mathcal{L}_{N,p}$ at time t, when the initial distribution is concentrated at $N\mathbf{e}_k \in \mathcal{E}_{K,N}$. Note that $\chi_{N\mathbf{e}_k}^2(0) \to \infty$ and $d_{N\mathbf{e}_k}^{\mathrm{TV}}(0) \to 1$, when $N \to \infty$.

Definition 1.11 (Chi-square and total variation cutoff)**.** We say that $\{\chi_{N\mathbf{e}_k}^2(t), N \geq 2\}$ exhibits a (t_N, b_N) chi-square cutoff if $t_N \geq 0$, $b_N \geq 0$, $b_N = o(t_N)$ and

$$\lim_{c \to \infty} \limsup_{N \to \infty} \chi_{N\mathbf{e}_k}^2(t_N + c\, b_N) = 0, \quad \lim_{c \to -\infty} \liminf_{N \to \infty} \chi_{N\mathbf{e}_k}^2(t_N + c\, b_N) = \infty.$$

Analogously, we say that $\{d_{N\mathbf{e}_k}^{\mathrm{TV}}(t), N \geq 2\}$ exhibits a (t_N, b_N) total variation cutoff if $t_N \geq 0$, $b_N \geq 0$, $b_N = o(t_N)$ and

$$\lim_{c \to \infty} \limsup_{N \to \infty} d_{N\mathbf{e}_k}^{\mathrm{TV}}(t_N + c\, b_N) = 0, \quad \lim_{c \to -\infty} \liminf_{N \to \infty} d_{N\mathbf{e}_k}^{\mathrm{TV}}(t_N + c\, b_N) = 1.$$

The sequences $(t_N)_{N \geq 2}$ and $(b_N)_{N \geq 2}$ are called the *cutoff* and the *window sequences*, respectively.

See Definition 2.1 and Remark 2.1 in (Chen and Saloff-Coste, 2008).

The cutoff phenomenon describes an abrupt transition in the convergence to stationarity: over a negligible period given by the window sequence $(b_N)_{N>2}$, the distance from equilibrium drops from near its initial value to near zero at a time given by the cutoff sequence $(t_N)_{N \geq 2}$.

A stronger condition for the existence of a (t_N, b_N) chi-square cutoff (resp. total variation cutoff) is the existence of the limit

$$G_k(c) := \lim_{N \to \infty} \chi_{N\mathbf{e}_k}^2(t_N + c\, b_N) \quad \left(\text{resp. } H_k(c) := \lim_{N \to \infty} d_{N\mathbf{e}_k}^{\mathrm{TV}}(t_N + c\, b_N) \right),$$

for a function G_k (resp. H_k), for $k \in [K]$, satisfying:

$$\lim_{c \to -\infty} G_k(c) = \infty \text{ and } \lim_{c \to \infty} G_k(c) = 0, \quad \left(\text{resp. } \lim_{c \to -\infty} H_k(c) = 1 \text{ and } \lim_{c \to \infty} H_k(c) = 0 \right).$$

Actually, in this case the (t_N, b_N) cutoff is said to be *strongly optimal*, see e.g. Definition 2.2 and Proposition 2.2 in (Chen and Saloff-Coste, 2008). See Chapter 2 in (Chen, 2006) and Sections 2.1 and 2.2 in (Chen and Saloff-Coste, 2008) for more details about the definition of (t_N, b_N) cutoff and window optimality.

The next two results establish the existence of cutoff phenomena in the chi-square and the total variation distances for the multi-allelic Moran process driven by $\mathcal{L}_{N,p}$, for $p \geq 0$, when the initial distribution is concentrated at $N\mathbf{e}_k$, for $k \in [K]$. In the chi-square case, we are able to explicitly provide the limit profile of the distance. Moreover, we prove that the total variation distance to stationarity of the mutation process driven by $\mathcal{L}_N$, i.e. for $p = 0$, has a Gaussian profile when all the individuals are initially of the same type.

Theorem 1.12 (Strongly optimal chi-square cutoff when $N \to \infty$). *For $k \in [K]$, with $K \geq 2$, $p \geq 0$ and every $c \in \mathbb{R}$, we have*

$$\lim_{N \to \infty} \chi^2_{N\mathbf{e}_k}(t_{N,c}) = \exp\{K_{k,p}\mathrm{e}^{-c}\} - 1, \tag{1.14}$$

where $t_{N,c} = \dfrac{\ln N + c}{2|\boldsymbol{\mu}|}$ and $K_{k,p} = \dfrac{|\boldsymbol{\mu}|(|\boldsymbol{\mu}| - \mu_k)}{\mu_k(|\boldsymbol{\mu}| + p)}$. Consequently, the Markov process driven by $\mathcal{L}_{N,p}$ has a strongly optimal $\left(\dfrac{\ln N}{2|\boldsymbol{\mu}|}, 1\right)$ chi-square cutoff when $N \to \infty$.

Theorem 1.13 (Total variation cutoff when $N \to \infty$). *For every $k \in [K]$, with $K \geq 2$, $p \geq 0$ and every $c > 0$, we have*

$$\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}\left(\frac{\ln N - c}{2|\boldsymbol{\mu}|}\right) \geq 1 - 32\kappa_k \mathrm{e}^{-c},$$

$$\lim_{N \to \infty} \mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}\left(\frac{\ln N + c}{2|\boldsymbol{\mu}|}\right) \leq \frac{1}{2}\sqrt{\exp\{K_{k,p}\mathrm{e}^{-c}\} - 1},$$

where $\kappa_k = \max\limits_{r:r \neq k} \dfrac{\mu_r}{\mu_k} \wedge 1$ and $K_{k,p} = \dfrac{|\boldsymbol{\mu}|(|\boldsymbol{\mu}| - \mu_k)}{\mu_k(|\boldsymbol{\mu}| + p)}$. Consequently, the Markov process driven by $\mathcal{L}_{N,p}$ exhibits a $\left(\dfrac{\ln N}{2|\boldsymbol{\mu}|}, 1\right)$ total variation cutoff when $N \to \infty$.

Moreover, when $p = 0$ and for every $c \in \mathbb{R}$, the limit profile of the total variation distance satisfies

$$\lim_{N \to \infty} \mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t_{N,c}) = 2\Phi\left(\frac{1}{2}\sqrt{K_{k,0}\mathrm{e}^{-c}}\right) - 1,$$

where Φ is the cumulative distribution function of the standard normal distribution. Thus, there exists a strongly optimal $\left(\dfrac{\ln N}{2|\boldsymbol{\mu}|}, 1\right)$ total variation cutoff for the process driven by $\mathcal{L}_N$ when $N \to \infty$.

Proof of Theorem 1.12 and 1.13 will be given in Section 5.1.

During the proof of Theorem 1.12, we prove the following result which is of independent interest.

Corollary 1.14 (Law of the process driven $\mathcal{L}_N$ ($p = 0$)). *The law of the process driven by $\mathcal{L}_N$ at time t when initially all the individuals are of type $k \in [K]$ is multinomial*

$$\mathcal{M}\left(\cdot \mid N, \frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|}(1 - \mathrm{e}^{-|\boldsymbol{\mu}|t}) + \mathrm{e}^{-|\boldsymbol{\mu}|t}\mathbf{e}_k\right).$$

Several authors have studied the existence of a cutoff in Moran type models. For instance, Donnelly and Rodrigues (2000) proved the existence of a cutoff for the two-allelic neutral Moran model in the separation distance. In order to do that, they used a duality property of the Moran process and found an asymptotic expression for the convergence in separation distance for a suitable scaled time, when the number of individuals tends to infinity. Khare and Zhou (2009) proved bounds for the chi-square distance in a discrete-time multi-allelic Moran process that implies the existence of a cutoff. Diaconis and Griffiths (2019) studied the existence of a chi-square and total variation cutoffs for a discrete-time analogous of the mutation process generated by $\mathcal{L}_N$, i.e. when $p = 0$. See also the study of similar models arried out by Diaconis et al. (2008) using spectral theory. Theorems 1.12 and 1.13 sharpen the results in (Khare and Zhou, 2009) and (Diaconis and Griffiths, 2019), since they provide the explicit limit profiles for the chi-square and the total variation distances, for $p \geq 0$ and $p = 0$, respectively.

The main difficulty in proving the existence of a total variation cutoff is that even if the Markov chain is reversible, the spectral decomposition of the generator of the Markov chains may only provide a good-enough upper bound for the total variation distance. In general, one needs to find a sufficiently sharp lower bound using another method. In our case, we could obtain such a bound by using the detailed description of the eigenvectors of $\mathcal{Q}_{N,p}$ of lower modulus, i.e. Theorem 1.7.

Theorem 1.13 is, to the best of our knowledge, the first result ensuring the existence of a total variation cutoff phenomenon for the neutral Moran model with parent independent mutation and neutral reproduction, i.e. with $p > 0$. Notice that the bound given by Theorem 1.7 applies to a general class of Moran-type process with neutral reproduction (not necessarily in the parent independent mutation setting). Hence, this result should facilitate the proof of the existence of total variation cutoff phenomena for other (non-reversible) Markov chains, using other methods in mixing times theory, such as coupling or spectral theory.

Results for the discrete-time Moran model

We next briefly discuss some results that can be obtained for an analogous discrete-time multi-allelic Moran model. Let $(\mu_{i,j})_{i,j \in [K]}$ be a K-dimensional stochastic matrix. We define the mutation and reproduction transition matrices $\boldsymbol{Q}_N$ and $\boldsymbol{A}_N$ as follows

$$\boldsymbol{Q}_N(\eta, \eta - \mathbf{e}_i + \mathbf{e}_j) = \frac{\eta(i)}{N}\mu_{i,j}, \quad \text{and} \quad \boldsymbol{A}_N(\eta, \eta - \mathbf{e}_i + \mathbf{e}_j) = \frac{\eta(i)}{N}\frac{\eta(j)}{N},$$

for every $\eta \in \mathcal{E}_{K,N}$ and $i \neq j \in [K]$. In addition,

$$\boldsymbol{Q}_N(\eta, \eta) = 1 - \sum_{i \neq j}\boldsymbol{Q}_N(\eta, \eta - \mathbf{e}_i + \mathbf{e}_j), \quad \text{and} \quad \boldsymbol{A}_N(\eta, \eta) = 1 - \sum_{i \neq j}\boldsymbol{A}_N(\eta, \eta - \mathbf{e}_i + \mathbf{e}_j).$$

Besides, every other entry of $\boldsymbol{Q}_N$ and $\boldsymbol{A}_N$ is null. Then, $\boldsymbol{Q}_N$ and $\boldsymbol{A}_N$ are well-defined transition matrices defining two discrete-time Markov chains, similar to the ones generated by $\mathcal{Q}_N$ and $\mathcal{A}_N$. The discrete-time multi-allelic Moran model is the Markov chain with transition matrix $\boldsymbol{Q}_{N,p} := p\boldsymbol{Q}_N + (1 - p)\boldsymbol{A}_N$ for some $p \in (0,1]$. Note that this is the Moran process defined by Khare and Zhou (2009, § 4.1.1) when the mutation probabilities of their model take the form $m_{i,j} = (1 - p)\mathbb{1}_{\{i=j\}} + p\,\mu_{i,j}$.

The transition matrices of the discrete-time mutation and reproduction processes and the generator of their continuous-time analogous viewed as matrices, are connected through the identities

$$\boldsymbol{Q}_N = \mathrm{I} + \frac{1}{N}\mathcal{Q}_N \quad \text{and} \quad \boldsymbol{A}_N = \mathrm{I} + \frac{1}{N}\mathcal{A}_N, \tag{1.15}$$

where I denotes the identity matrix. The factor $1/N$ is necessary in order to ensure that the diagonal elements in $\boldsymbol{Q}_N$ and $\boldsymbol{A}_N$ are positive.

Thus, given an irreducible stochastic matrix $(\mu_{i,j})_{i,j \in [K]}$, the results on the eigenvalues and eigenvectors of $\mathcal{Q}_N$ and $\mathcal{A}_N$ (namely, Theorems 1.1, 1.3 and 1.4) are easily translated to those of $\boldsymbol{Q}_N$ and $\boldsymbol{A}_N$. In particular, if $\lambda_0 = 1$ and λ_k, for $k \in [K-1]$, are the K roots, counting algebraic multiplicities, of the characteristic polynomial of $(\mu_{i,j})_{i,j \in [K]}$. For every $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$, let us define

$$\lambda_{\eta,p} := 1 - p\sum_{k=1}^{K-1}\frac{\eta(k)}{N}\lambda_k - (1 - p)\frac{|\eta|(|\eta| - 1)}{N^2}.$$

Then, the eigenvalues of $\boldsymbol{Q}_{N,p}$, counting algebraic multiplicities, are 1 and $\lambda_{\eta,p}$, for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$.

In addition, we have some information on the eigenvectors of $\boldsymbol{Q}_{N,p}$, in terms of the eigenvectors of $(\mu_{i,j})_{i,j \in [K]}$. Furthermore, when $(\mu_{i,j})_{i,j \in [K]}$ allows a real eigenvalue $\lambda < 1$ we can find an estimate that in the same spirit of Theorem 1.7, provides a lower bound for the total variation distance to stationarity at time $(N \ln(N) - Nc)/2$, for every $c \geq 0$.

Adhering strictly to the demonstration of Corollary 1.9, we also get that the discrete-time multi-allelic Moran process is reversible if and only if its mutation matrix is parent independent, i.e. if

$\mu_{i,j} = \mu_j$. Besides, the stationary distribution $\nu_{N,p}$ of the Markov chain with transition rates $\boldsymbol{Q}_{N,p}$ satisfies

$$\nu_{N,p}(\eta) := \begin{cases} \mathcal{DM}(\eta \mid N, N\boldsymbol{\mu}\,p/(1-p)) & \text{if} \quad p \in (0,1) \\ \mathcal{M}(\eta \mid \boldsymbol{\mu}) & \text{if} \quad p = 1, \end{cases}$$

We can also extrapolate the results on the ergodicity for those processes with parent independent mutation rates, obtaining analogous results to Theorems 1.12 and 1.13. In particular, we can ensure that discrete-time Moran model with parent independent mutation exhibits a $(N \ln N/2, N)$ total variation cutoff when $N \to \infty$, when initially all the individuals are of the same type. The extra factor N in the cutoff time and window is directly linked to the identity (1.15) and the expression for the eigenvalues of $\boldsymbol{Q}_{N,p}$. One can check that in the discrete time setting when $p \in (0,1)$ and taking $l_{N,c} = \lceil (N\ln(N) + Nc)/2 \rceil$, for $c \in \mathbb{R}$, we get

$$\lim_{N\to\infty} \chi^2_{N\mathbf{e}_k}(l_{N,c}) = \exp\left\{ (1-p)\frac{1-\mu_k}{\mu_k}\mathrm{e}^{-c} \right\} - 1, \tag{1.16}$$

and when $p = 0$,

$$\lim_{N\to\infty} \mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(l_{N,c}) = 2\Phi\left(\frac{1}{2}\sqrt{\frac{1-\mu_k}{\mu_k}}\mathrm{e}^{-c/2} \right) - 1.$$

The explicit expression (1.16) for the chi-square limit profile strengthen Proposition 4.7 in (Khare and Zhou, 2009).

Discussion and open problems

There are several future directions to explore in order to better understand Moran models. Despite the fact that it is non-reversible in general, the neutral multi-allelic Moran model with reversible mutation process seems an interesting model for both theoretical and practical reasons (cf. Schrempf and Hobolth, 2017). One possible first step to study the eigenfunctions of $\mathcal{Q}_{N,p}$ when Q is reversible, could be the study of the eigenfunctions of the generator of the reproduction process $\mathcal{A}_N$, for $K \geq 3$, extending the results in (Zhou, 2008, §4.2.2).

There are several ways to continue the study of the existence of cutoff phenomena for Moran processes. For example, using the results of Zhou and Lange (2009), it could be possible to prove the existence of a (strongly optimal) chi-square cutoff for the composition chain, when the process driven by the mutation matrix is reversible. A possible generalisation of Theorems 1.12 and 1.13 would be to prove the existence of a cutoff phenomenon for the Moran process with parent independent mutation, when initially all the individuals are not of the same type.

Another interesting problem to address is the study of the spectrum of the multi-allelic Moran process with selection, i.e. when the parameter p in (1.2) may depend on i and j. Under parent independent settings and selection at birth (cf. Durrett, 2008; Etheridge, 2011; Muirhead and Wakeley, 2009) the infinitesimal rate matrix of the process is reversible, but an explicit expression for its spectral gap is unknown. The multi-allelic Moran process with selection at death (cf. Muirhead and Wakeley, 2009) seems more complicated from the spectral point of view because it is non-reversible. However, this process is very interesting in population genetics but also because of its interpretation as a Fleming–Viot particle system, which approximates the quasi-stationary distribution of a continuous-time Markov chain (see e.g. Asselah et al., 2011; Cloez and Thai, 2016b; Ferrari and Marić, 2007). We believe the exact results exhibit here for the neutral Moran process will offer clues on the study of the spectrum of the more complicated Moran processes with selection.

Structure of the article. The rest of the paper is organised as follows. In Section 2 we study the state spaces of the neutral multi-allelic Moran models, when the individuals are assumed distinguishable or indistinguishable, respectively. We particularly focus on the study of the vector spaces of real functions defined on the state spaces of these two models. The notations and results in Section 2 are

used to prove our main theorems in Section 3. Sections 3.1, 3.2 and 3.3 are devoted to the proofs of Theorems 1.1, 1.3 and 1.4, respectively. In Section 4 we focus on the applications of our main results to the asymptotic exponential ergodicity in total variation distance of the process driven by $\mathcal{Q}_{N,p}$ to its stationary distribution, using the eigenstructure of Q. In particular, we prove Corollary 1.6 and Theorem 1.7. In Section 5 we consider the neutral multi-allelic Moran process with parent independent mutation and provide a complete description of its eigenvalues and eigenfunctions. We also prove Theorems 1.12 and 1.13 about the existence of a cutoff phenomena in the chi-square and the total variation distances, when initially all the individuals are of the same type.

2. State spaces for distinguishable and indistinguishable particle processes

The Moran model can be seen as a system of N interacting particles on K sites moving according to a continuous-time Markov chain. For the same model, we study two different situations. Although the sites themselves are supposed to be distinguishable, the N particles can be considered either *distinguishable* or *indistinguishable*. According to both interpretations we describe two state spaces for the two Markov chains modelling the N independent particle systems. We study how the vector spaces of the real functions defined on those state spaces are related.

For N *distinguishable* particles on K sites, the state space of the model describes the location of each particle, i.e. it is the set $[K]^N$. This is the state space considered by Etheridge (2011) and Ferrari and Marić (2007). The set of real functions on $[K]$, denoted $\mathbb{R}^{[K]}$, may be endowed with a vector space structure. Thus, the set of real functions on $[K]^N$ may be considered as a tensor product of N vectors in $\mathbb{R}^K$ as we commented in the introduction.

When the N particles are considered *indistinguishable*, what matters is the number of particles present at each of the K sites. The state space for this second model, as in (Cloez and Thai, 2016b) and (Etheridge and Griffiths, 2009), is the set $\mathcal{E}_{K,N}$ defined by (1.1) with cardinality equal to $\mathrm{Card}\,(\mathcal{E}_{K,N}) = \binom{K-1+N}{N}$.

For any k, $1 \leq k \leq K$, let us denote by x_k the k-th coordinate function defined by

$$x_k : \eta = (\eta(1), \eta(2), \ldots, \eta(K)) \in \mathcal{E}_{K,N} \mapsto \eta(k) \in \mathbb{R}.$$

Let us also denote by $\mathbf{x}^\alpha$ the monomial on $\mathcal{E}_{K,N}$ defined by

$$\mathbf{x}^\alpha := x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_K^{\alpha_K}, \tag{2.1}$$

where $\alpha \in \mathcal{E}_{K,L}$, for $L \in [N]$.

