On the spectrum of a neutral multi-allelic Moran model

Josué M. Corujo

ON THE SPECTRUM OF A NEUTRAL MULTI-ALLELIC MORAN MODEL

JOSUÉ CORUJO

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, which is non-reversible in general, continuous-time Markov chain with unknown stationary distribution. Specifically, we consider N individuals, where 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. We also study in detail the spectral decomposition of the neutral multi-allelic Moran model with parent independent mutation scheme, which turns to be the unique mutation scheme that makes the neutral Moran process reversible. Under the parent independent mutation assumption we provide a complete description of the left and right eigenfunctions of the infinitesimal generator of the Markov process. We use spectral techniques to provide non-asymptotic bounds for the rate of convergence of the neutral multi-allelic Moran process with parent independent mutation to stationarity in the total variation and chi-square distances. Moreover, we prove the existence of a strongly optimal cutoff phenomenon in the chi-square distance 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 the mutation process presents a strongly optimal total variation cutoff with Gaussian profile when initially all the individuals are of the same type.

Contents

1. Introduction and main results 1
Applications to the neutral multi-allelic Moran model with parent independent mutation 4
Cutoff phenomenon 7
Links with other models 9
Structure of the article 10
2. State spaces for distinguishable and indistinguishable particle processes 10
3. Spectrum of the neutral multi-allelic Moran process 13
3.1. Proof of Theorem 1.1 13
3.2. Proof of Theorem 1.2 17
3.3. Proof of Theorem 1.3 18
3.4. Applications to the exponential ergodicity of neutral multi-allelic Moran processes 19
4. Spectral decomposition of the neutral multi-allelic Moran type process with parent independent mutation 21
4.1. Background on reversible finite continuous-time Markov chains 21
4.2. Proof of Theorems 1.6 and 1.7 23
Multivariate orthogonal Hahn and Krawtchouk polynomials 24
Kernel polynomials for Dirichlet multinomial and multinomial distributions 25
5. Discussion and open problems 30
Appendix A. Proofs of Lemmas 2.1 and 2.3, and Proposition 2.4 31
Appendix B. Proof of Lemma 3.3 32
Appendix C. Proof of Theorem 1.5 33
Acknowledgement 36
References 36

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 population genetics literature [EG09]: the K sites correspond to K allelic types in a population of N individuals. The state space of

Date: October 2020.
1991 *Mathematics Subject Classification.* Primary 60J27; Secondary 37A30, 92D10, 33C50.
Key words and phrases. neutral multi-allelic Moran process; Fleming – Viot type particle system; interacting particle system; convergence rate to stationarity; finite continuous-time Markov chains; multivariate Hahn polynomials; cutoff.

the process is the K-dimensional N-discrete simplex:

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

where $[N]_0 := \{0, 1, \ldots, N\}$ and $|\cdot|$ stands for the sum of elements in a vector. $\mathcal{E}_{K,N}$ 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:

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

for all $\eta \in \mathcal{E}_{K,N}$, where $\mathbf{e}_k$ is the k-th canonical vector of $\mathbb{R}^K$ (cf. [EG09]). 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 [Mor58], the same individual removed from the population can be duplicated, in this case the state of the system does not change. In the instance 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 follows:

$$\mathcal{Q}_{N,p} = \mathcal{Q}_N + \frac{p}{N} \mathcal{A}_N,$$

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

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

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

for every $\eta \in \mathcal{E}_{K,N}$. The processes driven by $\mathcal{Q}_N$ and $\mathcal{A}_N$ are called *mutation process* and *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 *compound chain* (cf. [ZL09]). 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 for 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 $1 \leq n \leq N$. See e.g. [Lan02] and [Pea65, Chapter XIV] for an introduction to the concept of tensor product and the related concept of Kronecker product, respectively. 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:

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

Symmetric functions on $[K]^N$ are highly important in the sequel because they are very related to the functions on $\mathcal{E}_{K,N}$. Symmetric functions help us to clarify the relationship between the eigenstructures of Q and $\mathcal{Q}_N$. Consider the application $\psi_{K,N} : \mathcal{E}_{K,N} \to [K]^N$ defined by