For $0 \leq L \leq N$, let us denote by $H_{K,L}$ the vector space of homogeneous polynomial functions of degree L in variables x_k, $1 \leq k \leq K$ on $\mathcal{E}_{K,N}$ and the null function. From the definition of $\mathcal{E}_{K,N}$, it follows that the function $\sum_{k=1}^K x_k$ is equal to the constant function equal to N. $H_{K,L}$ may be considered as a subspace of $H_{K,L'}$ when $0 \leq L < L' \leq N$ by identifying $P(x_1, x_2, \ldots, x_K) \in H_{K,L}$ with

$$\frac{1}{N^{L'-L}} \left(\sum_{k=1}^K x_k \right)^{L'-L} P(x_1, x_2, \ldots, x_K) \in H_{K,L'}.$$

We will say that the *degree of homogeneity* of a homogeneous polynomial P is L, if P is the sum of monomials $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_K^{\alpha_K}$ with the same value of $|\alpha| = L$, and the value L is the smallest as possible. This corresponds to the fact that there is no factor equal to $x_1 + x_2 + \cdots + x_K$ in the factorisation of P. The *total degree* of a polynomial P is the minimum value of L such that $P = P_L + R$ where P_L is homogeneous of degree L and all the monomials in R have a maximum degree strictly less than L. Such an expression for P, which is not unique, may be obtained by replacing x_K by $N - \sum_{k=1}^{K-1} x_k$ in P and adding the monomials in $P(x_1, \ldots, x_{K-1}, N - \sum_{k=1}^{K-1} x_k)$ of maximum total degree to define P_L.

The following result shows that each element in $\mathbb{R}^{\mathcal{E}_{K,N}}$ can be identify to a polynomial in $H_{K,N}$.

Lemma 2.1. *Let $K \geq 2$ and $N \geq 1$. Then*

    (a) *For any real function f on $\mathcal{E}_{K,N}$ there exists a unique homogeneous polynomial $P \in H_{K,N}$ of degree N such that $f(\eta) = P(\eta)$, for all $\eta \in \mathcal{E}_{K,N}$.*

    (b) *The set of monomials of degree N*

$$\mathcal{B}_{H_{K,N}} := \{\mathbf{x}^\alpha, \alpha \in \mathcal{E}_{K,N}\}$$

where $\mathbf{x}^\alpha$ is defined by (2.1), is a basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$.

The proof of Lemma 2.1 is mostly technical and is deferred to Appendix A.

Remark 2.2 (Dimension of $H_{K,N}$). As a consequence of Lemma 2.1-(b) we have that the dimension of $H_{K,N}$ equals $\binom{K+N-1}{N}$.

A natural link between the two state spaces of distinguishable and indistinguishable particles is

$$\phi_{K,N} : (k_1, k_2, \ldots, k_N) \in [K]^N \mapsto (\eta(1), \eta(2), \ldots, \eta(K)) \in \mathcal{E}_{K,N}, \tag{2.2}$$

where $\eta(k) = \text{Card}(\{n, \ 1 \leq n \leq N, \ k_n = k\})$, for all $k \in [K]$. The function $\phi_{K,N}$ forgets the identity of the N particles. Note that $\psi_{K,N}$, defined in (1.6), is a right inverse of $\phi_{K,N}$, i.e. $\phi_{K,N} \circ \psi_{K,N} = \text{Id}_{\mathcal{E}_{K,N}}$, where $\text{Id}_{\mathcal{E}_{K,N}}$ stands for the identity function on $\mathcal{E}_{K,N}$.

Let us denote by Sym the *symmetrisation* endomorphism, acting on function $f \in \mathbb{R}^{[K]^N}$ as defined by (1.5). In fact, Sym is the projection onto the subspace of symmetric functions, denoted $\text{Sym}\big(\mathbb{R}^{[K]^N}\big)$.

Note that $\phi_{K,N}$ is a symmetric function on $[K]^N$. Furthermore, the identity $\phi_{K,N}(\mathbf{x}) = \phi_{K,N}(\mathbf{y})$ holds if and only if $\mathbf{y}$ is obtained from $\mathbf{x}$ by a permutation of its components. Hence, if f is symmetric and $\mathbf{x}$ and $\mathbf{y}$ are elements in $[K]^N$ such that $\phi_{K,N}(\mathbf{x}) = \phi_{K,N}(\mathbf{y})$, then $f(\mathbf{x}) = f(\mathbf{y})$.

In general, for every function f on $[K]^N$ it is not always possible to define a function $\tilde{f}$ on $\mathcal{E}_{K,N}$ such that $f = \tilde{f} \circ \phi_{K,N}$ holds. We claim that such a function $\tilde{f}$ exists if and only if f is symmetric.

Lemma 2.3 (Link between $\mathbb{R}^{\mathcal{E}_{K,N}}$ and $\text{Sym}(\mathbb{R}^{[K]^N})$). *The linear operator*

$$\Phi_{K,N} : f \in \text{Sym}\left(\mathbb{R}^{[K]^N}\right) \mapsto f \circ \psi_{K,N} \in \mathbb{R}^{\mathcal{E}_{K,N}}, \tag{2.3}$$

where $\psi_{K,N}$ is defined by (1.6), is an isomorphism. In particular, the dimension of the space of symmetric functions on $[K]^N$ is

$$\dim\left(\text{Sym}\left(\mathbb{R}^{[K]^N}\right)\right) = \binom{K+N-1}{N}.$$

Proof: Note that $\Phi_{K,N}$ is linear and well defined. Moreover, for any function h on $\mathcal{E}_{K,N}$, the function $h \circ \phi_{K,N}$ is symmetric on $[K]^N$ and satisfies $\Phi_{K,N}(h \circ \phi_{K,N}) = h$, proving that $\Phi_{K,N}$ is an isomorphism. $\square$

Lemma 2.3 justifies the well definiteness of $\tilde{V}_\eta$, defined by (1.9), for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$. The relationship between f and $\tilde{f}$ is shown in the following diagram:

$$
\begin{array}{ccc}
 & [K]^N & \\
\phi_{K,N} \downarrow & \searrow^{f} & \\
 & \mathcal{E}_{K,N} \xrightarrow[\tilde{f}]{} & \mathbb{R}.
\end{array}
$$

We denote by U_0 the K-dimensional all-one vector, which is always a right eigenvector associated to zero of every K-dimensional rate matrix of a continuous-time Markov chain. Let $K \geq 2$, $N \geq 2$

and $1 \leq L \leq N$ and let us consider L vectors $V_1, V_2, \ldots, V_L$ in $\mathbb{R}^K$, non-proportional to U_0, and f the function equal to the following symmetrised tensor product

$$f := \mathrm{Sym}(V_1 \otimes V_2 \otimes \cdots \otimes V_L \otimes \underbrace{U_0 \otimes \cdots \otimes U_0}_{N-L}) \in \mathrm{Sym}\left(\mathbb{R}^{[K]^N}\right).$$

Note that,

$$f(k_1, k_2, \ldots, k_N) = \frac{1}{N!} \sum_{\sigma \in \mathcal{S}_N} V_1(k_{\sigma(1)}) V_2(k_{\sigma(2)}) \times \cdots \times V_L(k_{\sigma(L)}). \qquad (2.4)$$

We denote by $\mathcal{I}_{L,N}$, for $1 \leq L \leq N$, the set of all injective applications from $[L]$ to $[N]$. For every $\sigma \in \mathcal{S}_N$, the map $s_\sigma : n \in [L] \mapsto \sigma(n) \in \{\sigma(1), \ldots, \sigma(L)\}$ is in $\mathcal{I}_{L,N}$ and σ is completely determined by this function s_σ and a bijective application $\beta : (L+1, \ldots, N) \to [N] \setminus s_\sigma([L])$. For each s_σ, there are $(N-L)!$ such applications β. Thus, using (2.4) we obtain

$$f(k_1, k_2, \ldots, k_N) = \frac{(N-L)!}{N!} \sum_{s \in \mathcal{I}_{L,N}} V_1(k_{s(1)}) V_2(k_{s(2)}) \times \cdots \times V_L(k_{s(L)}).$$

In order to simplify the calculations we denote by $\xi(V_1, V_2, \ldots, V_L)$ the function on $[K]^N$ defined by

$$\xi(V_1, V_2, \ldots, V_L) : (k_1, k_2, \ldots, k_N) \mapsto \sum_{s \in \mathcal{I}_{L,N}} V_1(k_{s(1)}) V_2(k_{s(2)}) \ldots V_L(k_{s(L)}). \qquad (2.5)$$

Note that $\xi(V_1, V_2, \ldots, V_L) = \frac{N!}{(N-L)!} f$. Since $\xi(V_1, V_2, \ldots, V_L)$ is symmetric, Lemma 2.3 ensures the existence of a unique function $\tilde{\xi}(V_1, V_2, \ldots, V_L)$ on $\mathcal{E}_{K,N}$ given by

$$\tilde{\xi}(V_1, V_2, \ldots, V_L) = \Phi_{K,N} \, \xi(V_1, V_2, \ldots, V_L). \qquad (2.6)$$

The following two identities are thus satisfied:

$$\xi(V_1, V_2, \ldots, V_L) = \tilde{\xi}(V_1, V_2, \ldots, V_L) \circ \phi_{K,N}, \quad \tilde{\xi}(V_1, V_2, \ldots, V_L) = \xi(V_1, V_2, \ldots, V_L) \circ \psi_{K,N}, \quad (2.7)$$

where $\phi_{K,N}$ and $\psi_{K,N}$ are defined in (2.2) and (1.6), respectively.

The next result provides recursive expressions for the functions $\xi(V_1, \ldots, V_L)$ and $\tilde{\xi}(V_1, \ldots, V_L)$, for $L \in [N]$. Furthermore, we prove that $\tilde{V}_\eta$, as defined by (1.9), is a polynomial of total degree $|\eta|$, for $\eta \in \bigcup_{L=1}^N \mathcal{E}_{K-1,L}$.

Lemma 2.4. *The following properties are verified:*

(a) *For $L = 1$: if $V_1 = [a_1, a_2, \ldots, a_K]^T$ is non-proportional to U_0, then $\xi(V_1)$ and $\tilde{\xi}(V_1)$, defined by (2.5) and (2.6), satisfy*

$$\xi(V_1) : (k_1, k_2, \ldots, k_N) \mapsto \sum_{i=1}^N V_1(k_i),$$

$$\tilde{\xi}(V_1) : (\eta(1), \eta(2), \ldots, \eta(K)) \mapsto \sum_{j=1}^K a_j \eta(j). \qquad (2.8)$$

(b) *For any L, $2 \leq L \leq N-1$: if the L vectors $V_i = [a_{i,1}, a_{i,2}, \ldots, a_{i,K}]^T$, $1 \leq i \leq L$, are non-proportional to U_0, then $\xi(V_1, \ldots, V_L)$ and $\tilde{\xi}(V_1, \ldots, V_L)$ satisfy*

$$\xi(V_1, \ldots, V_L) = \xi(V_1, \ldots, V_{L-1})\xi(V_L) - \sum_{i=1}^{L-1} \xi(V_1, \ldots, V_{i-1}, V_i \odot V_L, V_{i+1}, \ldots, V_{L-1}),$$

$$\tilde{\xi}(V_1, \ldots, V_L) = \tilde{\xi}(V_1, \ldots, V_{L-1})\tilde{\xi}(V_L) - \sum_{i=1}^{L-1} \tilde{\xi}(V_1, \ldots, V_{i-1}, V_i \odot V_L, V_{i+1}, \ldots, V_{L-1}),$$

where $V_i \odot V_L$ stands for the Hadamard (componentwise) product of the vectors V_i and V_L.

In particular, when $L = 2$ and the two vectors $V_1 = [a_1, a_2, \ldots, a_K]^T$ and $V_2 = [b_1, b_2, \ldots, b_K]^T$ are non-proportional to U_0, then $\tilde{\xi}(V_1, V_2)$ is the quadratic polynomial given by

$$\tilde{\xi}(V_1, V_2) = \tilde{\xi}(V_1)\tilde{\xi}(V_2) - \tilde{\xi}(V_1 \odot V_2). \tag{2.9}$$

(c) For any L, $1 \leq L \leq N$: if the L vectors $V_i = [a_{i,1}, a_{i,2}, \ldots, a_{i,K}]^T$, $1 \leq i \leq L$, are non-proportional to U_0, then $\tilde{\xi}(V_1, V_2, \ldots, V_L)$ is a polynomial of total degree L satisfying

$$\tilde{\xi}(V_1, V_2, \ldots, V_L) = \prod_{i=1}^{L} \tilde{\xi}(V_i) + q, \tag{2.10}$$

where q is a polynomial of total degree strictly less than L. In particular, $\tilde{V}_\eta$, as defined by (1.9), is a polynomial of total degree $|\eta|$, for $\eta \in \bigcup_{L=0}^{N} \mathcal{E}_{K-1,L}$.

The proof of Lemma 2.4 can be found in Appendix A.

The following result helps us to construct from a given basis of $\mathbb{R}^K$, the three bases for the three vector spaces $\mathbb{R}^{[K]^N}$, $\mathrm{Sym}(\mathbb{R}^{[K]^N})$ and $\mathbb{R}^{\mathcal{E}_{K,N}}$, respectively.

Proposition 2.5. *Let U_0 be the all-one vector in $\mathbb{R}^K$ and $U_1, U_2, \ldots, U_{K-1} \in \mathbb{R}^K$ such that*

$$\mathcal{U} = \{U_0, U_1, \ldots, U_{K-1}\}$$

is a basis of $\mathbb{R}^K$. The following statements are true:

a) *$\mathcal{U}^N$, defined as $\mathcal{U}^N := \{W_1 \otimes W_2 \otimes \cdots \otimes W_N,$ where $W_i \in \mathcal{U}$, for $i \in [N]\}$ is a basis of $\mathbb{R}^{[K]^N}$.*

b) *$\mathcal{S}^N$, defined as*

$$\mathcal{S}^N := \{\underbrace{U_0 \otimes \cdots \otimes U_0}_{N \ times}\} \cup \bigcup_{L=1}^{N} \{V_\eta, \eta \in \mathcal{E}_{K-1,L},\}$$

where V_η is defined by (1.8), is a basis of $\mathrm{Sym}\left(\mathbb{R}^{[K]^N}\right)$.

c) *$\tilde{\mathcal{S}}^N$, defined as*

$$\tilde{\mathcal{S}}^N := \{\underbrace{U_0 \otimes \cdots \otimes U_0}_{K \ times}\} \cup \bigcup_{L=1}^{N} \{\tilde{V}_\eta, \eta \in \mathcal{E}_{K-1,L}\}, \tag{2.11}$$

where $\tilde{V}_\eta$ is defined by (1.9), is a basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$.

The proof of Proposition 2.5 is deferred to Appendix A.

3. Spectrum of the neutral multi-allelic Moran process

The main goal of this section is to prove Theorem 1.4. In Section 3.1 we prove Theorem 1.1 describing the set of eigenvalues of the composition chain $\mathcal{Q}_N$ in terms of the eigenvalues of Q. Moreover, we construct right eigenvectors of $\mathcal{Q}_N$ using the symmetrised tensor product of right eigenvectors of Q. Later, in Section 3.2 we prove Theorem 1.3. Using the results in these two sections we prove Theorem 1.4 in Section 3.3.

3.1. *Proof of Theorem 1.1.* As we commented in Section 2, the N particles in the neutral multi-allelic Moran type process can be considered distinguishable or indistinguishable. Throughout the paper we suppose that Q is irreducible. Thus, 0 is a simple eigenvalue of Q with eigenvector U_0. The generator for the distinguishable case, denoted by $\mathcal{D}_N$, acts on a real function f on $[K]^N$ as follows

$$\left(\mathcal{D}_N f\right)(k_1, k_2, \ldots, k_N) := \sum_{i=1}^{N} \sum_{k=1}^{K} \mu_{k_i, k} \left[f(k_1, \ldots k_{i-1}, k, k_{i+1}, \ldots, k_N) - f(k_1, \ldots, k_N) \right],$$

for all $(k_1, k_2, \ldots, k_N) \in [K]^N$. If the function is given in a tensor product form, we get

$$\mathcal{D}_N \left(V_1 \otimes V_2 \otimes \cdots \otimes V_N \right) = \sum_{n=1}^{N} V_1 \otimes V_2 \otimes \cdots \otimes Q V_n \otimes \cdots \otimes V_N, \tag{3.1}$$

where $Q V_n(k) := \sum_{r=1}^{K} \mu_{k,r} V_n(r) = \sum_{r=1}^{K} \mu_{k,r}(V_n(r) - V_n(k))$, for all $k \in [K]$.

Remark 3.1 ($\mathcal{D}_N$ as a Kronecker sum). In fact, the infinitesimal generator satisfies $\mathcal{D}_N = Q \oplus Q \oplus \cdots \oplus Q$, where $\oplus$ denotes the Kronecker sum. The well-known relationship between the exponential of a Kronecker sum and the Kronecker product of exponential matrices, namely:

$$\exp\{Q \oplus Q \oplus \cdots \oplus Q\} = \exp\{Q\} \otimes \exp\{Q\} \otimes \cdots \otimes \exp\{Q\},$$

makes clearer the idea that $\mathcal{D}_N$ is the infinitesimal generator of the system of N particles moving independently according to the infinitesimal generator Q. See (Pease, 1965, Ch. XIV) and (Davis, 1979, §2.2) for further details on the Kronecker sum.

The Markov chain generated by $\mathcal{D}_N$ is usually called *product chain*. The infinitesimal generator $\mathcal{D}_N$ inherits its spectral properties from those of Q. Namely, if π is the stationary distribution of Q, then $\pi \otimes \pi \otimes \cdots \otimes \pi$ is the stationary distribution of $\mathcal{D}_N$. Moreover, if $V_1, V_2, \ldots, V_N$ are N (not necessarily distinct) eigenvectors of Q, then $V_1 \otimes V_2 \otimes \cdots \otimes V_N$ is an eigenvector of $\mathcal{D}_N$. Consequently, if Q is diagonalisable, then $\mathcal{D}_N$ is also diagonalisable and the tensors products of vectors in an eigen-basis of Q form an eigen-basis of $\mathcal{D}_N$, as in Proposition 2.5-(a). In particular, if $\lambda_0 = 0, \lambda_1, \ldots, \lambda_{K-1}$ are the K complex eigenvalues of Q, then the eigenvalues of $\mathcal{D}_N$ are given by the sums of eigenvalues of Q, i.e. the spectrum of $\mathcal{D}_N$ is

$$\left\{ z_0 + z_1 + \cdots + z_{K-1} : z_i \in \{\lambda_0, \lambda_1, \ldots, \lambda_{K-1}\} \right\}.$$

See Sections 12.4 and 20.4 in (Levin and Peres, 2017) for the proofs of these results and more details on product chains.

When the N particles are considered indistinguishable, the infinitesimal generator of the Markov chain, denoted by $\mathcal{Q}_N$, is that defined by (1.3), i.e.

$$\left(\mathcal{Q}_N f\right)(\eta) = \sum_{i,j \in [K]} \eta(i) \mu_{i,j} \left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta) \right],$$

for all $\eta \in \mathcal{E}_{K,N}$ and for every function f on $\mathcal{E}_{K,N}$. Zhou and Lange (2009) noticed that $\mathcal{Q}_N$ is a lumped chain of $\mathcal{D}_N$ and used this fact to study the relationship between the spectral properties of both chains. They studied the eigenvalues and the left eigenfunctions of $\mathcal{Q}_N$. In particular, they proved that the stationary distribution of $\mathcal{Q}_N$ is multinomial with probability vector π, denoted $\mathcal{M}(\cdot \mid N, \pi)$, where π is the unique stationary probability of Q. Our approach differs from that on (Zhou and Lange, 2009): we study the right eigenfunctions of $\mathcal{Q}_N$ using the connections between the real functions on $\mathcal{E}_{K,N}$ and the symmetric real functions on $[K]^N$ studied in Section 2. In addition, our methods allow us to explicitly describe the spectrum of $\mathcal{Q}_N$, for every mutation matrix Q generating an irreducible process, even when Q is non-diagonalisable. We first study the relationship between the generators $\mathcal{Q}_N$ and $\mathcal{D}_N$ through the operator $\Phi_{K,N}$.

Lemma 3.2 (Link between the generators $\mathcal{Q}_N$ and $\mathcal{D}_N$). *For any symmetric function ξ on $[K]^N$, the function $\mathcal{D}_N \xi$ is also symmetric. In addition,*

$$\mathcal{Q}_N \left(\Phi_{K,N} \, \xi \right) = \Phi_{K,N} \left(\mathcal{D}_N \, \xi \right),$$

where $\Phi_{K,N}$ is defined by (2.3).

Proof: The symmetry of $\mathcal{D}_N \xi$ is a consequence of the symmetry of ξ and the linearity of $\mathcal{D}_N$.

For $\eta \in \mathcal{E}_{K,N}$ let us define $(k_1, k_2, \ldots, k_N) = \psi_{K,N}(\eta)$, i.e. k_i is the position on $[K]$ of the i-th particle according to the definition of $\psi_{K,N}$. We have

$$\left(\mathcal{D}_N \, \xi \circ \psi_{K,N} \right)(\eta) = \sum_{i=1}^{N} \sum_{k=1}^{K} \mu_{k_i,k} \Big[\xi(k_1, \ldots, k_{i-1}, k, k_{i+1}, \ldots, k_N) - \xi\big(\psi_{K,N}(\eta)\big) \Big]$$

$$= \sum_{k=1}^{K} \sum_{r=1}^{K} \sum_{i: \, k_i = r} \mu_{k_i,k} \Big[\xi(k_1, \ldots, k_{i-1}, k, k_{i+1}, \ldots, k_N) - \xi\big(\psi_{K,N}(\eta)\big) \Big].$$

Using the symmetry of ξ, for all η such that $\psi_{K,N}(\eta)(i) = r$ we obtain

$$\xi(k_1, \ldots, k_{i-1}, k, k_{i+1}, \ldots, k_N) - \xi(\psi_{K,N}(\eta)) = \xi(\psi_{K,N}(\eta - \mathbf{e}_r + \mathbf{e}_k)) - \xi(\psi_{K,N}(\eta)).$$

Thus,

$$\left(\mathcal{D}_N \, \xi \circ \psi_{K,N} \right)(\eta) = \sum_{k=1}^{K} \sum_{r=1}^{K} \sum_{i: \, k_i = r} \mu_{k_i,k} \Big[\xi\big(\psi_{K,N}(\eta - \mathbf{e}_r + \mathbf{e}_k)\big) - \xi\big(\psi_{K,N}(\eta)\big) \Big]$$

$$= \sum_{k=1}^{K} \sum_{r=1}^{K} \eta(r) \mu_{r,k} \Big[\xi\big(\psi_{K,N}(\eta - \mathbf{e}_r + \mathbf{e}_k)\big) - \xi\big(\psi_{K,N}(\eta)\big) \Big]$$

$$= \left(\mathcal{Q}_N \, \xi \circ \psi_{K,N} \right)(\eta),$$

for every $\eta \in \mathcal{E}_{K,N}$. $\square$

The following lemma describes all the eigenvalues of $\mathcal{Q}_N$, defined by (1.3), in the case where the mutation matrix is diagonalisable.

Lemma 3.3 (Eigenvalues of $\mathcal{Q}_N$ for diagonalisable Q). *Assume Q is diagonalisable and*

$$\mathcal{U} = \{U_0, U_1, \ldots, U_{K-1}\}$$

is the basis of $\mathbb{R}^K$ formed by right eigenvectors of Q, such that U_0 is the all-one vector. Consider $\tilde{V}_\eta$ and λ_η defined as in (1.9) and (1.10), respectively. Then

(a) *λ_η is an eigenvalue of $\mathcal{Q}_N$ with right eigenvector $\tilde{V}_\eta$.*

(b) *The spectrum of $\mathcal{Q}_N$ is formed by 0 and all λ_η for $\eta \in \bigcup\limits_{L=1}^{N} \mathcal{E}_{K-1,L}$.*

(c) *$\mathcal{Q}_N$ is diagonalisable.*

Proof: (a) For $\eta \in \mathcal{E}_{K-1,L}$ let us denote U_η as in (1.7). Because $Q U_0 = 0$ and $Q U_k = \lambda_k U_k$, $1 \leq k \leq K - 1$, from (3.1), we get $\mathcal{D}_N(U_\eta) = \lambda_\eta U_\eta$. More generally, for every permutation $\sigma \in \mathcal{S}_N$, $\mathcal{D}_N(\sigma U_\eta) = \lambda_\eta(\sigma U_\eta)$, and thus, using the linearity of $\mathcal{D}_N$ we get

$$\mathcal{D}_N \, V_\eta = \lambda_\eta V_\eta,$$

where V_η is defined as in (1.8). Applying $\psi_{K,N}$ to both members of the previous equality we obtain $(\mathcal{D}_N V_\eta) \circ \psi_{K,N} = \lambda_\eta V_\eta \circ \psi_{K,N}$. Now, using Lemma (3.2), and the expressions (1.8) and (1.9), definitions of V_η and $\tilde{V}_\eta$, respectively, we obtain $\mathcal{Q}_N \tilde{V}_\eta = \lambda_\eta \tilde{V}_\eta$, which proves (a).

(b)-(c) Because $\mathcal{U}$ is a basis of $\mathbb{R}^K$, the set $\tilde{\mathcal{S}}^N$ as defined in (2.11) is a basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$, due to Proposition 2.5-(c). Therefore, all the eigenvalues of $\mathcal{Q}_N$ are those described in part (b) and $\mathcal{Q}_N$ is diagonalisable. $\square$

Remark 3.4. Note that the results in Lemma 3.3 remains valid for all operator $\mathcal{Q}_N$ defined using a diagonalisable matrix Q, not necessarily a rate matrix, with complex entries and such that $QU_0 = \mathbf{0}$ and $\lambda_0 = 0$ has algebraic multiplicity equal to one.

Lemma 3.3 provides all the eigenvalues and right eigenvectors of $\mathcal{Q}_N$ when Q is diagonalisable. However, an ergodic rate matrix is not necessarily diagonalisable. Thereby, we want to extend the results in Lemma 3.3 to the case where the matrix Q is non-diagonalisable, as stated in Theorem 1.1. Let us first recall two known facts in the theory of real matrices. We denote by $M_n(\mathbb{R})$ and $M_n(\mathbb{C})$ the vector space of n-dimensional real and complex matrices, respectively. For a matrix $M \in M_n(\mathbb{C})$ we denote by $\mathrm{Spec}(M) \in \mathbb{C}^n$ its spectrum counting the algebraic multiplicities of the eigenvalues. It is known that the set of diagonalisable complex matrices is dense in $M_n(\mathbb{C})$. Serre Serre (2010, Cor. 5.1), for instance, proves this result as a consequence of Schur's Theorem (Serre, 2010, Thm. 5.1). Using the same reasoning we can prove the following:

Fact 1: The set of diagonalisable complex matrices with each row summing to zero is dense in the set of the irreducible rate matrices: for every rate matrix $Q \in M_n(\mathbb{R})$ and $\epsilon > 0$ there exists a diagonalisable matrix $\bar{Q} \in M_n(\mathbb{C})$ such that $\|Q - \bar{Q}\| < \epsilon$. Moreover, $\bar{Q}$ can be chosen such that $0 \in \mathrm{Spec}(\bar{Q})$, with 0 having geometric multiplicity 1 and $\bar{Q} U_0 = \mathbf{0}$, where $\mathbf{0}$ denotes the K dimensional null column vector, i.e. each row of $\bar{Q}$ sums to zero.

The idea of the proof of Fact 1 is to modify diagonal elements in the upper-triangular matrix obtained by Schur's Theorem to get a matrix with n different eigenvalues, and thus diagonalisable. Indeed, since Q is an irreducible rate matrix, the eigenspace associated to the eigenvalue $\lambda_0 = 0$ has dimension one and it is generated by U_0. Moreover, the other $n - 1$ complex eigenvalues have strictly negative real parts. Thus, it is possible to modify the diagonal of the upper triangular matrix obtained by the Schur's Theorem in such a way that the eigenvalues of the modified matrix, denoted $\bar{Q}$, are zero and $n - 1$ complex numbers with different and strictly negative real parts. Furthermore, because of the Schur's factorisation, U_0 is also an eigenvector of $\bar{Q}$ associated to the null eigenvalue, i.e. $\bar{Q} U_0 = \mathbf{0}$.

Note that, since $M_n(\mathbb{C})$ is a finite dimensional vector space, the result in Fact 1 holds for every norm defined on $M_n(\mathbb{C})$. In the sequel we will use the uniform norm, denoted $\|\cdot\|_{\mathrm{Unif}}$, and defined as follows

$$\|A\|_{\mathrm{Unif}} := \max_{i,j} |a_{i,j}|,$$

for every matrix $A = (a_{i,j})_{i,j} \in M_n(\mathbb{C})$.

The second fact is related to the continuity of the eigenvalues of a matrix with respect to its entries. Consider the following distance between two sets of n elements in $\mathbb{C}$:

$$\mathrm{D}\left(\{z_i\}_{i=1}^n, \{\omega_i\}_{i=1}^n\right) := \inf_{\sigma \in \mathcal{S}_n} \max_j |z_j - \omega_{\sigma(j)}|,$$

where $\mathcal{S}_n$ denotes de symmetric group on $[n]$, for every $n \in \mathbb{N}$.

Fact 2: The eigenvalues are continuous with respect to the entries of the matrix in the following sense: consider $M \in M_n(\mathbb{C})$, then for all $\epsilon > 0$ there exists a $\delta > 0$ such that for every matrix $N \in M_n(\mathbb{C})$ such that $\|M - N\| < \delta$, then $\mathrm{D}\left(\mathrm{Spec}(M), \mathrm{Spec}(N)\right) < \epsilon$.

See e.g. (Harris and Martin, 1987) and (Serre, 2010, Thm. 5.2) for a proof of Fact 2.

Proof of Theorem 1.1: From Lemma 3.3 we know that the statement of Theorem 1.1 holds for a diagonalisable rate matrix Q. Let us prove it in the general case using the Facts 1 and 2 we previously discussed.

For a mutation rate matrix $Q \in M_K(\mathbb{R})$ with spectrum $\mathrm{Spec}(Q) = \{0, \lambda_1, \ldots, \lambda_{K-1}\}$, let us define by $\sigma_N(Q)$ the set formed by 0 and λ_η, for $\eta \in \bigcup_{L=1}^{K-1} \mathcal{E}_{K-1,L}$, where the values λ_k in the definition (1.10) of λ_η are those in $\mathrm{Spec}(Q)$. Then, proving Theorem 1.1-(a) is equivalent to prove that $\sigma_N(Q)$ is the spectrum of $\mathcal{Q}_N$, i.e. $\mathrm{D}\left(\mathrm{Spec}(\mathcal{Q}_N), \sigma_N(Q)\right) = 0$.

For a matrix $\bar{Q} \in M_K(\mathbb{C})$ whose rows sum to zero (not necessarily a rate matrix), let us define $\bar{\mathcal{Q}}_N$ similarly to the definition of $\mathcal{Q}_N$ (1.3), but with $\bar{Q}$ as mutation matrix instead of Q. As we commented in Remark 3.4, Lemma 3.3 remains valid and it ensures us that $\mathrm{Spec}(\bar{\mathcal{Q}}_N) = \sigma_N(\bar{Q})$. Thus, using the triangular inequality we get

$$\mathrm{D}\left(\mathrm{Spec}(\mathcal{Q}_N), \sigma_N(Q)\right) \leq \mathrm{D}\left(\mathrm{Spec}(\mathcal{Q}_N), \mathrm{Spec}(\bar{\mathcal{Q}}_N)\right) + \mathrm{D}\left(\mathrm{Spec}(\bar{\mathcal{Q}}_N), \sigma_N(Q)\right).$$

Moreover,

$$\|\mathcal{Q}_N - \bar{\mathcal{Q}}_N\|_{\mathrm{Unif}} \leq N\|Q - \bar{Q}\|_{\mathrm{Unif}},$$
$$\mathrm{D}\left(\mathrm{Spec}(\bar{\mathcal{Q}}_N), \sigma_N(Q)\right) \leq N\,\mathrm{D}\left(\mathrm{Spec}(\bar{Q}), \mathrm{Spec}(Q)\right).$$

Fix $\epsilon > 0$. Using Fact 2, we know there exist $\delta_1, \delta_2 > 0$ such that

$$\mathrm{D}\left(\mathrm{Spec}(\mathcal{Q}_N), \mathrm{Spec}(\bar{\mathcal{Q}}_N)\right) \leq \frac{\epsilon}{2} \quad \text{if} \quad \|\mathcal{Q}_N - \bar{\mathcal{Q}}_N\|_{\mathrm{Unif}} < \delta_1,$$
$$\mathrm{D}\left(\mathrm{Spec}(\bar{Q}), \mathrm{Spec}(Q)\right) \leq \frac{\epsilon}{2N} \quad \text{if} \quad \|Q - \bar{Q}\|_{\mathrm{Unif}} < \delta_2.$$

Thus,

$$\mathrm{D}\left(\mathrm{Spec}(\mathcal{Q}_N), \sigma_N(Q)\right) \leq \frac{\epsilon}{2} + N\,\mathrm{D}(\mathrm{Spec}(\bar{Q}), \mathrm{Spec}(Q)) < \epsilon,$$

whenever $\|Q - \bar{Q}\|_{\mathrm{Unif}} < \min\{\delta_1/N, \delta_2\}$. Since ϵ can be taken arbitrary small, by Fact 1, the proof of (a) is finished.

The proof of (b) is exactly the same as the proof of (a) in Lemma 3.3. Note that, since $\eta(r) = \cdots = \eta(K-1) = 0$, the definition of $\tilde{V}_\eta$ only depends on the r linearly independent vectors forming $\mathcal{U}$. Finally, the result in (c) trivially comes from Lemma 3.3. $\qquad\square$

Remark 3.5 (Alternative proof for Theorem 1.1). The *Jordan–Chevalley decomposition* is an elegant tool to find the eigenvalues of $\mathcal{Q}_N$ and prove Theorem 1.1. The Jordan-Chevalley decomposition ensures the existence of two matrices Q_{Diag} and Q_{Nil} such that $Q = Q_{\mathrm{Diag}} + Q_{\mathrm{Nil}}$. Moreover, Q_{Diag} is diagonalisable, Q_{Nil} is nilpotent, they commute and such a decomposition is unique. See (Serre, 2010, Prop. 3.20) and (Couty et al., 2011) for more details about the Jordan-Chevalley decomposition. Then, it can be proved that the Jordan-Chevalley decomposition of $\mathcal{Q}_N$ is $\mathcal{Q}_N = (\mathcal{Q}_{\mathrm{Diag}})_N + (\mathcal{Q}_{\mathrm{Nil}})_N$, where $(\mathcal{Q}_{\mathrm{Diag}})_N$ and $(\mathcal{Q}_{\mathrm{Nil}})_N$ are defined similarly to $\mathcal{Q}_N$ in (1.3), substituting Q by Q_{Diag} and Q_{Nil}, respectively. Now, since the spectrum of $\mathcal{Q}_N$ is that of $(\mathcal{Q}_{\mathrm{Diag}})_N$, the proof of Theorem 1.1 follows from Lemma 3.3.

3.2. *Proof of Theorem 1.3.* In this section, given $K \geq 2$ and $N \geq 2$, we consider the continuous-time Markov chain of N indistinguishable particles on K sites, with state space $\mathcal{E}_{K,N}$, where, with rate 1, any particle jumps to one of the positions of another particle chosen at random. We denote by $\mathcal{A}_N$ the infinitesimal generator of this reproduction process, which is defined in (1.4) as

$$\left(\mathcal{A}_N f\right)(\eta) = \sum_{i,j \in [K]} \eta(i)\eta(j)\left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)\right]$$

for every real function f and all $\eta \in \mathcal{E}_{K,N}$.

Remark 3.6 (First degree eigenfunctions of $\mathcal{A}_N$). Note that the states $\{N\,\mathbf{e}_k\}_{k=1}^K \subset \mathcal{E}_{K,N}$ are the only absorbing states for the interaction process generated by $\mathcal{A}_N$. Thus, the distribution concentrated at $N\,\mathbf{e}_k$, denoted $\delta_{\{N\mathbf{e}_k\}}$, is stationary for $\mathcal{A}_N$, for $k \in [K]$. It is not difficult to check that the real functions on $\mathcal{E}_{K,N}$, $x_0 \equiv 1$ and $x_k : \eta \mapsto \eta_k$, for $k \in [K-1]$, are linearly independent vectors of $\mathbb{R}^{\mathcal{E}_{K,N}}$ and they satisfy $\mathcal{A}_N x_k = 0$, for all $k = 0, 1, \ldots, K-1$. Thus, the right eigenspace associated to 0 is the space of homogeneous polynomials of degree 1, which has dimension K.

Actually, it can be proved that the generator $\mathcal{A}_N$ preserves the total degree of a polynomial, in the sense that the image of a polynomial is another polynomial of the same total degree. To prove Theorem 1.3 we first formally describe the preserving degree polynomial property satisfied by $\mathcal{A}_N$.

Lemma 3.7 ($\mathcal{A}_N$ preserves polynomial total degree). *Assume $K \geq 2$ and $N \geq 2$. Let P be a polynomial on $\mathcal{E}_{K,N}$ of total degree L with $1 \leq L \leq N$. Then,*

$$\mathcal{A}_N V_P = -L(L-1)V_P + V_R,$$

where R is a polynomial with a total degree strictly less than L.

The proof of Lemma 3.7 is technical and it is deferred to Appendix B. We next prove Theorem 1.3.

Proof of Theorem 1.3: (a) For $K \geq 2$ and $N \geq 2$, let us define the sets $\mathcal{B}_L$ of monomials in $\mathcal{E}_{K,N}$ as follows

$$\mathcal{B}_0 := \{1\}, \quad \mathcal{B}_1 := \{x_1, x_2, \ldots, x_{K-1}\}, \quad \mathcal{B}_L := \{\mathbf{x}^\alpha, \ \alpha \in \mathcal{E}_{K-1,L}\},$$

for $2 \leq L \leq N$, where $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \ldots x_{K-1}^{\alpha_{K-1}}$ for $\alpha := (\alpha_1, \alpha_2, \ldots, \alpha_{K-1})$. Then, consider the ordered set

$$\mathcal{B} = \mathcal{B}_0 \cup \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_N.$$

The set $\mathcal{B}$ is a basis of the space of real functions on $\mathcal{E}_{K,N}$, due to Lemma 2.1-(b). The matrix similar to $\mathcal{A}_N$ with respect to this basis is $\bar{\mathcal{A}}_N = W^{-1}\mathcal{A}_N W$, where W is the matrix with P, with $P \in \mathcal{B}$, as column vectors. Thanks to the result in Lemma 3.7-(a), $\bar{\mathcal{A}}_N$ is a block upper triangular matrix, where the first diagonal block has size K and is a null matrix. The other diagonal blocks have size $\mathrm{Card}(\mathcal{E}_{K-1,L}) = \binom{K-2+L}{L}$ and are diagonal matrices with constant diagonal elements equal to $-L(L-1)$, with $2 \leq L \leq N$. This analysis gives us the eigenvalues of $\mathcal{A}_N$ are 0 with algebraic multiplicity K and $-L(L-1)$ with algebraic multiplicity $\binom{K-2+L}{L}$ for $2 \leq L \leq N$.