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

when the number of k in $\overbrace{k, k, \dots, k}^{\eta(k)}$ is 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 $U_1, U_2, \dots, U_{K-1} \in \mathbb{R}^K$ such that $\mathcal{U} := \{U_0, U_1, \dots, U_{K-1}\}$ is a basis of $\mathbb{R}^K$. Note that this is the type of basis given by the eigenvectors of the diagonalisable rate matrix of dimension K of a Markov chain on $[K]$. For every $\eta \in \mathcal{E}_{K-1,L}$, $1 \le L \le N$, let us also denote by $U_\eta \in \mathbb{R}^{[K]^N}$, $V_\eta \in \mathrm{Sym}\left(\mathbb{R}^{[K]^N}\right)$ and $\tilde{V}_\eta \in \mathbb{R}^{\mathcal{E}_{K,N}}$ the vectors defined by

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

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

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

where $(k_1, k_2, \dots, k_L) = \psi_{K-1,L}(\eta)$, $\eta \in \mathcal{E}_{K-1,L}$ and $L \in [N]$. In Section 2 we analyse the link between the spaces $\mathrm{Sym}\left(\mathbb{R}^{[K]^N}\right)$ and $\mathbb{R}^{\mathcal{E}_{K,N}}$, and we clarify the nature of the definitions previously introduced. Next theorem below uses the results in Section 2 to detail the connection between the eigenstructures of Q and $\mathcal{Q}_N$

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

$$\lambda_\eta := \sum_{k=1}^{K-1} \eta(k) \lambda_k. \tag{1.10}$$

Then,

(a) *The 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 $\eta \in \bigcup_{L=1}^{N} \mathcal{E}_{K-1,L}$ 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.*

The proof of Theorem 1.1 can be found in Section 3.1. Theorem 1.1 can be seen as a continuous-time generalisation of the results provided by Zhou and Lange [ZL09] for the discrete-time analogous of the mutation process driven by $\mathcal{Q}_N$. We emphasize that our hypotheses do not require the mutation rate matrix Q to be diagonalisable.

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

Theorem 1.2 (Spectrum of $\mathcal{A}_N$). *Let $K \ge 3$ and $N \ge 2$. The eigenvalues of $\mathcal{A}_N$ are*

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

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

The proof of Theorem 1.2 is deferred to Section 3.2. Theorem 1.2 can be seen as a generalisation, for $K \ge 3$, of the results in Section 4.2.2 of [Zho08] for the discrete analogous of the reproduction process driven by $\mathcal{A}_N$, for $K = 2$.

Unlike in the 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 harder the study of its spectrum. Our main result is precisely a complete description of the eigenvalues of the generator $\mathcal{Q}_{N,p}$, which is expressed in the following theorem.

Theorem 1.3 (Spectrum of $\mathcal{Q}_{N,p}$). *Assume $K \geq 2$, $N \geq 3$ and $p \in [0, \infty)$. Let us denote by λ_k, $k \in [K-1]$, the non-zero $K-1$ 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 1.3 is given in Section 3.3.

Remark 1.1 (Monotony in N of the spectrum of $\mathcal{Q}_{N,p}$). *Theorem 1.3 implies that the spectrum of $\mathcal{Q}_{N,p}$, for fixed values of K and $p \geq 0$, is an increasing function of N in the sense of the inclusion of sets.*

As a consequence of Theorem 1.3, the *second largest eigenvalue in modulus* (SLEM) of $\mathcal{Q}_{N,p}$ is equal to that of Q. The SLEM of the generator of the process is useful to study the asymptotic convergence of the process in total variation. Hence, in Section 3.4 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.

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

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

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

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

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. [Eth11]). 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 $\eta \in \mathcal{E}_{K,N}$.

The next result explicitly describes the spectrum of $\mathcal{L}_{N,p}$ and it is a consequence of Theorem 1.3.

Corollary 1.4 (Spectrum of $\mathcal{L}_{N,p}$). *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 $\rho = |\boldsymbol{\mu}|$.

Corollary 1.4 is proved in Section 4.2.

Remark 1.2 (Complete graph model)**.** *The* complete graph model *studied by Cloez and Thai* [CT16a] *is a particular case of the reversible process driven by* $Q_{\boldsymbol{\mu}}$ *above when* $\mu_j = \frac{1}{K}$*, 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, Corollary 1.4 improves the Lemma 2.14 in* [CT16a].

Remark 1.3 (Spectral gap and the exponential ergodicity coefficient in [CT16b])**.** *Cloez and Thai* [CT16b] *considere a Fleming – Viot (or Moran type) particle system and coupling arguments to prove a contraction result for a suitable Wasserstein distance. They thus obtain the exponential ergodicity of the process when the time goes to infinity. The exponential ergodicity rate in* [CT16b, Theorem 1.1] *for the process driven by (1.2) and adapted to our notation is*

$$\rho^\star := \inf_{i,i' \in [K]} \left(\mu_{i,i'} + \mu_{i',i} + \sum_{j \neq i,i'} \mu_{i,j} \wedge \mu_{i',j} \right),$$

where $Q = (\mu_{i,j})_{i,j \in [K]}$ *is the mutation rate matrix of the process driven by* $\mathcal{Q}_{N,p}$*, defined in (1.2). This constant* $\rho^\star$ *is a lower bound for the spectral gap of the process, i.e.* $\rho^\star \geq \rho$*, see e.g. Theorem 13.1 in* [LP17]*. For the neutral multi-allelic Moran model driven by* $\mathcal{L}_{N,p}$*, the ergodicity constant of Cloez and Thai is optimal in the sense that it is equal to the spectral gap of the process:* $\rho^\star = |\boldsymbol{\mu}|$*. This does not happen in general: consider for example the neutral multi-allelic Moran model where the mutation process is a simple asymmetric random walk on* $[K]$*. In this case, the coefficient* $\rho^\star$ *vanishes but, as we prove in Example 3, the* SLEM *of the Moran process with mutation rate matrix* Q_θ *is equal to* $2(1+\theta)\sin^2\left(\frac{\pi}{K}\right)$*.*

For a real x and $n \in \mathbb{N}_0$, 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

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

for all $n > 0$ and $\eta \in \mathcal{E}_{K,N}$, respectively. We set by convention $x_{(0)} := 1$ and $x_{[0]} := 1$, even for $x = 0$.

The *multinomial distribution* distribution on $\mathcal{E}_{K,N}$ with parameters N and $\mathbf{q} = (q_1, \dots, q_K) \in (0,1)^K$ 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 $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \dots, \alpha_K) \in (0,\infty)^K$, 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 $\eta \in \mathcal{E}_{K,N}$. $\mathcal{DM}(\cdot \mid N, \boldsymbol{\alpha})$ is a mixture, using a Dirichlet distribution, of $\mathcal{M}(\cdot \mid N, \mathbf{q})$. See Mosimann [Mos62] for the original reference to the Dirichlet multinomial distribution and Johnson et al. [JKK05, Section 13.1], a classical reference on multivariate discrete distributions, for more details.

It is known in 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. [EG09]. Moreover, the stationary distribution of the process driven by $\mathcal{L}_N$ is $\mathcal{M}(\cdot \mid N, \boldsymbol{\mu}/|\boldsymbol{\mu}|)$, see e.g. [ZL09]. 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 [ZL09] 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 4.2. Going further, the next result characterises the reversible neutral multi-allelic Moran processes as those with parent independent mutation.

Theorem 1.5 (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 $\boldsymbol{\mu}$, and consequently $\mathcal{Q}_{N,p}$ 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 proof of Theorem 1.5 is deferred to Appendix C.

Remark 1.4 (Case $K = 2$: beta-binomial distribution)**.** *For the special case where $K = 2$ and $p > 0$, the stationary distribution of the N-particle process is $\mathcal{DM}(\cdot \mid N, \mu_1 N/p, \mu_2 N/p)$. This distribution can be seen as a univariate distribution, in the sense that $\nu_{N,p}(n, N-n)$ only depends on n for $n \in [N]_0$. This univariate distribution is known as* beta-binomial distribution*, see e.g. Chapter 6 of* [JKK05]*. Analogously, the stationary distribution of $\mathcal{L}_N$ is binomial with parameters N and $\mu_1/(\mu_1+\mu_2)$.*

Section 4 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. [KM75, KZ09]). 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. [KM65, ZL09, DG14]). In addition, we study the speed of convergence of the N-particle process to stationarity when t tends towards infinity, providing explicit non-asymptotic bounds for the distance to stationarity, for an arbitrary initial configuration.