Now, using the block multiplication of matrices, it is not difficult to see that $(\bar{\mathcal{A}}_N)^n$ is also a block diagonal matrix, where the L-th block is a diagonal matrix of dimension $\binom{K-2+L}{L}$ with all the entries on the diagonal equal to $(-L(L-1))^n$, for $2 \leq L \leq N$. Thus, for every real polynomial Υ the matrix $\Upsilon(\bar{\mathcal{A}}_N) = W^{-1}\Upsilon(\mathcal{A}_N)W$ is a block diagonal matrix with diagonal elements $\Upsilon(-L(L-1))$. Taking

$$\Upsilon : s \mapsto s \prod_{L=2}^{N} [s + L(L-1)],$$

we get $\Upsilon(\bar{\mathcal{A}}_N) = \mathbf{0}_{K,N}$, where $\mathbf{0}_{K,N}$ is the $\binom{K-1+N}{N}$ dimensional null matrix. Thus, $\Upsilon(\mathcal{A}_N) = \mathbf{0}_{K,N}$ and Υ is necessarily the *minimal polynomial* of $\mathcal{A}_N$, which factors into distinct linear factors. We thus conclude that $\mathcal{A}_N$ is diagonalisable. $\qquad\square$

Remark 3.8 (On the right eigenfunctions of $\mathcal{A}_N$). Theorem 1.3 does not provide a characterisation of the eigenspace associated to the eigenvalue $-L(L-1)$, for $L \in [N]$. For the special case $K = 2$, Watterson (1961) does provide such a decomposition for the discrete analogue of $\mathcal{A}_N$ in terms of cumulative sums of discrete Chebyshev polynomials. In addition, Zhou (2008, §4.2.2) provides an equivalent but simpler expression for the eigenvectors of the equivalent analogous of $\mathcal{A}_N$, for $K = 2$, in terms of univariate Hahn polynomials. In the general case ($K \geq 3$), it is possible to describe the eigenspaces associated to the first three eigenvalues of $\mathcal{A}_N$. As we commented in Remark 3.6, the right eigenspace associated to 0 is the space of homogeneous polynomials of first degree. Moreover, the right eigenspace associated to -2 has dimension $K(K-1)/2$ and it is generated by the set of monomials $\{x_k x_r, \ 1 \leq k < r \leq K\}$. Additionally, for $L = 3$, it is possible to prove that a simple basis of the right eigenspace associated to -6 has dimension $K(K+1)(K-1)/6$ and is given by eigenvectors $\{x_k^2 x_r - x_k x_r^2, \ 1 \leq k < r \leq K\} \cup \{x_k x_r x_s, \ 1 \leq k < r < s \leq K\}$. The complete characterisation of the eigenvectors of $\mathcal{A}_N$, for $K \geq 3$, is a topic of further research.

3.3. *Proof of Theorem 1.4.* This section is devoted to the proof of Theorem 1.4 providing a description of the spectrum of the neutral multi-allelic Moran process with generator $\mathcal{Q}_{N,p}$, defined by (1.2) as

$$\left(\mathcal{Q}_{N,p}f\right)(\eta) = \sum_{i,j\in[K]} \eta(i)\left(\mu_{i,j} + \frac{p}{N}\eta(j)\right)\left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)\right],$$

for every real function f in $\mathcal{E}_{K,N}$ and every $\eta \in \mathcal{E}_{K,N}$.

Assume $K \geq 2$, $N \geq 2$ and $p \in [0,\infty)$ and suppose that Q is diagonalisable with eigenvalues 0 and λ_k, for $k \in [K-1]$. For any $\eta \in \mathcal{E}_{K-1,L}$, with $L \in [N]$, let us recall the definition of $\lambda_{\eta,p}$:

$$\lambda_{\eta,p} = -L(L-1)\frac{p}{N} + \sum_{k=1}^{K-1} \eta(k)\lambda_k. \tag{3.2}$$

Then, we will prove that the eigenvalues of $\mathcal{Q}_{N,p}$ are 0 and all $\lambda_{\eta,p}$ for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$.

Proof of Theorem 1.4: Recall that $\mathcal{Q}_{N,p} = \mathcal{Q}_N + \frac{p}{N}\mathcal{A}_N$, where $\mathcal{Q}_N$ and $\mathcal{A}_N$ are the generators of the mutation and the reproduction processes defined by (1.3) and (1.4), respectively.

Let us first prove the result when the mutation rate matrix Q is diagonalisable. As proved in Lemma 3.3, the vector $\tilde{V}_\eta$ is an eigenvector of $\mathcal{Q}_N$ with eigenvalue λ_η, for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$. Let us denote by $\tilde{V}_0$ the all-one vector in $\mathbb{R}^{\mathcal{E}_{K,N}}$. Then, the set $\mathcal{B} = \{V_0\} \cup \{\tilde{V}_\eta,\ \eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}\}$ is a basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$, thanks to Proposition 2.5-(c). Let us denote by W the matrix with the elements of $\mathcal{B}$ as columns such that $W^{-1}\mathcal{Q}_N W$ is a diagonal matrix with diagonal entries equal to 0 and λ_η, for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$.

For $1 \leq L \leq N$ and $\eta \in \mathcal{E}_{K-1,L}$, the expression (1.9) and Lemma 2.4-(c) ensure that $\tilde{V}_\eta$ is a polynomial of total degree equal to L. Using now Theorem 1.3-(b), we get

$$\mathcal{A}_N \tilde{V}_\eta = -L(L-1)\tilde{V}_\eta + R,$$

where R is a polynomial of total degree strictly less than L. This fact means that, like in Theorem 1.3-(c), $W^{-1}\mathcal{A}_N W$ is a block upper triangular matrix, where the diagonal blocks of size $\mathrm{Card}(\mathcal{E}_{K-1,L}) = \binom{K+L-2}{L}$ are diagonal matrices with constant diagonal elements equal to $-L(L-1)$, for $2 \leq L \leq N$. The first diagonal block of size K is a null matrix. It follows that

$$W^{-1}\mathcal{Q}_{N,p}W = W^{-1}\mathcal{Q}_N W + \frac{p}{N}W^{-1}\mathcal{A}_N W$$

is a block upper triangular matrix, where the first diagonal block has dimension one and is null, i.e. the first column is null. Moreover, the L-th diagonal block has dimension $\binom{K+L-2}{L}$ and its diagonal elements are the eigenvalues $\lambda_{\eta,p}$ with $\eta \in \mathcal{E}_{K-1,L}$, for $L \in [N]$. Thus, these are the eigenvalues of $\mathcal{Q}_{N,p}$.

Now, consider a general mutation matrix $Q \in M_K(\mathbb{R})$, not necessarily diagonalisable, with spectrum $\mathrm{Spec}(Q) = \{0, \lambda_1, \ldots, \lambda_{K-1}\}$. Let us define by $\sigma_{N,p}(Q)$ the set formed by 0 and $\lambda_{\eta,p}$, for $\eta \in \bigcup_{L=1}^{K-1} \mathcal{E}_{K-1,L}$, where the values λ_k in the definition (3.2) of $\lambda_{\eta,p}$, are those in $\mathrm{Spec}(Q)$. Define $\bar{\mathcal{Q}}_{N,p}$ similarly to (1.2) but with a diagonalisable matrix $\bar{Q} \in M_K(\mathbb{C})$, whose rows have null sum (not necessarily a rate matrix), instead of Q. Then,

$$\|\mathcal{Q}_{N,p} - \bar{\mathcal{Q}}_{N,p}\|_{\mathrm{Unif}} = \|\mathcal{Q}_N - \bar{\mathcal{Q}}_N\|_{\mathrm{Unif}},$$

$$\mathrm{D}\left(\mathrm{Spec}(\bar{\mathcal{Q}}_{N,p}), \sigma_{N,p}(Q)\right) = \mathrm{D}\left(\mathrm{Spec}(\bar{\mathcal{Q}}_N), \sigma_N(Q)\right).$$

Hence, $\sigma_{N,p}(Q)$ is proved to be the spectrum of $\mathcal{Q}_{N,p}$, analogously to the proof of Theorem 1.1-(a). □

Remark 3.9 (Alternative proof of Theorem 1.4). Another proof of Theorem 1.4 can be carried out using the Jordan form of the mutation rate matrix Q. Indeed, the vectors $\tilde{V}_\eta \in \mathbb{R}^{\mathcal{E}_{K,N}}$ can be defined

using the basis of $\mathbb{R}^K$ that transforms Q in its normal Jordan form. Then, defining a suitable order among the vectors $\tilde{V}_\eta$, for $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$, it is possible to show that $\mathcal{Q}_{N,p}$ is similar to an upper triangular matrix with the values $\lambda_{\eta,p}$ on the diagonal.

4. Applications to the convergences to stationarity

This section devoted to some applications of the results in Section 3 to the study of the ergodicity of the process driven by $\mathcal{Q}_{N,p}$ in total variation, using spectral properties of Q. In this section we prove Corollary 1.6 and Theorem 1.7. First, let us establish that the Jordan form of Q is a diagonal block in the Jordan form of $\mathcal{Q}_{N,p}$.

Corollary 4.1 (Jordan forms of Q and $\mathcal{Q}_{N,p}$). *Consider $K \geq 2$, $N \geq 2$ and $p \geq 0$. If J is the Jordan form of Q, then the Jordan normal form of $\mathcal{Q}_{N,p}$ is $J \oplus J'$, where J' is a Jordan matrix of dimension $\binom{K-1+N}{N} - K$. In particular, e^{tQ} and $\mathrm{e}^{t\mathcal{Q}_{N,p}}$ have that same second largest eigenvalue in modulus* (SLEM), *for every $t \geq 0$.*

Proof: The image by $\mathcal{Q}_{N,p}$ of a first degree polynomial is also a first degree polynomial, i.e. the space of first degree polynomials is invariant by $\mathcal{Q}_{N,p}$. Moreover, as a consequence of Lemma 3.2 we obtain

$$\mathcal{Q}_{N,p}\,\tilde{\xi}(V) = \mathcal{Q}_N\,\tilde{\xi}(V) = \Phi_{K,N}\,\mathcal{D}_N \xi(V) = \Phi_{K,N}\,\xi(QV) = \tilde{\xi}(QV).$$

Let $\mathcal{U} = \{U_0, \ldots, U_{K-1}\}$ by a Jordan basis of Q formed by generalised eigenvectors of Q. Since $\mathcal{Q}_{N,p}\,\tilde{\xi}(U_k) = \tilde{\xi}(Q\,U_k)$, for every $k \in [K-1]_0$, we have that $\{\tilde{\xi}(U_0), \ldots, \tilde{\xi}(U_{K-1})\}$ is a system of linearly independent generalised eigenvectors of $Q_{N,p}$. They are precisely the generalised eigenvectors of $\mathcal{Q}_{N,p}$ associated to the eigenvalues in $\mathrm{Spec}(Q) \subset \mathrm{Spec}(\mathcal{Q}_{N,p})$. We can complete this system to a Jordan basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$, adding the generalised eigenvectors of the other eigenvalues on $\mathcal{Q}_{N,p}$. With respect to this Jordan basis $\mathcal{Q}_{N,p}$ becomes similar to $J \oplus J'$, where J is the Jordan matrix of Q and J' is a Jordan matrix of dimension $\binom{K-1+N}{N} - K$. Note that the eigenvalues $\{\lambda_0, \lambda_1, \ldots, \lambda_{K-1}\}$ are those eigenvalues of $\mathcal{Q}_{N,p}$ of smallest modulus. We thus get that e^{tQ} and $\mathrm{e}^{t\mathcal{Q}_{N,p}}$ have the same SLEM, for every $t \geq 0$. $\qquad\square$

Every irreducible finite Markov chain convergences exponentially to stationarity, see e.g. (Levin and Peres, 2017, Thm. 4.9). In addition, the sharpest asymptotic speed of convergence is associated to the SLEM and the size of the largest Jordan block corresponding to any eigenvalue with this modulus. We recall that the size of the largest Jordan block associated to an eigenvalue λ is equal to the multiplicity of λ in the minimal polynomial of the rate matrix of the Markov chain.

Proof of Corollary 1.6: Let $\mathrm{e}^{-\rho t}$ be the SLEM of e^{tQ} and s the largest multiplicity in the minimal polynomial of e^{tQ} of all the eigenvalues with modulus $\mathrm{e}^{-\rho t}$, or equivalently, the size of the largest Jordan block associated to eigenvalues with modulus $\mathrm{e}^{-\rho t}$. Then,

$$\mathrm{D}_Q^{\mathrm{TV}}(t) = \Theta(t^{s-1}\mathrm{e}^{-\rho t}), \tag{4.1}$$

see e.g. (Szehr et al., 2015, Thm. 3.2). The result follows as a consequence of Corollary 4.1 and (4.1).

$\qquad\square$

The following example uses Corollary 1.6 to provide the rates for the exponential convergence to stationarity of the neutral multi-allelic Moran process considered by Corujo (2021).

Example 4.2 (Circulant mutation rate matrix). Consider the following mutation rate matrix

$$Q_\theta = \begin{pmatrix} -(1+\theta) & 1 & 0 & \ldots & 0 & \theta \\ \theta & -(1+\theta) & 1 & \ldots & 0 & 0 \\ 0 & \theta & -(1+\theta) & \ldots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & \theta & -(1+\theta) \end{pmatrix},$$

where $\theta \geq 0$. Q_θ is the infinitesimal generator of a simple asymmetric random walk on the K-cycle graph. The neutral multi-allelic Moran type process with mutation rate Q_θ was considered in (Corujo, 2021). Since Q_θ is circulant, it is possible to explicitly diagonalise it using the Fourier matrix. The eigenvalues of Q_θ are

$$\lambda_k = -2(1+\theta)\sin^2\left(\frac{\pi k}{K}\right) + i(1-\theta)\sin\left(\frac{2\pi k}{K}\right),$$

for $0 \leq k \leq K-1$. Thus, the SLEM of e^{tQ_θ} is $e^{-2(1+\theta)\sin^2\left(\frac{\pi}{K}\right)t}$, which is attained for two eigenvalues, each one of them with algebraic multiplicity equals to 1, for $\theta \neq 1$. When $\theta = 1$, the SLEM of e^{tQ_θ} is $e^{-4\sin^2\left(\frac{\pi}{K}\right)}$ and it is attained for a unique eigenvalue with algebraic and geometric multiplicities equal to 2. Let $\mathcal{Q}_\theta$ be the infinitesimal generator of the neutral multi-allelic Moran process with mutation rate Q_θ. Then,

$$\mathrm{D}_{\mathcal{Q}_\theta}^{\mathrm{TV}}(t) = \Theta\left(e^{-2(1+\theta)\sin^2\left(\frac{\pi}{K}\right)t}\right).$$

Example 4.3 (Convergence rate for a process with non-diagonalisable mutation rate matrix). Consider the rate matrix Q defined as

$$Q = \begin{pmatrix} -9 & 7 & 2 \\ 1 & -7 & 6 \\ 5 & 7 & -12 \end{pmatrix}$$

and $\mathcal{Q}_{N,p}$ the infinitesimal generator of the associated neutral multi-allelic Moran process with mutation rate matrix Q. Then, $\lambda_0 = 0$ and $\lambda_1 = \lambda_2 = -14$, because -14 has algebraic multiplicity 2. Then, for N fixed, the eigenvalues of $\mathcal{Q}_{N,p}$ are

$$\lambda_{L,p} := \eta(1)\lambda_1 + \eta(2)\lambda_2 - L(L-1)\frac{p}{N} = -14L - L(L-1)\frac{p}{N},$$

for $L \in [N]_0$. In addition, $\lambda_{L,p}$ has algebraic multiplicity $\mathrm{Card}(\mathcal{E}_{2,L}) = L+1$.

Note that the minimal polynomial of Q is $m_Q : s \mapsto s(s+14)^2$ and according to the notation in Corollary 1.6 we get $\rho = 14$ and $s = 2$. Then,

$$\mathrm{D}_{\mathcal{Q}_{N,p}}^{\mathrm{TV}}(t) = \Theta\left(t\,e^{-14t}\right).$$

4.1. *Proof of Theorem 1.7.* Let us denote by $\Gamma_{\mathcal{L}}$ the *carré-du-champ* operator associated to the Markov generator $\mathcal{L}$ on a state space $\mathcal{E}$, i.e.

$$\Gamma_{\mathcal{L}}f : \eta \mapsto \left(\mathcal{L}f^2\right)(\eta) - 2f(\eta)(\mathcal{L}f)(\eta),$$

for all $\eta \in \mathcal{E}$.

The carré-du-champ operator is associated to the evolution in time of the variance of the test function. Indeed,

$$\mathrm{Var}_\eta(f(\eta_t)) = \int_0^t e^{s\mathcal{L}}\left(\Gamma_{\mathcal{L}}\left(e^{(t-s)\mathcal{L}}f\right)\right)(\eta)\mathrm{d}s,$$

where $(e^{t\mathcal{L}})_{t \geq 0}$ denotes the semigroup generated by $\mathcal{L}$. See, for example, (Cloez and Thai, 2016b, p. 695).

Proof of Theorem 1.7: Our method of proof is based on Wilson's method (cf. (Levin and Peres, 2017, Thm. 13.28)). Let us denote $V = [v_1, v_2, \ldots, v_K]$ a real right-eigenvector satisfying $QV = -\lambda V$. Then, using Theorem 1.4 and Lemma 2.4 (specifically equations (2.8) and (2.9)) we get that $\tilde{\xi}(V)$ and $\tilde{\xi}(V, V)$ are right-eigenfunctions of $\mathcal{Q}_{N,p}$ satisfying

$$\left(e^{t\mathcal{Q}_{N,p}}\tilde{\xi}(V)\right)(\eta) = e^{-t\lambda}\tilde{\xi}(V)(\eta)$$

$$\left(e^{t\mathcal{Q}_{N,p}}\tilde{\xi}(V,V)\right)(\eta) = e^{-2(\lambda+p/N)t}\tilde{\xi}(V,V)(\eta),$$

for every $\eta \in \mathcal{E}_{K,N}$. We recall that from (2.9) we have

$$\tilde{\xi}(V,V) = \tilde{\xi}(V)^2 - \tilde{\xi}(V \odot V).$$

where $V \odot V = [v_1^2, \ldots, v_K^2]$ is the componentwise square vector of V.

Thereafter, using $\tilde{\xi}(V)$ as a test function we get

$$\mathrm{d}^{\mathrm{TV}}(\delta_{N\mathbf{e}_k}e^{t_{N,c}\mathcal{Q}_{N,p}}, \nu_{N,p}) \geq \mathbb{P}_{N\mathbf{e}_k}\left[\tilde{\xi}(V)(\eta_t) \geq \mu_t/2\right] - \mathbb{P}_{\nu_{N,p}}\left[\tilde{\xi}(V)(\eta_\infty) \geq \mu_t/2\right], \qquad (4.2)$$

where $\mu_t = \mathbb{E}_{N\mathbf{e}_k}\left[\tilde{\xi}(V)(\eta_t)\right] = e^{-t\lambda}Nv_k$. Using Markov and Chebyshev inequalities we obtain

$$\mathbb{P}_{\nu_{N,p}}\left[\tilde{\xi}(V)(\eta_\infty) \geq \frac{\mu_t}{2}\right] \leq 4e^{2t\lambda}\frac{\mathrm{Var}_{\nu_{N,p}}\left[\tilde{\xi}(V)(\eta_\infty)\right]}{|v_k|^2N^2},$$

$$\mathbb{P}_{N\mathbf{e}_k}\left[\tilde{\xi}(V)(\eta_t) \geq \frac{\mu_t}{2}\right] \geq 1 - 4e^{2t\lambda}\frac{\mathrm{Var}_{N\mathbf{e}_k}\left[\tilde{\xi}(V)(\eta_t)\right]}{|v_k|^2N^2}.$$

Thus, plugging these last expressions into (4.2) we get

$$\mathrm{d}^{\mathrm{TV}}(\delta_{N\mathbf{e}_k}e^{t\mathcal{Q}_{N,p}}, \nu_{N,p}) \geq 1 - 8\frac{e^{2\lambda t}}{|v_k|^2N}\sup_{t\geq 0}\frac{\mathrm{Var}_{N\mathbf{e}_k}\left[\tilde{\xi}(V)(\eta_t)\right]}{N}.$$

We are interested in finding a lower bound for $\mathrm{D}_{\mathcal{Q}_N}^{\mathrm{TV}}$ at time $t_{N,c} = (\ln(N) - c)/2\lambda$. It remains to prove a bound for the last factor in the previous expression.

Note that

$$\Gamma_{\mathcal{Q}_{N,p}}\tilde{\xi}(V) = \mathcal{Q}_{N,p}(\tilde{\xi}(V)^2) - 2\tilde{\xi}(V)\mathcal{Q}_{N,p}(\tilde{\xi}(V))$$

$$= \mathcal{Q}_{N,p}(\tilde{\xi}(V,V)) + \mathcal{Q}_{N,p}\tilde{\xi}(V \odot V) - 2\tilde{\xi}(V)\mathcal{Q}_{N,p}(\tilde{\xi}(V))$$

$$= -2\left(\lambda + \frac{p}{N}\right)\tilde{\xi}(V,V) + 2\lambda\tilde{\xi}(V)^2 + \tilde{\xi}(Q(V \odot V))$$

$$= -2\frac{p}{N}\tilde{\xi}(V,V) + 2\lambda\tilde{\xi}(V \odot V) + \tilde{\xi}(Q(V \odot V)).$$

Thus,

$$\frac{\mathrm{Var}_{N\mathbf{e}_k}\left(\tilde{\xi}(V)(\eta_t)\right)}{N} = \frac{1}{N}\int_0^t e^{s\mathcal{Q}_{N,p}}\left(\Gamma_{\mathcal{Q}_{N,p}}\left(e^{(t-s)\mathcal{Q}_{N,p}}\tilde{\xi}(V)\right)\right)(N\mathbf{e}_k)\mathrm{d}s$$

$$= \frac{1}{N}\int_0^t e^{-\lambda(t-s)}e^{s\mathcal{Q}_{N,p}}\left(\Gamma_{\mathcal{Q}_{N,p}}\tilde{\xi}(V)\right)(N\mathbf{e}_k)\mathrm{d}s$$

$$= \frac{1}{N}\int_0^t e^{-\lambda(t-s)}e^{s\mathcal{Q}_{N,p}}\left(-2\frac{p}{N}\tilde{\xi}(V,V) + 2\lambda\tilde{\xi}(V \odot V) + \tilde{\xi}(Q(V \odot V))\right)(N\mathbf{e}_k)\mathrm{d}s.$$

Note that

$$\frac{1}{N}\int_0^t e^{-\lambda(t-s)}e^{s\mathcal{Q}_{N,p}}\left(-2\frac{p}{N}\tilde{\xi}(V,V)\right)(N\mathbf{e}_k)\mathrm{d}s = -2\left(1 - \frac{1}{N}\right)p\,v_k^2 e^{-\lambda t}\int_0^t e^{-(\lambda+2p/N)s}\mathrm{d}s \leq 0.$$

Hence,

$$\frac{\operatorname{Var}_{N\mathbf{e}_k}\left(\tilde{\xi}(V)(\eta_t)\right)}{N} \leq \frac{1}{N}\int_0^t e^{-\lambda(t-s)}e^{s\mathcal{Q}_{N,p}}\left(2\lambda\tilde{\xi}(V\odot V) + \tilde{\xi}(Q(V\odot V))\right)(N\mathbf{e}_k)\mathrm{d}s$$

$$\leq \left(2\lambda\left\|\frac{\tilde{\xi}(V\odot V)}{N}\right\|_\infty + \left\|\frac{\tilde{\xi}(Q(V\odot V))}{N}\right\|_\infty\right)\int_0^t e^{-\lambda(t-s)}\mathrm{d}s$$

$$\leq (2 + \|Q\|_\infty/\lambda)\|V\|_\infty.$$

The lower bound for $\mathrm{D}_{\mathcal{Q}_{N,p}}^{\mathrm{TV}}$ is obtained considering the initial distribution concentrated at $N\mathbf{e}_{k^\star}$, where $k^\star$ satisfies $|v_{k^\star}| = \|V\|_\infty$. $\qquad\square$

5. Neutral multi-allelic Moran type process with parent independent mutation

In this section we discuss some applications of the Theorem 1.4 and its consequences to the neutral multi-allelic Moran model with parent independent mutation scheme. We will use some well-known results on finite state reversible Markov chains and their convergence to stationarity. We refer the interested reader to (Brémaud, 2020), (Levin and Peres, 2017) and (Saloff-Coste, 1997), for further details. We will focus on the case where the Moran process has *parent independent mutation* (Etheridge, 2011), and thus it is reversible. In fact, as we claimed in Lemma 1.9, the neutral Moran process with $p > 0$ is reversible if and only if its mutation matrix satisfies the *parent independent* condition. We explicitly diagonalise the infinitesimal generator of the neutral multi-allelic Moran process with parent independent mutation rate using the multivariate Hahn and Krawtchouk polynomials, which allows us to provide an explicit expression for the transition function of this process. Using these results we prove Theorems 1.12 and 1.13.

5.1. *Proof of Theorems 1.12 and 1.13.* Let us recall that the generator of the neutral multi-allelic Moran process with parent independent mutation defined by (5.1), which acts on a real function f on $\mathcal{E}_{K,N}$ as follows

$$(\mathcal{L}_{N,p}f)(\eta) := \sum_{i,j=1}^K \eta(i)\left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)\right]\left(\mu_j + p\frac{\eta(j)}{N}\right),$$

for all $\eta \in \mathcal{E}_{K,N}$.

Multivariate orthogonal Hahn and Krawtchouk polynomials. The rest of the section is devoted to the characterisation of the eigenfunctions of $\mathcal{L}_{N,p}$ and the proof of Theorem 1.12. Let us establish some notation that will be useful in the sequel to study the eigenfunctions of $\mathcal{L}_{N,p}$. For a K-dimensional real vector $\mathbf{x}$ we define the following quantities:

$$|\mathbf{x}_i| := \sum_{j=1}^i x_j, \quad |\mathbf{x}^i| := \sum_{j=i}^K x_j.$$

We set by convention $|\mathbf{x}^i| := 0$, for all $i > K$.

The orthogonal polynomials we define below are indexed by the set $\bigcup_{L=0}^N \mathcal{E}_{K-1,L}$, where $\mathcal{E}_{K-1,0} = \{\mathbf{0}\}$ is the set formed by the $K-1$ dimensional null vector. We define the *multivariate Hahn polynomials* on $\mathcal{E}_{K,N}$, indexed by $\eta \in \mathcal{E}_{K-1,L}$, for $L \in [N]_0$, and denoted $H_\eta(\mathbf{x}; N, \boldsymbol{\alpha})$, as follows

$$H_\eta(\mathbf{x}; N, \boldsymbol{\alpha}) := \frac{1}{(N)_{[|\eta|]}}\prod_{k=1}^{K-1}(-N + |\mathbf{x}_{k-1}| + |\eta^{k+1}|)_{(\eta(k))}H_{\eta(k)}(x_k; M_k, \alpha_k, \gamma_k) \qquad (5.1)$$

where $M_k = N - |\mathbf{x}_{k-1}| - |\eta^{k+1}|$, $\gamma_k = |\boldsymbol{\alpha}^{k+1}| + 2|\eta^{k+1}|$ and $H_n(x; M, \beta, \gamma)$ is the *univariate Hahn polynomial* defined by

$$H_n(x; M, \beta, \gamma) := {}_3F_2 \left(\begin{array}{cc} -n, & n+\beta+\gamma-1, -x \\ \beta, & -M \end{array} \middle| 1 \right) \tag{5.2}$$

$$= \sum_{j=0}^{n} \frac{(-n)_{(j)}(n+\beta+\gamma-1)_{(j)}(-x)_{(j)}}{\beta_{(j)}(-M)_{(j)}} \frac{1}{j!}.$$

Note that for $\mathbf{0} \in \mathcal{E}_{K-1,0}$ we obtain $H_\mathbf{0}(\cdot; N, \boldsymbol{\alpha}) \equiv 1$. In addition, it is not difficult to check that $H_\eta(N\mathbf{e}_K; N, \boldsymbol{\alpha}) \equiv 1$, for all $\eta \in \bigcup_{L=0}^{N} \mathcal{E}_{K-1,L}$.

We also define the *multivariate Krawtchouk polynomials* on $\mathcal{E}_{K,N}$ denoted $K_\eta(\mathbf{x}; N, \mathbf{q})$, indexed by $\eta \in \bigcup_{L=0}^{N} \mathcal{E}_{K-1,L}$, with $\mathbf{q} \in (0,1)^K$ such that $|\mathbf{q}| = 1$, as the multivariate polynomials satisfying:

$$K_\eta(\mathbf{x}; N, \mathbf{q}) := \frac{1}{(N)_{[|\eta|]}} \prod_{k=1}^{K-1} (-N + |\mathbf{x}_{k-1}| + |\eta^{k+1}|)_{(\eta(k))} K_{\eta(k)} \left(x_k; M_k, \frac{q_k}{|\mathbf{q}^k|} \right) \tag{5.3}$$

where $M_k = N - |\mathbf{x}_{k-1}| - |\eta^{k+1}|$, and $K_n(x; N, q)$ is the *univariate Krawtchouk polynomial* defined by

$$K_n(x; N, q) := {}_2F_1 \left(\begin{array}{c} -n, & -x \\ -N \end{array} \middle| \frac{1}{q} \right) \tag{5.4}$$

$$= \sum_{j=0}^{n} \frac{(-n)_{(j)}(-x)_{(j)}}{(-N)_{(j)}} \frac{1}{j! q^j}.$$

In addition, $K_\mathbf{0}(\cdot; N, \mathbf{q}) \equiv 1$, for $\mathbf{0} \in \mathcal{E}_{K-1,0}$, and $K_\eta(N\mathbf{e}_K; N, \mathbf{q}) \equiv 1$, for all $\eta \in \bigcup_{L=0}^{N} \mathcal{E}_{K-1,L}$.

See (Ismail, 2005, Ch. 6) and (Koekoek et al., 2010, Ch. 9) for more details about the univariate Hahn and Krawtchouk polynomials. We define the univariate Hahn and Krawtchouk polynomials in (5.2) and (5.4), respectively, using the hypergeometric functions notation which could be very useful for algebraic manipulations (cf. Koekoek et al., 2010, Ch. 10). For instance, consider $\boldsymbol{\alpha} = N\boldsymbol{\mu}/p$ in the definition of Hahn polynomials, then

$$\lim_{p\to 0^+} H_{\eta(k)}(x_k; M_k, \alpha_k, |\boldsymbol{\alpha}^{k+1}| + 2|\eta^{k+1}|) = \lim_{p\to 0^+} H_{\eta(k)} \left(x_k; M_k, \frac{N\mu_k}{p}, \frac{N|\boldsymbol{\mu}^{k+1}|}{p} + 2|\eta^{k+1}| \right)$$

$$= \lim_{p\to 0^+} {}_3F_2 \left(\begin{array}{cc} -\eta(k), & \eta(k) + N\mu_k/p + N|\boldsymbol{\mu}^{k+1}|/p + 2|\eta^{k+1}| - 1, -x_k \\ N\mu_k/p, & -M_k \end{array} \middle| 1 \right)$$

$$= {}_2F_1 \left(\begin{array}{c} -\eta(k), -x_k \\ -M_k \end{array} \middle| \frac{\mu_k + |\boldsymbol{\mu}^{k+1}|}{\mu_k} \right)$$

$$= K_{\eta(k)} \left(x_k; N, \frac{\mu_k}{|\boldsymbol{\mu}^k|} \right),$$

for every $k \in [K]$, where the calculation of the limit in the third equation follows from (Koekoek et al., 2010, Eq. (1.4.5)) and the last inequality follows from the definition of univariate Krawtchouk polynomials in (5.4). Now, using the previous limit and the definitions (5.1) and (5.3) of the multivariate Hahn and Krawtchouk polynomials we get

$$\lim_{p\to 0^+} H_\eta \left(\mathbf{x}; N, N\frac{\boldsymbol{\mu}}{p} \right) = K_\eta \left(\mathbf{x}; N, \frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|} \right).$$

Thus, similarly to how we define $\nu_{N,p}$ in (1.13), we define the multivariate polynomial $Q_\eta(\,\cdot\,; N, \boldsymbol{\mu}, p)$ by

$$Q_\eta(\mathbf{x}; N, \boldsymbol{\mu}, p) := \begin{cases} H_\eta\left(\mathbf{x}; N, \frac{N\boldsymbol{\mu}}{p}\right) & \text{if} \quad p > 0 \\ K_\eta\left(\mathbf{x}; N, \frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|}\right) & \text{if} \quad p = 0, \end{cases} \tag{5.5}$$

for every $\eta \in \bigcup_{L=0}^N \mathcal{E}_{K-1,L}$, and for all $\mathbf{x} \in \mathcal{E}_{K,N}$. Note that the functions $Q_\eta(\mathbf{x}; N, \boldsymbol{\mu}, p)$ are continuous when p tends towards zero, in the sense that:

$$\lim_{p \to 0^+} Q_\eta(\mathbf{x}; N, \boldsymbol{\mu}, p) = Q_\eta(\mathbf{x}; N, \boldsymbol{\mu}, 0),$$

for every $\mathbf{x} \in \mathcal{E}_{K,N}$. The following result sets some important properties of the multivariate Hahn and Krawtchouk polynomials.

Proposition 5.1 (Orthogonality of the Hahn and Krawtchouk polynomials). *The multivariate polynomials Q_η defined by (5.5) satisfy the following properties:*

a) *$Q_\eta(\,\cdot\,; N, \boldsymbol{\mu}, p)$ is a polynomial on $\mathcal{E}_{K,N}$ of total degree $|\eta|$, for every $\eta \in \bigcup_{L=0}^N \mathcal{E}_{K-1,L}$.*

b) *The polynomials $Q_\eta(\,\cdot\,; N, \boldsymbol{\mu}, p)$ are orthogonal on $\mathcal{E}_{K,N}$ with respect to the probability distribution $\nu_{N,p}$, defined by (1.13), i.e.*

$$\mathbb{E}_{\nu_{N,p}}\left[Q_\eta(\,\cdot\,; N, \boldsymbol{\mu}, p) Q_{\eta'}(\,\cdot\,; N, \boldsymbol{\mu}, p)\right] = \sum_{\xi \in \mathcal{E}_{K,N}} Q_\eta(\xi; N, \boldsymbol{\mu}, p) Q_{\eta'}(\xi; N, \boldsymbol{\mu}, p) \nu_{N,p}(\xi)$$

$$= d_{\eta,p}^2 \, \delta_{\eta,\eta'},$$

for every $\eta, \eta' \in \bigcup_{L=0}^N \mathcal{E}_{K-1,L}$, where $\delta_{\eta,\eta'}$ stands for the Kronecker delta function and

$$d_{\eta,p}^2 = \begin{cases} \dfrac{(|\boldsymbol{\alpha}| + N)_{(|\eta|)}}{(N)_{[|\eta|]} |\boldsymbol{\alpha}|_{(2|\eta|)}} \displaystyle\prod_{j=1}^{K-1} \dfrac{(|\boldsymbol{\alpha}^j| + |\eta^j| + |\eta^{j+1}| - 1)_{(\eta(j))} (|\boldsymbol{\alpha}^{j+1}| + 2|\eta^{j+1}|)_{(\eta(j))} \eta(j)!}{(\alpha_j)_{(\eta(j))}}, & p > 0 \\ \dfrac{1}{(N)_{[|\eta|]}} \displaystyle\prod_{j=1}^{K-1} \dfrac{(|\boldsymbol{\pi}^j|)^{\eta(j)} (|\boldsymbol{\pi}^{j+1}|)^{\eta(j)}}{\pi_j^{\eta(j)}} \eta(j)!, & p = 0, \end{cases}$$

where $\boldsymbol{\alpha} = N\boldsymbol{\mu}/p$ and $\boldsymbol{\pi} = \boldsymbol{\mu}/|\boldsymbol{\mu}|$.

See Theorem 5.4 in (Iliev and Xu, 2007) and Proposition 2.1, also Remark 2.2, in (Khare and Zhou, 2009) for the proofs of these results on multivariate Hahn polynomials. See Theorem 6.2 in (Iliev and Xu, 2007) and Proposition 2.4 in (Khare and Zhou, 2009) for the proofs for the multivariate Krawtchouk polynomials. The system of orthogonal polynomials for a fixed multinomial distribution is not unique. A general construction of the multivariate Krawtchouk polynomials can be found in (Diaconis and Griffiths, 2014).

Kernel polynomials for Dirichlet multinomial and multinomial distributions. Consider ν a multivariate distribution on $\mathcal{E}_{K,N}$ and $\{Q_\eta^0\}$ an orthonormal system of polynomials in $l^2(\mathbb{R}^{\mathcal{E}_{K,N}}, \nu)$. Then, the *kernel polynomial* associated to ν is defined by

$$h_n(\mathbf{x}, \mathbf{y}) := \sum_{|\eta|=n} Q_\eta^0(\mathbf{x}) Q_\eta^0(\mathbf{y}),$$

for all $\mathbf{x}, \mathbf{y} \in \mathcal{E}_{K,N}$ and for every $n \in [N]_0$. The kernel polynomials are invariant under the choice of the orthonormal systems, i.e. they only depend on the distribution ν. Kernel polynomials are used for manipulating sums of products of orthogonal polynomials. They are especially useful to

obtain explicit expressions for the transition function of a reversible Markov chain with polynomial eigenfunctions, as we show in Remark 5.2 below.

We next review the expressions for the kernel polynomials of the Dirichlet multinomial and the multinomial distributions. Let us denote by $h_n(\mathbf{x}, \mathbf{y}; p)$ the n-th kernel polynomial of $\nu_{N,p}$, for all $n \in [N]_0$. Then, it can be proven that

$$h_n(N\mathbf{e}_k, N\mathbf{e}_k; p) = \binom{N}{n} \frac{(|\boldsymbol{\alpha}| + 2n - 1)(|\boldsymbol{\alpha}|)_{(n-1)}(|\boldsymbol{\alpha}| - \alpha_k)_{(n)}}{(|\boldsymbol{\alpha}| + N)_{(n)}(\alpha_k)_{(n)}}, \tag{5.6}$$

for all $p > 0$, see (Khare and Zhou, 2009, Eq. (2.18)).

For $p = 0$, $\nu_{N,0}$ follows a $\mathcal{M}(\cdot \mid N, \boldsymbol{\mu}/|\boldsymbol{\mu}|)$ distribution and its n-th kernel polynomial satisfies

$$h_n(\mathbf{x}, N\mathbf{e}_k; 0) = \sum_{m=0}^{n} \binom{N}{m}\binom{N-m}{n-m}(-1)^{n-m} \frac{(x_k)_{[m]}}{N_{[m]}} \left(\frac{\mu_k}{|\boldsymbol{\mu}|}\right)^{-m}, \tag{5.7}$$

and

$$h_n(N\mathbf{e}_k, N\mathbf{e}_k; 0) = \binom{N}{n}\left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1\right)^{n}. \tag{5.8}$$

For more details on the kernel polynomials for the multinomial distribution see e.g. (Khare and Zhou, 2009, Prop. 2.8) and (Diaconis and Griffiths, 2019). Also, for more details on the kernel polynomials for the Dirichlet multinomial distribution see e.g. (Khare and Zhou, 2009, Prop. 2.6) and (Griffiths and Spanò, 2013).

We next show that the right eigenfunctions of $\mathcal{L}_{N,p}$ are given by multivariate orthogonal polynomials defined by (5.5).

Remark 5.2 (Eigenfunctions of $\mathcal{L}_{N,p}$). The right eigenfunctions of $\mathcal{L}_{N,p}$ are the multivariate polynomials $Q_\eta(\cdot; N, \boldsymbol{\mu}, p)$ with associated eigenvalue $\lambda_{L,p}$, for $\eta \in \mathcal{E}_{K-1,L}$, for $L \in [N]_0$. Moreover, the set of right eigenfunctions

$$\left\{ Q_\eta(\cdot; N, \boldsymbol{\mu}, p), \eta \in \bigcup_{L=0}^{N} \mathcal{E}_{K-1,L} \right\}$$

is orthogonal in $l^2(\nu_{N,p})$, for all $p \geq 0$. In addition, the functions $\phi_\eta(\cdot; N, \boldsymbol{\mu}, p)$ defined by

$$\phi_\eta(\eta'; N, \boldsymbol{\mu}, p) := \nu_{N,p}(\eta')Q_\eta(\eta'; N, \boldsymbol{\mu}, p)$$

are left eigenfunctions of $\mathcal{L}_{N,p}$ and the set of left eigenfunctions is orthogonal in $l^2(1/\nu_{N,p})$.

Furthermore, the transition kernel of the Markov chain driven by $\mathcal{L}_{N,p}$ can be decomposed as follows:

$$(e^{t\mathcal{L}_{N,p}}\delta_\xi)(\eta) = \nu_{N,p}(\xi)\left(1 + \sum_{L=1}^{N} e^{\lambda_{L,p}t}h_L(\eta, \xi; p)\right), \tag{5.9}$$

where $h_L(\eta, \xi; p)$ is the kernel polynomial associated to $\nu_{N,p}$.

Griffiths and Spanò (2013) give the expression (5.9) for the transition kernel of the process driven by $\mathcal{L}_{N,p}$, for $p > 0$, as an example of the usefulness of the kernel polynomials for the Dirichlet multinomial distribution.

The following result provides an explicit expression for the chi-square distance between the distribution of the Markov process driven by $\mathcal{L}_{N,p}$ starting at $N\mathbf{e}_k$ and its stationary distribution at a given time t.

Corollary 5.3 (Explicit expression for the chi-square distance)**.** *For $K \geq 2$, $N \geq 2$ and $p \geq 0$, we obtain the following explicit expression for the chi-square distance between the distribution of the*

reversible process driven by $\mathcal{L}_{N,p}$ at time t when the initial distribution is concentrated at $N\mathbf{e}_k$, for $k \in [K]$:

$$\chi^2_{N\mathbf{e}_k}(t) = \begin{cases} \left[1 + e^{-2|\boldsymbol{\mu}|t}\left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1\right)\right]^N - 1 & \text{if } p = 0 \\ \displaystyle\sum_{L=1}^{N} e^{2\lambda_{L,p}t}\binom{N}{L}\frac{(|\boldsymbol{\alpha}| + 2L - 1)(|\boldsymbol{\alpha}|)_{(L-1)}(|\boldsymbol{\alpha}| - \alpha_k)_{(L)}}{(|\boldsymbol{\alpha}| + N)_{(L)}(\alpha_k)_{(L)}} & \text{if } p > 0 \end{cases} \tag{5.10}$$

Proof: Using classical results on reversible Markov chains, see e.g. (Khare and Zhou, 2009, Eq. (2.1)), we obtain the following equality for the chi-square distance:

$$\chi^2_{N\mathbf{e}_k}(t) = \sum_{L=1}^{N} e^{2\lambda_{L,p}t}h_L(N\mathbf{e}_k, N\mathbf{e}_k; p),$$

where $h_L(N\mathbf{e}_k, N\mathbf{e}_k; p)$ stands for the kernel polynomials associated to $\nu_{N,p}$, as defined in (5.6) and (5.8). Thus, the expression for $\chi^2_{N\mathbf{e}_k}(t)$ in (5.10) simply comes from (5.6), when $p > 0$.

To prove the case when $p = 0$, note that (5.8) implies

$$\chi^2_{N\mathbf{e}_k}(t) = \sum_{L=1}^{N} e^{-2L|\boldsymbol{\mu}|t}\binom{N}{L}\left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1\right)^L = \left[1 + e^{-2|\boldsymbol{\mu}|t}\left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1\right)\right]^N - 1.$$

$\square$

We now take advantage of the explicit expression in (5.10) to prove the existence of a strongly optimal cutoff in the chi-square distance for the multi-allelic Moran process with parent independent mutation when $N \to \infty$.

Proof of Theorem 1.12: Let us first prove the existence of the chi-square cutoff. When $p = 0$, for $t_{N,c} = \dfrac{\ln N + c}{2|\boldsymbol{\mu}|}$ we obtain

$$\lim_{N\to\infty} \chi^2_{N\mathbf{e}_k}(t_{N,c}) = \lim_{N\to\infty}\left[1 + \frac{e^{-c}}{N}\left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1\right)\right]^N - 1$$

$$= \exp\left\{-\left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1\right)e^{-c}\right\} - 1.$$

Now, since $K_{k,0} = |\boldsymbol{\mu}|/\mu_k - 1$, we have proved the existence of the limit (1.14) for $p = 0$.

Now, for $p > 0$ let us focus on expression (5.10). For every $L \in \mathbb{N}$ and $k \in [K]$, let us denote

$$\phi_{L,k}(N) := \frac{(|\boldsymbol{\alpha}| + 2L - 1)(|\boldsymbol{\alpha}|)_{(L-1)}(|\boldsymbol{\alpha}| - \alpha_k)_{(L)}}{(|\boldsymbol{\alpha}| + N)_{(L)}(\alpha_k)_{(L)}}.$$

We thus have

$$\phi_{L,k}(N) := \frac{|\boldsymbol{\alpha}| + 2L - 1}{|\boldsymbol{\alpha}| + L - 1}\frac{\prod_{r=0}^{L-1}(|\boldsymbol{\alpha}| + r)(|\boldsymbol{\alpha}| - \alpha_k + r)}{\prod_{r=0}^{L-1}(|\boldsymbol{\alpha}| + N + r)(\alpha_k + r)}$$

$$= \frac{N|\boldsymbol{\mu}|/p + 2L - 1}{N|\boldsymbol{\mu}|/p + L - 1}\left[\frac{|\boldsymbol{\mu}|(|\boldsymbol{\mu}| - \mu_k)}{\mu_k(|\boldsymbol{\mu}| + p)}\right]^L\frac{\prod_{r=0}^{L-1}\left(1 + \frac{p}{N|\boldsymbol{\mu}|}r\right)\left(1 + \frac{p}{N(|\boldsymbol{\mu}| - \mu_k)}r\right)}{\prod_{r=0}^{L-1}\left(1 + \frac{p}{N(|\boldsymbol{\mu}| + p)}r\right)\left(1 + \frac{p}{N\mu_k}r\right)}.$$

Hence, for all $L \in \mathbb{N}$ we get

$$\lim_{N \to \infty} \phi_{L,k}(N) = \left[\frac{|\boldsymbol{\mu}|(|\boldsymbol{\mu}| - \mu_k)}{\mu_k(|\boldsymbol{\mu}| + p)} \right]^L = (K_{k,p})^L.$$

Moreover,

$$\binom{N}{L} \underset{N}{\sim} \frac{N^L}{L!} \quad \text{and} \quad e^{2\lambda_L t_N} \underset{N}{\sim} \frac{(e^{-c})^L}{N^L},$$

where for two sequences (f_N) and (g_N) the notation $f_N \underset{N}{\sim} g_N$ means $f_N - g_N = o(g_N)$. According to (5.6) we have

$$h_L(N\mathbf{e}_k, N\mathbf{e}_k; p) = \binom{N}{L} \frac{(|\boldsymbol{\alpha}| + 2L - 1)(|\boldsymbol{\alpha}|)_{(L-1)}(|\boldsymbol{\alpha}| - \alpha_k)_{(L)}}{(|\boldsymbol{\alpha}| + N)_{(L)}(\alpha_k)_{(L)}}.$$

Plugging these asymptotic expressions in the L-th summand of (5.10) yields

$$\lim_{N \to \infty} e^{2\lambda_L t_N} h_L(N\mathbf{e}_k, N\mathbf{e}_k; p) = \frac{(K_{k,p} e^c)^L}{L!}.$$

Moreover,

$$e^{2\lambda_L t_N} h_L(N\mathbf{e}_k, N\mathbf{e}_k; p) \le e^{-L(c + \ln(N))} h_L(N\mathbf{e}_k, N\mathbf{e}_k; p)$$

$$= \frac{e^{-cL}}{N^L} \binom{N}{L} \frac{|\boldsymbol{\alpha}| + 2L - 1}{\boldsymbol{\alpha}| + L - 1} \frac{|\boldsymbol{\alpha}|_{(L)}(|\boldsymbol{\alpha}| - \alpha_k)_{(L)}}{(|\alpha| + N)_{(L)}(\alpha_k)_{(L)}}$$

$$= \frac{e^{-cL}}{L!} \frac{|\boldsymbol{\alpha}| + 2L - 1}{\boldsymbol{\alpha}| + L - 1} \prod_{r=0}^{L-1} \left[\frac{N - r}{N} \frac{|\boldsymbol{\alpha}| + r}{|\boldsymbol{\alpha}| + N + r} \frac{|\boldsymbol{\alpha}| - \alpha_k + r}{\alpha_k + r} \right]$$

$$\le 3 \frac{(\gamma e^{-c})^L}{L!},$$

where $\gamma = \max\{1, K_{k,0}\}$.

For an arbitrary small $\epsilon > 0$ let us consider $M \in \mathbb{N}$ such that $3 \sum_{L=M+1}^{\infty} \frac{(\gamma e^{-c})^L}{L!} \le \frac{\epsilon}{3}$, and let N_ϵ be a positive integer such that

$$\left| \sum_{L=1}^{M} e^{2\lambda_L t_N} h_L(N\mathbf{e}_k, N\mathbf{e}_k; p) - \sum_{L=1}^{M} \frac{(K_{k,p})^L}{L!} \right| \le \frac{\epsilon}{3},$$

for all $N \ge N_\epsilon$. Note that

$$\sum_{L=M+1}^{\infty} \frac{(K_{k,p} e^{-c})^L}{L!} \le \frac{\epsilon}{3}.$$

Then, for all $N \ge N_\epsilon$, using the triangular inequality we have

$$\left| \sum_{L=1}^{N} e^{2\lambda_L t_N} h_L(N\mathbf{e}_k) - \left(\exp\{K_{k,p} e^{-c} - 1\} \right) \right| \le \epsilon,$$

which concludes the proof for the chi-square cutoff for the process driven by $\mathcal{L}_{N,p}$, for $p \ge 0$. $\square$

Let us establish a result that will be very useful during the proof of Theorem 1.13.

Lemma 5.4 (Lemma A.2 in Nestoridi and Olesker-Taylor, 2020). *Let $\psi_N \in (0,1)$, for all $N \in \mathbb{N}$, such that $N\psi_N \to \infty$, when $N \to \infty$. Then, for all $y \in \mathbb{R}$ we have*

$$\lim_{N \to \infty} d^{\mathrm{TV}} \left(\mathrm{Bin}(N, \psi_N), \mathrm{Bin}\left(N, \psi_N + \sqrt{\frac{\psi_N(1 - \psi_N)}{N}} y \right) \right) = 2\Phi\left(\frac{1}{2}|y| \right) - 1,$$

where where $\mathrm{Bin}(N, \psi)$ *stands for the binomial distribution with N trials and probability of success ψ, and Φ is the cumulative distribution function of the standard normal distribution, i.e.*

$$\Phi : t \mapsto \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-s^2/2} \mathrm{d}s.$$

This lemma characterises the limit profile of the total variation distance between two random variables B_1 and B_2, following binomial distributions, when the difference between their means is of the same order of the standard deviation of B_1/N. The proof can be found in the Appendix A.2 (Nestoridi and Olesker-Taylor, 2020), see also (Nestoridi and Olesker-Taylor, 2022), the published version of this paper.

Proof of Theorem 1.13: First note that the lower and upper bounds for $\mathrm{d}_{N\mathbf{e}_k}^{\mathrm{TV}}(t_{N,c})$ are simply consequences of Theorems 1.7 and 1.12, respectively. Indeed, for $c < 0$ and using Theorems 1.7 we have

$$\lim_{N\to\infty} \mathrm{d}_{N\mathbf{e}_k}^{\mathrm{TV}}(t_{N,c}) \geq 1 - \kappa \frac{\|V\|_\infty}{v_k} \mathrm{e}^{-c},$$

where $\kappa = 8(2+\|Q_{\boldsymbol{\mu}}\|_\infty/|\boldsymbol{\mu}|) = 32$, and V is any right eigenvector of $Q_{\boldsymbol{\mu}}$ with eigenvalue $|\boldsymbol{\mu}|$. Finally, the desired inequality is obtained considering the eigenvector $V = 1/\mu_k \mathbf{e}_k - 1/\mu_s \mathbf{e}_s$, where $s \in [K]$ satisfies $\mu_s \wedge \mu_k = \min_{r:r\neq k} \mu_r \wedge \mu_k$.

In addition, using the classical inequality between the chi-square and the total variation distances and Theorem 1.12 we get

$$\lim_{N\to\infty} \mathrm{d}_{N\mathbf{e}_k}^{\mathrm{TV}}(t_{N,c}) \leq \lim_{N\to\infty} \frac{1}{2}\sqrt{\chi_{N\mathbf{e}_k}^2(t_{N,c})} = \frac{1}{2}\sqrt{\exp\{K_{k,p}\mathrm{e}^{-c}\} - 1}.$$

This concludes to proof of the existence of the $\left(\dfrac{\ln N}{2|\boldsymbol{\mu}|}, 1\right)$ total variation cutoff.

Let us now prove the limit profile for the total variation distance when $p = 0$. Using (5.7) and (5.9) we get

$$
\begin{aligned}
(\mathrm{e}^{t\mathcal{L}_N}\delta_\xi)(N\mathbf{e}_k) &= \nu_N(\xi) \sum_{L=0}^{N} \mathrm{e}^{-|\boldsymbol{\mu}|Lt} \sum_{m=0}^{L} \binom{N}{m}\binom{N-m}{L-m}(-1)^{L-m}\frac{(\xi_k)_{[m]}}{N_{[m]}}\left(\frac{\mu_k}{|\boldsymbol{\mu}|}\right)^{-m} \\
&= \nu_N(\xi) \sum_{m=0}^{N} \binom{N}{m}\frac{(\xi_k)_{[m]}}{N_{[m]}}\left(\frac{\mu_k}{|\boldsymbol{\mu}|}\right)^{-m} \sum_{L=m}^{N} \mathrm{e}^{-|\boldsymbol{\mu}|Lt}\binom{N-m}{L-m}(-1)^{L-m} \\
&= \nu_N(\xi) \sum_{m=0}^{N} \binom{N}{m}\frac{(\xi_k)_{[m]}}{N_{[m]}}\left(\frac{\mu_k}{|\boldsymbol{\mu}|}\right)^{-m} \mathrm{e}^{-|\boldsymbol{\mu}|mt}(1-\mathrm{e}^{-|\boldsymbol{\mu}|t})^{N-m} \\
&= \nu_N(\xi)(1-\mathrm{e}^{-|\boldsymbol{\mu}|t})^N \sum_{m=0}^{\xi_k} \binom{\xi_k}{m}\left[\frac{\mu_k}{|\boldsymbol{\mu}|}\mathrm{e}^{|\boldsymbol{\mu}|t}(1-\mathrm{e}^{-|\boldsymbol{\mu}|t})\right]^{-m} \\
&= \nu_N(\xi)(1-\mathrm{e}^{-|\boldsymbol{\mu}|t})^{N-\xi_k}\left[(1-\mathrm{e}^{-|\boldsymbol{\mu}|t}) + \frac{|\boldsymbol{\mu}|\mathrm{e}^{-|\boldsymbol{\mu}|t}}{\mu_k}\right]^{\xi_k}.
\end{aligned}
$$

Thus, the process driven by $\mathcal{L}_N$ starting at $N\mathbf{e}_k$ at time t follows a $\mathcal{M}\left(\cdot \mid N, (1 - e^{-|\boldsymbol{\mu}|t})\frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|} + e^{-|\boldsymbol{\mu}|t}\mathbf{e}_k\right)$ distribution, which proves Corollary 1.14. Moreover,

$$d_{N\mathbf{e}_k}^{\mathrm{TV}}(t) = \frac{1}{2}\sum_{\xi \in \mathcal{E}_{K,N}} \left|(e^{t\mathcal{L}_N}\delta_\xi)(N\mathbf{e}_k) - \nu_N(\xi)\right|$$

$$= \frac{1}{2}\sum_{L=0}^{N}\sum_{\substack{\xi \in \mathcal{E}_{K,N}:\\ \xi_k = L}}\nu_N(\xi)\left|(1-e^{-|\boldsymbol{\mu}|t})^{N-\xi_k}\left[1 - e^{-|\boldsymbol{\mu}|t} + \frac{|\boldsymbol{\mu}|e^{-|\boldsymbol{\mu}|t}}{\mu_k}\right]^{\xi_k} - 1\right|$$

$$= \frac{1}{2}\sum_{L=0}^{N}\binom{N}{L}\left(\frac{\mu_k}{|\boldsymbol{\mu}|}\right)^L\left(1 - \frac{\mu_k}{|\boldsymbol{\mu}|}\right)^{N-L}\left|(1-e^{-|\boldsymbol{\mu}|t})^{N-L}\left[1 - e^{-|\boldsymbol{\mu}|t} + \frac{|\boldsymbol{\mu}|e^{-|\boldsymbol{\mu}|t}}{\mu_k}\right]^{L} - 1\right|$$

$$= d^{\mathrm{TV}}\left(\mathrm{Bin}\left(N, \frac{\mu_k}{|\boldsymbol{\mu}|}\right), \mathrm{Bin}\left(N, \frac{\mu_k}{|\boldsymbol{\mu}|}(1 - e^{-|\boldsymbol{\mu}|t}) + e^{-|\boldsymbol{\mu}|t}\right)\right).$$

Then, we have proved that we can write $d^{\mathrm{TV}}(t, N\mathbf{e}_k)$ as the total variation distance between two binomial distributions with parameters N both and probabilities of success $\pi_k = \mu_k/|\boldsymbol{\mu}|$ and $\tilde{\pi}_k = \pi_k(1 - e^{-|\boldsymbol{\mu}|t}) + e^{-|\boldsymbol{\mu}|t}$, respectively. For $t_{N,c} = \frac{\ln N + c}{2|\boldsymbol{\mu}|}$ we get

$$\tilde{\pi}_k = \pi_k + \frac{\sqrt{\pi_k(1-\pi_k)}}{\sqrt{N}}\sqrt{\frac{1-\pi_k}{\pi_k}}e^{-c/2}.$$

Therefore, using Lemma 5.4 we obtain

$$\lim_{N\to\infty} d_{N\mathbf{e}_k}^{\mathrm{TV}}(t) = 2\Phi\left(\frac{1}{2}\sqrt{K_{k,0}e^{-c}}\right) - 1,$$

where $K_{k,0} = \frac{1-\pi_k}{\pi_k} = \frac{|\boldsymbol{\mu}|}{\mu_k} - 1$.

$\square$

Appendix A. Proofs of Lemmas 2.1 and 2.4, and Proposition 2.5

This section is devoted to the proofs of Lemmas 2.1 and 2.4, and Proposition 2.5.

Proof of Lemma 2.1: (a) Let us first prove that for any $\alpha \in \mathcal{E}_{K,N}$, there exists a unique polynomial $P_\alpha \in H_{K,N}$, product of N linear functions on $H_{K,1}$, such that $P_\alpha(\eta) = 1$ if $\eta = \alpha$ and 0 otherwise. Indeed, let us define the polynomial P_α by

$$P_\alpha : \mathbf{x} \in \mathcal{E}_{K,N} \mapsto \prod_{k=1}^{K}\prod_{a=0}^{\alpha_k-1}\frac{x_k - a}{\alpha_k - a},$$

where $\prod_{a=0}^{\alpha_k-1}(x_k - a) = 1$ when $\alpha_k = 0$. Note that $P_\alpha = \mathbf{1}_\alpha$, for every $\alpha \in \mathcal{E}_{K,N}$. There are $\sum_{k=1}^{K}\alpha_k = N$ linear factors in the numerator. Also, each term $x_k - a$ may be replaced by $x_k - \frac{a}{N}\sum_{k=1}^{K}x_k$ when $a \neq 0$, so $P_\alpha(x)$ may be considered as a product of N linear functions on $H_{K,1}$, and because the uniqueness of such a function P_α is straightforward, (a) is proved.

Now, for every real function f on $\mathcal{E}_{K,N}$, the result is immediately obtained from (a) by setting

$$P := \sum_{\alpha \in \mathcal{E}_{K,N}} f(\alpha)P_\alpha.$$

(b) From part (b) we have that $\mathcal{B}_{H_{K,N}}$ is a generator system of $\mathbb{R}^{\mathcal{E}_{K,N}}$. Moreover,

$$\mathrm{Card}(\mathcal{B}_{H_{K,N}}) = \mathrm{Card}(\mathcal{E}_{K,N}) = \dim(\mathbb{R}^{\mathcal{E}_{K,N}}) = \binom{K-1+N}{N},$$

thus $\mathcal{B}_{H_{K,N}}$ is necessarily a basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$.   □

Proof of Lemma 2.4: (a) For $L = 1$: An injection $s : \{1\} \to \{1, 2, \ldots, N\}$ is characterised by $s(1) = i$. It follows from (2.5) that

$$\xi(V_1)(k_1, k_2, \ldots, k_N) = \sum_{i=1}^{N} V_1(k_i),$$

which is a symmetric function. For every $\eta = (\eta(1), \eta(2), \ldots, \eta(K)) \in \mathcal{E}_{K,N}$, we have

$$\tilde{\xi}(V_1)(\eta) = \big(\xi(V_1) \circ \psi_{K,N}\big)(\eta) = \sum_{j=1}^{K} V_1(j)\eta(j),$$

which finishes the proof of part (a).
   (b) From (2.5), we get

$$\xi(V_1, V_2, \ldots, V_L)(k_1, \ldots, k_N) = \sum_{s \in \mathcal{I}_{L-1,N}} V_1(k_{s(1)}) \ldots V_{L-1}(k_{s(L-1)}) \sum_{i \in [N] \setminus s([L-1])} V_L(k_i)$$

$$= \sum_{s \in \mathcal{I}_{L-1,N}} V_1(k_{s(1)}) \ldots V_{L-1}(k_{s(L-1)}) \left(\sum_{i=1}^{N} V_L(k_i) - \sum_{i=1}^{L-1} V_L(k_{s(i)}) \right)$$

$$= \xi(V_1, V_2, \ldots, V_{L-1})(k_1, \ldots, k_N)\xi(V_L)(k_1, \ldots, k_N)$$

$$- \sum_{i=1}^{L-1} \xi(V_1, \ldots, V_i \odot V_L, \ldots, V_{L-1})(k_1, \ldots, k_N).$$

Using (2.7) we obtain the result for $\tilde{\xi}(V_1, V_2, \ldots, V_L)$. The particular case $L = 2$ comes from part (a).
   (c) We can prove equation (2.10) by induction on L. For $L = 1$ the result easily comes by (a). If we suppose that (2.10) is satisfied for L, for $2 \le L < N - 1$, then, using (b) and (a), we can check that (2.10) holds for $L + 1$.   □

Proof of Proposition 2.5: Since $\mathcal{U}$ is a basis of $\mathbb{R}^K$ we trivially have that $\mathcal{U}^N$ is a basis of $\mathbb{R}^{[K]^N}$, proving (a) (cf. Levin and Peres, 2017, Lemma 12.12). To prove (b) we prove that each element of $\mathcal{U}^N$ has image in $\mathcal{S}^N$ by Sym, defined as in (1.5). First, $\mathrm{Sym}(U_0 \otimes \cdots \otimes U_0) = U_0 \otimes \cdots \otimes U_0$, since the constant function equal to one is symmetric. Furthermore, for every $W = W_1 \otimes W_2 \otimes \cdots \otimes W_N \in \mathcal{U}^N$ there is a permutation $\sigma \in \mathcal{S}_N$ such that $\sigma W = U_\eta$, with $\eta \in \mathcal{E}_{K-1,L}$, where $L \in [N]$ is the number of components in the expression of W different from U_0. Thus, $\mathrm{Sym}(W) = \mathrm{Sym}(\sigma W) = V_\eta$, for $\eta \in \mathcal{E}_{K-1,L}$. We have not proved that $V_\eta \ne V_\alpha$, for $\eta \ne \alpha$. However, $\mathcal{S}^N$ is a generator system of $\mathrm{Sym}(\mathbb{R}^{[K]^N})$ satisfying

$$\mathrm{Card}\left(\mathcal{S}^N\right) \le 1 + \sum_{L=1}^{N} \mathrm{Card}(\mathcal{E}_{K-1,L})$$

$$= \sum_{L=0}^{N} \binom{K-2+L}{L} = \binom{K-1+N}{N},$$

where the last equality is the well-known *Hockey–Stick identity* in combinatorics, see e.g. (Lovász et al., 2003). Now, since

$$\dim\left(\mathrm{Sym}\left(\mathbb{R}^{[K]^N}\right)\right) = \binom{K-1+N}{N},$$

we have that $\mathcal{S}^N$ is a generator system with a minimal number of vectors, therefore it is a basis of $\mathrm{Sym}(\mathbb{R}^{[K]^N})$. To prove (c) simply note that each element in $\tilde{\mathcal{S}}^N$ is the image by the isomorphism $\Phi_{K,N}$ of an element in $\mathcal{S}^N$. $\qquad\square$

Appendix B. Proof of Lemma 3.7

Proof of Lemma 3.7: Without lost of generality we can only prove the result for the monomials on $\mathcal{E}_{K,N}$. Consider m a monomial on $\mathcal{E}_{K,N}$ of total degree $|\alpha| = L$ with $0 \le L \le N$. Then, we want to prove that
$$\mathcal{A}_N V_m = -L(L-1)V_m + V_q,$$
where q is a polynomial with a total degree strictly less than L.

As we commented in Remark 3.6, the result is true for $L = 1$. Let us assume $L \ge 2$ and consider the monomial $m : \eta \mapsto \prod_{r=1}^K \eta(r)^{\alpha_r}$. Evaluating V_m in $\mathcal{A}_N$, defined by (1.4), we obtain

$$(\mathcal{A}_N V_m)(\eta) = \sum_{k,r:k\ne r} \left(\prod_{s\notin\{k,r\}} \eta(s)^{\alpha_s} \right) \left[(\eta(k)-1)^{\alpha_k}(\eta(r)+1)^{\alpha_r} - \eta(k)^{\alpha_k}\eta(r)^{\alpha_r} \right] \eta(k)\eta(r), \quad \text{(B.1)}$$

for all $\eta \in \mathcal{E}_{K,N}$. Then, from the Newton's binomial formula, we get

$$\eta(k)(\eta(k)-1)^{\alpha_k} = \eta(k)^{\alpha_k+1} - \alpha_k\eta(k)^{\alpha_k} + \frac{\alpha_k(\alpha_k-1)}{2}\eta(k)^{\alpha_k-1} + a(\eta(k)),$$

where $a(\eta(k))$ is a polynomial in $\eta(k)$ with degree strictly less than $\alpha_k - 1$ if $\alpha_k \ge 2$ and null otherwise. In the same way, we get

$$\eta(r)(\eta(r)+1)^{\alpha_r} = \eta(r)^{\alpha_r+1} + \alpha_r\eta(r)^{\alpha_r} + \frac{\alpha_r(\alpha_r-1)}{2}\eta(r)^{\alpha_r-1} + b(\eta(r)),$$

where $b(\eta(r))$ is a polynomial in $\eta(r)$ with degree strictly less than $\alpha_r - 1$ if $\alpha_r \ge 2$ and null otherwise.

Using this expansion in (B.1) and regrouping terms with total degree in $\eta(k)$ and $\eta(r)$ strictly less than $\alpha_k + \alpha_r$ give

$$
\begin{aligned}
(\mathcal{A}_N V_m)(\eta) = & \sum_{k,r:k\ne r} \left(\prod_{s\notin\{k,r\}} \eta(s)^{\alpha_s} \right) \left(\alpha_r\eta(k)^{\alpha_k+1}\eta(r)^{\alpha_r} - \alpha_k\eta(k)^{\alpha_k}\eta(r)^{\alpha_r+1} \right) \\
& + \sum_{k,r:k\ne r} \left(\prod_{s\notin\{k,r\}} \eta(s)^{\alpha_s} \right) \frac{\alpha_r(\alpha_r-1)}{2}\eta(k)^{\alpha_k+1}\eta(r)^{\alpha_r-1} \\
& - \sum_{k,r:k\ne r} \left(\prod_{s\notin\{k,r\}} \eta(s)^{\alpha_s} \right) \alpha_k\alpha_r\eta(k)^{\alpha_k}\eta(r)^{\alpha_r} \qquad\qquad \text{(B.2)} \\
& + \sum_{k,r:k\ne r} \left(\prod_{s\notin\{k,r\}} \eta(k)^{\alpha_s} \right) \frac{\alpha_k(\alpha_k-1)}{2}\eta(k)^{\alpha_k-1}\eta(r)^{\alpha_r+1} + w(\eta),
\end{aligned}
$$

where w is a polynomial in η of total degree strictly less than $\sum_k \alpha_k = L$. The first sum in the right member of (B.2) is null because the antisymmetry in k, r of its summands. The third term is

$$- \sum_{k,r:k\ne r} \left(\prod_{s\notin\{k,r\}} \eta(s)^{\alpha_s} \right) \alpha_k\alpha_r\eta(k)^{\alpha_k}\eta(r)^{\alpha_r} = -c_1 p(\eta),$$

with

$$c_1 = \sum_{k,r:k\neq r} \alpha_k\alpha_r = \left(\sum_{k=1}^{K}\alpha_k\right)^2 - \sum_{k=1}^{K}\alpha_k^2 = L^2 - \sum_{k=1}^{K}\alpha_k^2.$$

By symmetry in k and r, it is obvious that the second and the fourth sums in the right member of (B.2) are equal. Using

$$\alpha_r(\alpha_r - 1)\eta(r)^{\alpha_r-1} = \eta(r)\frac{\partial^2}{\partial\eta(r)^2}\eta(r)^{\alpha_r},$$

it follows that

$$\sum_{k\neq r}\left(\prod_{s\notin\{k,r\}}\eta(s)^{\alpha_s}\right)\alpha_r(\alpha_r-1)\eta(k)^{\alpha_k+1}\eta(r)^{\alpha_r-1} = \sum_{k,r:k\neq r}\eta(k)\eta(r)\frac{\partial^2}{\partial\eta(r)^2}m(\eta)$$

$$= \sum_{k,r}\eta(k)\eta(r)\frac{\partial^2}{\partial\eta(r)^2}m(\eta) - \sum_{r=1}^{K}\eta(r)^2\frac{\partial^2}{\partial\eta(r)^2}m(\eta)$$

$$= N\sum_{r=1}^{K}\eta(r)\frac{\partial^2}{\partial\eta(r)^2}m(\eta) - \sum_{r=1}^{K}\eta(r)^2\frac{\partial^2}{\partial\eta(r)^2}m(\eta).$$

The first summand in the last equality is an homogeneous polynomial of degree $L-1$ and the second one satisfies

$$-\sum_{r=1}^{K}\eta(r)^2\frac{\partial^2}{\partial\eta(r)^2}m(\eta) = -c_2\,m(\eta),$$

with

$$c_2 = \sum_{r=1}^{K}\alpha_r(\alpha_r-1) = \sum_{r=1}^{K}\alpha_r^2 - L.$$

As a conclusion, it comes from (B.2) that

$$\mathcal{A}_N V_m = -(c_1+c_2)V_m + V_q = -L(L-1)V_m + V_q,$$

where q is a polynomial of total degree strictly less than L, which proves (a).

$\square$

Appendix C. Proof of Lemma 1.9

First we prove Lemma C.1 showing that the neutral multi-allelic Moran process driven by $\mathcal{Q}_{N,p}$ is reversible if and only if its mutation rate matrix can be written in the form of $Q_{\boldsymbol{\mu}}$, given by (1.11). We start by proving that when the neutral multi-allelic Moran process is reversible, then all the entries of the mutation matrix are positive and it can be written in the form of $Q_{\boldsymbol{\mu}}$, i.e. the *"only if part"*. Later, in Lemma C.2 we prove that the process driven by $\mathcal{L}_{N,p}$ is reversible and we provide the explicit expression for its stationary distribution, i.e. we prove the *"if part"*. Actually, the results in Lemma C.2 are proved for a more general Moral model with *selection at birth*.

Lemma C.1. *If the process driven by the generator (1.2) is reversible, then $\mu_{i,j} = \mu_j > 0$, for all $i \in [K]$, and every $j \in [K]$, $j \neq i$.*

Proof: We first prove that if the process is reversible, then all the entries of the mutation matrix are positive. Let us denote by $\nu_{N,p}$ the stationary probability measure of the process driven by $\mathcal{Q}_{N,p}$, which is assumed to be reversible. We denote $\mathcal{Q}_{N,p}[\eta,\xi] := (\mathcal{Q}_{N,p}\delta_\xi)(\eta)$, for all $\eta,\xi \in \mathcal{E}_{K,N}$.

Consider the states $\eta^{(1)}$ and $\eta^{(2)}$ defined as $\eta^{(1)} := N\mathbf{e}_i$ and $\eta^{(2)} := \eta^{(1)} - \mathbf{e}_i + \mathbf{e}_j$, for $i, j \in [K]$ such that $i \neq j$. Since the process is reversible, the measure ν_N satisfies the balance equation

$$\nu_{N,p}(\eta^{(1)})\mathcal{Q}_{N,p}[\eta^{(1)}, \eta^{(2)}] = \nu_{N,p}(\eta^{(2)})\mathcal{Q}_{N,p}[\eta^{(2)}, \eta^{(1)}],$$

see e.g. (Kelly, 1979, Thm. 1.3). We have

$$\mathcal{Q}_{N,p}[\eta^{(1)}, \eta^{(2)}] = N\mu_{i,j}, \text{ and}$$

$$\mathcal{Q}_{N,p}[\eta^{(2)}, \eta^{(1)}] = \mu_{j,i} + p(N-1)/N > 0.$$

Furthermore, since the process is irreducible we have that $\nu_N(\eta) > 0$, for all $\eta \in \mathcal{E}_{K,N}$. Finally, the balance equation implies that $\mu_{i,j} > 0$, for all $i \neq j$.

Now, we prove that for every $j \in [K]$ we have $\mu_{i,j} = \mu_j > 0$, for all $i \in [K]$. For $K = 2$, there is nothing to prove. For $K \geq 3$, $N \geq 2$, let us consider a general model with a reversible stationary probability. Let i, j, k be three different indices on $[K]$ and consider the four states $\eta^{(1)}$, $\eta^{(2)}$, $\eta^{(3)}$ and $\eta^{(4)}$ in $\mathcal{E}_{K,N}$ defined by

$$\eta^{(1)} := N\mathbf{e}_i, \quad \eta^{(2)} := \eta^{(1)} - \mathbf{e}_i + \mathbf{e}_j, \quad \eta^{(3)} := \eta^{(1)} - 2\,\mathbf{e}_i + \mathbf{e}_j + \mathbf{e}_k, \quad \eta^{(4)} := \eta^{(1)} - \mathbf{e}_i + \mathbf{e}_k.$$

Note that

$$\begin{array}{ll}
\mathcal{Q}_{N,p}[\eta^{(1)}, \eta^{(2)}] = N\mu_{i,j}, & \mathcal{Q}_{N,p}[\eta^{(2)}, \eta^{(1)}] = \mu_{j,i} + (N-1)p/N, \\
\mathcal{Q}_{N,p}[\eta^{(2)}, \eta^{(3)}] = (N-1)\mu_{i,k}, & \mathcal{Q}_{N,p}[\eta^{(3)}, \eta^{(2)}] = \mu_{k,i} + (N-2)p/N, \\
\mathcal{Q}_{N,p}[\eta^{(3)}, \eta^{(4)}] = \mu_{j,i} + (N-2)p/N, & \mathcal{Q}_{N,p}[\eta^{(4)}, \eta^{(3)}] = (N-1)\mu_{i,j}, \\
\mathcal{Q}_{N,p}[\eta^{(4)}, \eta^{(1)}] = \mu_{k,i} + (N-1)p/N, & \mathcal{Q}_{N,p}[\eta^{(1)}, \eta^{(4)}] = N\mu_{i,k}.
\end{array}$$

Then,

$$\frac{\mathcal{Q}_{N,p}[\eta^{(4)}, \eta^{(1)}]}{N(N-1)} \prod_{r=1}^{3} \mathcal{Q}_{N,p}[\eta^{(r)}, \eta^{(r+1)}] = \mu_{i,j}\mu_{i,k}\left(\mu_{j,i} + p\frac{N-2}{N}\right)\left(\mu_{k,i} + p\frac{N-1}{N}\right),$$

$$\frac{\mathcal{Q}_{N,p}[\eta^{(1)}, \eta^{(4)}]}{N(N-1)} \prod_{r=1}^{3} \mathcal{Q}_{N,p}[\eta^{(r+1)}, \eta^{(r)}] = \mu_{i,k}\mu_{i,j}\left(\mu_{j,i} + p\frac{N-1}{N}\right)\left(\mu_{k,i} + p\frac{N-2}{N}\right).$$

Therefore, since the stationary probability is reversible, the *Kolmogorov cycle reversibility criterion* (Kelly, 1979, Thm. 1.8) holds:

$$\mathcal{Q}_{N,p}[\eta^{(4)}, \eta^{(1)}] \prod_{r=1}^{3} \mathcal{Q}_{N,p}[\eta^{(r)}, \eta^{(r+1)}] = \mathcal{Q}_{N,p}[\eta^{(1)}, \eta^{(4)}] \prod_{r=1}^{3} \mathcal{Q}_{N,p}[\eta^{(r+1)}, \eta^{(r)}],$$

and we get $p(N-1)\mu_{i,j}\mu_{i,k}(\mu_{j,i} - \mu_{k,i}) = 0$. We know that $\mu_{i,j} > 0$ for all $i, j \in [K]$, thus $\mu_{j,i} = \mu_{k,i}$, for all $j, k \in [K]$, with $j \neq k$, and every $i \in [K]$, with $i \notin \{j, k\}$. Denoting $\mu_j := \mu_{i,j}$ for any $i \in [K]$, with $i \neq j$, we prove that the mutation matrix is of the form of $Q_{\boldsymbol{\mu}}$ for a suitable vector $\boldsymbol{\mu}$. $\qquad\square$

It remains to prove that the stationary distribution of $\mathcal{L}_{N,p}$ is compound Dirichlet multinomial with suitable parameters. Actually, a more general version of Lemma 1.9 can be proved, where the values of the parameter p in (5.1) also depend on j, i.e. a model with selection at birth or fecundity selection (Muirhead and Wakeley, 2009). Abusing notation, for two vectors $\boldsymbol{p} = (p_1, p_2, \ldots, p_K)$ and $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_K)$ such that $p_j, \mu_j > 0$, for all $j \in [K]$, let us denote by $\mathcal{L}_{N,\boldsymbol{p}}$ the infinitesimal generator satisfying

$$(\mathcal{L}_{N,\boldsymbol{p}}f)(\eta) := \sum_{i,j=1}^{K} \eta(i)\left(\mu_j + p_j\frac{\eta(j)}{N}\right)[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta)], \qquad (C.1)$$

for every function f on $\mathcal{E}_{K,N}$ and all $\eta \in \mathcal{E}_{K,N}$. We define the *weighted Dirichlet-compound multinomial distribution* with parameters N, $\boldsymbol{\mu}$ and $\boldsymbol{p}$, denoted $\mathcal{WDM}(\cdot \mid N, \boldsymbol{\mu}, \boldsymbol{p})$, as follows

$$\mathcal{WDM}(\eta \mid N, \boldsymbol{\mu}, \boldsymbol{p}) := Z^{-1} \binom{N}{\eta} \prod_{k=1}^{K} p_k^{\eta(k)} (\alpha_k)_{(\eta(k))}, \tag{C.2}$$

for all $\eta \in \mathcal{E}_{K,N}$, where $\alpha_k = \mu_k / p_k$, for all $k \in [K]$ and Z is a normalisation constant satisfying

$$Z = \mathbb{E}\left[\left(\sum_{j=1}^{K} p_j X_j \right)^N \right], \tag{C.3}$$

where $(X_1, X_2, \ldots, X_K)$ follows a $\mathcal{DM}(\cdot \mid N, N\boldsymbol{\mu})$. Note that the measure defined by (C.2) with the normalisation constant (C.3) is a probability distribution. See (Johnson et al., 1997) and (Navarro et al., 2006) for more details about the weighted multinomial distributions.

Lemma C.2 (Reversible probability of $\mathcal{L}_{N,\boldsymbol{p}}$). *The process driven by (C.1) is reversible and its stationary distribution is $\mathcal{WDM}(\cdot \mid N, \boldsymbol{\alpha}, \boldsymbol{p})$, where $\alpha_k = N\mu_k$, for all $k \in [K]$.*

Remark C.3. This result is known for multi-allelic Moran models with parent independent mutation. See e.g. (Etheridge and Griffiths, 2009, Section 3). However, we have not found a proof in the literature. So, for the sake of completeness we provide a proof. When the vector $\boldsymbol{p}$ is constant we obtain the stationary distribution of the neutral case and we thus conclude the proof of Lemma 1.9.

Proof of Lemma C.2: Let us define $q_k := p_k/N$, for $k \in [K]$ and, abusing notation, $\mathcal{L}_{N,\boldsymbol{p}}[\eta, \xi] := \mathcal{L}_{N,\boldsymbol{p}} \delta_\xi(\eta)$, for all $\eta, \xi \in \mathcal{E}_{K,N}$. Note that for $\eta, \xi \in \mathcal{E}_{K,N}$ with $\eta \neq \xi$, we have $\mathcal{L}_{N,\boldsymbol{p}}[\eta, \xi] \neq 0$ if and only if there exist $i, j \in [K]$, such that $i \neq j$, $\eta(i) > 0$ and $\xi = \eta - \mathbf{e}_i + \mathbf{e}_j$. In this case

$$\mathcal{L}_{N,\boldsymbol{p}}[\eta, \xi] = \eta(i)[\mu_j + \eta(j)q_j].$$

This implies that $\xi(j) = \eta(j) + 1 > 0$ and $\eta = \xi - \mathbf{e}_j + \mathbf{e}_i$. As a consequence

$$\mathcal{L}_{N,\boldsymbol{p}}[\xi, \eta] = \xi(j)[\mu_i + \xi(i)q_i] = (\eta(j) + 1)[\mu_i + (\eta(i) - 1)q_i].$$

Also $\eta(k) = \xi(k)$, for all $k \neq i$, $k \neq j$.

Therefore we get,

$$Z \cdot \mathcal{WDM}(\eta \mid N, \boldsymbol{\mu}, \boldsymbol{p}) \mathcal{L}_{N,\boldsymbol{p}}[\eta, \xi] = \binom{N}{\eta} \left[\prod_{k=1}^{K} p_k^{\eta(k)} \left(\frac{\mu_k}{q_k} \right)_{(\eta(k))} \right] \eta(i)[\mu_j + \eta(j)q_j]$$

$$= \frac{N!}{\prod_{k \notin \{i,j\}} \eta(k)!} \frac{1}{\eta(i)!\eta(j)!} \left[\prod_{k=1}^{K} \prod_{l=0}^{\eta(k)-1} (\mu_k + l\,q_k) \right] \eta(i)[\mu_j + \eta(j)\,q_j], \tag{C.4}$$

where Z is the normalisation constant given by (C.3). Note that

$$\frac{N!}{\prod_{k \notin \{i,j\}} \eta(k)!} \prod_{k \notin \{i,j\}}^{K} \prod_{l=0}^{\eta(k)-1} (\mu_k + l\,q_k) = \frac{N!}{\prod_{k \notin \{i,j\}} \xi(k)!} \prod_{k \notin \{i,j\}}^{K} \prod_{l=0}^{\xi(k)-1} (\mu_k + l\,q_k), \tag{C.5}$$

because $\eta(k) = \xi(k)$, for $k \notin \{i,j\}$. Moreover,

$$\frac{1}{\eta(i)!\,\eta(j)!} \eta(i) = \frac{1}{(\eta(i) - 1)!\,\eta(j)!} = \frac{1}{\xi(i)!\,(\xi(j) - 1)!} = \frac{1}{\xi(i)!\,\xi(j)!} \xi(j), \tag{C.6}$$

because $\xi(i) = \eta(i) - 1$ and $\xi(j) = \eta(j) + 1$. In addition,

$$\prod_{l=0}^{\eta(i)-1} (\mu_i + l\,q_i) = \prod_{l=0}^{\xi(i)} (\mu_i + l\,q_i) = (\mu_i + \xi(i)\,q_i) \prod_{l=0}^{\xi(i)-1} (\mu_i + l\,q_i), \tag{C.7}$$

and

$$\left[\prod_{l=0}^{\eta(j)-1}(\mu_j + l\,q_j)\right][\mu_j + \eta(j)\,q_j] = \prod_{l=0}^{\eta(j)}(\mu_j + l\,q_j) = \prod_{l=0}^{\xi(j)-1}(\mu_j + l\,q_j). \tag{C.8}$$

Using (C.5), (C.6), (C.7) and (C.8) in (C.4) gives

$$\mathcal{WDM}(\eta \mid N, \boldsymbol{\mu}, \boldsymbol{p})\mathcal{L}_{N,\boldsymbol{p}}[\eta,\xi] = Z^{-1}\binom{N}{\xi}\left[\prod_{k=1}^{K}p_k^{\xi(k)}\left(\frac{\mu_k}{p_k}\right)_{(\xi(k))}\right]\xi(j)[\mu_i + \xi(i)q_i]$$
$$= \mathcal{WDM}(\xi \mid N, \boldsymbol{\mu}, \boldsymbol{p})\mathcal{L}_{N,\boldsymbol{p}}[\xi,\eta],$$

for all $\eta, \xi \in \mathcal{E}_{K,N}$. The distribution ν_N satisfies the detailed balance property, thus it is reversible for $\mathcal{L}_{N,\boldsymbol{p}}$, and it is the unique stationary measure, because the process generated by $\mathcal{L}_{N,\boldsymbol{p}}$ is irreducible. $\qquad\square$

Acknowledgements

References

Aldous, D. Random walks on finite groups and rapidly mixing Markov chains. In *Seminar on probability, XVII*, volume 986 of *Lecture Notes in Math.*, pp. 243–297. Springer, Berlin (1983). MR770418.

Aldous, D. and Diaconis, P. Shuffling cards and stopping times. *Amer. Math. Monthly*, **93** (5), 333–348 (1986). MR841111.

Asselah, A., Ferrari, P. A., and Groisman, P. Quasistationary distributions and Fleming-Viot processes in finite spaces. *J. Appl. Probab.*, **48** (2), 322–332 (2011). MR2840302.

Barrera, J., Lachaud, B., and Ycart, B. Cut-off for *n*-tuples of exponentially converging processes. *Stochastic Process. Appl.*, **116** (10), 1433–1446 (2006). MR2260742.

Brémaud, P. *Markov chains—Gibbs fields, Monte Carlo simulation and queues*, volume 31 of *Texts in Applied Mathematics*. Springer, Cham (2020). ISBN 978-3-030-45981-9; 978-3-030-45982-6. MR4174390.

Chen, G.-Y. *The cutoff phenomenon for finite Markov chains*. Ph.D. thesis, Cornell University (2006). MR2708835.

Chen, G.-Y., Hsu, J.-M., and Sheu, Y.-C. The L^2-cutoffs for reversible Markov chains. *Ann. Appl. Probab.*, **27** (4), 2305–2341 (2017). MR3693527.

Chen, G.-Y. and Kumagai, T. Cutoffs for product chains. *Stochastic Process. Appl.*, **128** (11), 3840–3879 (2018). MR3860012.

Chen, G.-Y. and Saloff-Coste, L. The cutoff phenomenon for ergodic Markov processes. *Electron. J. Probab.*, **13**, no. 3, 26–78 (2008). MR2375599.

Chen, G.-Y. and Saloff-Coste, L. The L^2-cutoff for reversible Markov processes. *J. Funct. Anal.*, **258** (7), 2246–2315 (2010). MR2584746.

Cloez, B. and Thai, M.-N. Fleming-Viot processes: two explicit examples. *ALEA Lat. Am. J. Probab. Math. Stat.*, **13** (1), 337–356 (2016a). MR3487076.

Cloez, B. and Thai, M.-N. Quantitative results for the Fleming-Viot particle system and quasi-stationary distributions in discrete space. *Stochastic Process. Appl.*, **126** (3), 680–702 (2016b). MR3452809.

Connor, S. B. Separation and coupling cutoffs for tuples of independent Markov processes. *ALEA Lat. Am. J. Probab. Math. Stat.*, **7**, 65–77 (2010). MR2644042.

Corujo, J. Dynamics of a Fleming-Viot type particle system on the cycle graph. *Stochastic Process. Appl.*, **136**, 57–91 (2021). MR4229456.

Couty, D., Esterle, J., and Zarouf, R. Décomposition effective de Jordan-Chevalley. *Gaz. Math.*, **129**, 29–49 (2011). MR2850401.

Davis, P. J. *Circulant matrices*. A Wiley-Interscience Publication. John Wiley & Sons, New York-Chichester-Brisbane (1979). ISBN 0-471-05771-1. MR543191.

Diaconis, P. and Griffiths, R. An introduction to multivariate Krawtchouk polynomials and their applications. *J. Statist. Plann. Inference*, **154**, 39–53 (2014). MR3258404.

Diaconis, P. and Griffiths, R. C. Reproducing kernel orthogonal polynomials on the multinomial distribution. *J. Approx. Theory*, **242**, 1–30 (2019). MR3915331.

Diaconis, P., Khare, K., and Saloff-Coste, L. Gibbs sampling, exponential families and orthogonal polynomials. *Statist. Sci.*, **23** (2), 151–178 (2008). MR2446500.

Diaconis, P. and Shahshahani, M. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, **57** (2), 159–179 (1981). MR626813.

Donnelly, P. and Rodrigues, E. R. Convergence to stationarity in the Moran model. *J. Appl. Probab.*, **37** (3), 705–717 (2000). MR1782447.

Durrett, R. *Probability models for DNA sequence evolution*. Probability and its Applications (New York). Springer, New York, second edition (2008). ISBN 978-0-387-78168-6. MR2439767.

Etheridge, A. *Some mathematical models from population genetics*, volume 2012 of *Lecture Notes in Mathematics*. Springer, Heidelberg (2011). ISBN 978-3-642-16631-0. Lectures from the 39th Probability Summer School held in Saint-Flour, 2009, École d'Été de Probabilités de Saint-Flour. MR2759587.

Etheridge, A. and Griffiths, R. A coalescent dual process in a Moran model with genic selection. *Theor. Popul. Biol.*, **75** (4), 320–330 (2009). Sam Karlin: Special Issue. DOI: 10.1016/j.tpb.2009.03.004.

Ethier, S. N. and Kurtz, T. G. Fleming-Viot processes in population genetics. *SIAM J. Control Optim.*, **31** (2), 345–386 (1993). MR1205982.

Ferrari, P. A. and Marić, N. Quasi stationary distributions and Fleming-Viot processes in countable spaces. *Electron. J. Probab.*, **12**, no. 24, 684–702 (2007). MR2318407.

Griffiths, R. C. The Λ-Fleming-Viot process and a connection with Wright-Fisher diffusion. *Adv. in Appl. Probab.*, **46** (4), 1009–1035 (2014). MR3290427.

Griffiths, R. C. and Spanò, D. Orthogonal polynomial kernels and canonical correlations for Dirichlet measures. *Bernoulli*, **19** (2), 548–598 (2013). MR3037164.

Harris, G. and Martin, C. The roots of a polynomial vary continuously as a function of the coefficients. *Proc. Amer. Math. Soc.*, **100** (2), 390–392 (1987). MR884486.

Iliev, P. and Xu, Y. Discrete orthogonal polynomials and difference equations of several variables. *Adv. Math.*, **212** (1), 1–36 (2007). MR2319761.

Ismail, M. E. H. *Classical and quantum orthogonal polynomials in one variable*, volume 98 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge (2005). ISBN 978-0-521-78201-2; 0-521-78201-5. MR2191786.

Johnson, N. L., Kemp, A. W., and Kotz, S. *Univariate discrete distributions*. Wiley Series in Probability and Statistics. Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, third edition (2005). ISBN 978-0-471-27246-5; 0-471-27246-9. MR2163227.

Johnson, N. L., Kotz, S., and Balakrishnan, N. *Discrete multivariate distributions*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New

York (1997). ISBN 0-471-12844-9. MR1429617.

Karlin, S. and McGregor, J. Ehrenfest urn models. *J. Appl. Probability*, **2**, 352–376 (1965). MR184284.

Karlin, S. and McGregor, J. Linear growth models with many types and multidimensional Hahn polynomials. In *Theory and application of special functions (Proc. Advanced Sem., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1975)*, Math. Res. Center, Univ. Wisconsin, Publ. No. 35, pp. 261–288. Academic Press, New York (1975). MR0406574.

Kelly, F. P. *Reversibility and stochastic networks.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, Ltd., Chichester (1979). ISBN 0-471-27601-4. MR554920.

Khare, K. and Zhou, H. Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions. *Ann. Appl. Probab.*, **19** (2), 737–777 (2009). MR2521887.

Koekoek, R., Lesky, P. A., and Swarttouw, R. F. *Hypergeometric orthogonal polynomials and their q-analogues.* Springer Monographs in Mathematics. Springer-Verlag, Berlin (2010). ISBN 978-3-642-05013-8. MR2656096.

Lacoin, H. A product chain without cutoff. *Electron. Commun. Probab.*, **20**, no. 19, 9 (2015). MR3320407.

Levin, D. A. and Peres, Y. *Markov chains and mixing times.* American Mathematical Society, Providence, RI (2017). ISBN 978-1-4704-2962-1. MR3726904.

Li, Z., Shiga, T., and Yao, L. A reversibility problem for Fleming-Viot processes. *Electron. Comm. Probab.*, **4**, 65–76 (1999). MR1711591.

Lovász, L., Pelikán, J., and Vesztergombi, K. *Discrete mathematics: Elementary and beyond.* Undergraduate Texts in Mathematics. Springer-Verlag, New York (2003). ISBN 0-387-95584-4. MR1952453.

Moran, P. A. P. Random processes in genetics. *Proc. Cambridge Philos. Soc.*, **54** (1), 60–71 (1958). DOI: 10.1017/S0305004100033193.

Mosimann, J. E. On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions. *Biometrika*, **49**, 65–82 (1962). MR143299.

Muirhead, C. A. and Wakeley, J. Modeling multiallelic selection using a Moran model. *Genetics*, **182** (4), 1141–1157 (2009). DOI: 10.1534/genetics.108.089474.

Navarro, J., Ruiz, J. M., and del Aguila, Y. Multivariate weighted distributions: a review and some extensions. *Statistics*, **40** (1), 51–64 (2006). MR2207404.

Nestoridi, E. and Olesker-Taylor, S. Limit profiles for reversible Markov chains. *ArXiv Mathematics e-prints* (2020). arXiv: 2005.13437.

Nestoridi, E. and Olesker-Taylor, S. Limit profiles for reversible Markov chains. *Probab. Theory Related Fields*, **182** (1-2), 157–188 (2022). MR4367947.

Nielsen, F. and Nock, R. On the chi square and higher-order chi distances for approximating f-divergences. *IEEE Signal Process. Lett.*, **21** (1), 10–13 (2014). DOI: 10.1109/LSP.2013.2288355.

Pease, M. C., III. *Methods of matrix algebra.* Mathematics in Science and Engineering, Vol. 16. Academic Press, New York-London (1965). MR0207719.

Saloff-Coste, L. Lectures on finite Markov chains. In *Lectures on probability theory and statistics (Saint-Flour, 1996)*, volume 1665 of *Lecture Notes in Math.*, pp. 301–413. Springer, Berlin (1997). MR1490046.

Schrempf, D. and Hobolth, A. An alternative derivation of the stationary distribution of the multivariate neutral Wright–Fisher model for low mutation rates with a view to mutation rate estimation from site frequency data. *Theor. Popul. Biol.*, **114**, 88–94 (2017). DOI: 10.1016/j.tpb.2016.12.001.

Serre, D. *Matrices: Theory and applications*, volume 216 of *Graduate Texts in Mathematics.* Springer, New York, second edition (2010). ISBN 978-1-4419-7682-6. MR2744852.

Szehr, O., Reeb, D., and Wolf, M. M. Spectral convergence bounds for classical and quantum Markov processes. *Comm. Math. Phys.*, **333** (2), 565–595 (2015). MR3296158.

Watterson, G. A. Markov chains with absorbing states: A genetic example. *Ann. Math. Statist.*, **32**, 716–729 (1961). MR125633.

Watterson, G. A. Lines of descent and the coalescent. *Theoret. Population Biol.*, **26** (1), 77–92 (1984). MR760232.

Ycart, B. Cutoff for samples of Markov chains. *ESAIM Probab. Statist.*, **3**, 89–106 (1999). MR1716128.

Zhou, H. *Examples of multivariate Markov chains with orthogonal polynomial eigenfunctions*. Ph.D. thesis, Stanford University (2008). MR2712374.

Zhou, H. and Lange, K. Composition Markov chains of multinomial type. *Adv. in Appl. Probab.*, **41** (1), 270–291 (2009). MR2514954.