For two probability measures μ_1 and μ_2 defined on the same discrete space Ω, we define the *total variation distance* as follows:

$$\mathrm{d}^{\mathrm{TV}}(\mu_1,\mu_2) := \sup_{A\subset\Omega} |\mu_1(A)-\mu_2(A)| = \sup_{f:\Omega\to[-1,1]} \left| \int f \mathrm{d}\mu_1 - \int f \mathrm{d}\mu_2 \right| = \frac{1}{2}\|\mu_1-\mu_2\|_1, \tag{1.14}$$

where $\|\cdot\|_1$ denotes the 1-norm in $\mathbb{R}^\Omega$. Moreover, the *chi-square divergence* of μ_2 with respect to the target distribution μ_1 is defined by

$$\chi^2(\mu_2 \mid \mu_1) := \sum_{\omega\in\Omega} \frac{[\mu_2(\omega)-\mu_1(\omega)]^2}{\mu_1(\omega)} = \|\mu_2-\mu_1\|^2_{\frac{1}{\mu_1}},$$

where $\|\cdot\|_{\frac{1}{\mu_1}}$ stands for the norm in $l^2(\mathbb{R}^\Omega, \frac{1}{\mu_1})$, and $\frac{1}{\mu_1}$ 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 [NN14]. In this context, the chi-square divergence is also known as *Pearson chi-square divergence*.

The functions $\mathrm{d}^{\mathrm{TV}}_\eta$ and χ^2_η, defined by

$$\begin{aligned}
\mathrm{d}^{\mathrm{TV}}_\eta(t) &:= \mathrm{d}^{\mathrm{TV}}(\delta_\eta \mathrm{e}^{t\mathcal{L}_{N,p}}, \nu_{N,p}) = \frac{1}{2}\sum_{\xi\in\mathcal{E}_{K,N}} \left| \left(\mathrm{e}^{t\mathcal{L}_{N,p}}\delta_\xi\right)(\eta) - \nu_{N,p}(\xi)\right|, \\
\chi^2_\eta(t) &:= \chi^2(\delta_\eta \mathrm{e}^{t\mathcal{L}_{N,p}} \mid \nu_{N,p}) = \sum_{\xi\in\mathcal{E}_{K,N}} \frac{\left[\left(\mathrm{e}^{t\mathcal{L}_{N,p}}\delta_\xi\right)(\eta) - \nu_{N,p}(\xi)\right]^2}{\nu_{N,p}(\xi)},
\end{aligned} \tag{1.15}$$

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 [Zho08, KZ09] we call $\mathrm{d}^{\mathrm{TV}}_\eta$ and χ^2_η the *total variation* and the *chi-square distances* to stationarity, respectively.

For any function f in $\mathbb{R}^{[K]}$, let us denote by $\tilde{\xi}(f)$ the linear function on $\mathcal{E}_{K,N}$ defined as

$$\tilde{\xi}(f) : \eta \mapsto f(1)\eta(1) + f(2)\eta(2) + \cdots + f(K)\eta(K).$$

Denote $\langle \boldsymbol{\mu}, f\rangle := \sum_{k=1}^K f(k)\mu_k$. The next theorem provides non-asymptotic bounds for the exponential speed of convergence of the process driven by $\mathcal{L}_{N,p}$ to stationarity.

Theorem 1.6 (Exponential ergodicity in total variation and chi-square distances)**.** *For $K \geq 2$, $N \geq 2$ and $p \geq 0$, we obtain the following non-asymptotic bounds for the speed of convergence to the stationary of the process driven by $\mathcal{L}_{N,p}$:*

$$\left(\max_{f:\langle\boldsymbol{\mu},f\rangle=0} \frac{|\tilde{\xi}(f)(\eta)|}{2N\|f\|_\infty}\right)\mathrm{e}^{-|\boldsymbol{\mu}|t} \leq \mathrm{d}^{\mathrm{TV}}_\eta(t) \leq \frac{1}{2}\sqrt{\chi^2_\eta(t)} \leq \frac{1}{2}\sqrt{\frac{1-\nu_{N,p}(\eta)}{\nu_{N,p}(\eta)}}\mathrm{e}^{-|\boldsymbol{\mu}|t}, \tag{1.16}$$

where the maximum is taken over all the functions f on $\mathbb{R}^K$ which are orthogonal to $\boldsymbol{\mu}$ and $\|f\|_\infty = \max_{k\in[K]} |f(k)|$. In particular, for $\eta = N\mathbf{e}_k$ we get

$$\frac{\kappa_k}{2}\mathrm{e}^{-|\boldsymbol{\mu}|t} \le \mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t) \le \frac{1}{2}\sqrt{\chi^2_{N\mathbf{e}_k}(t)} \le \frac{1}{2}\sqrt{C_{k,p}}\,\mathrm{e}^{-|\boldsymbol{\mu}|t}, \tag{1.17}$$

where

$$\kappa_k = \max_{r\neq k} \frac{\mu_k \wedge \mu_r}{\mu_k},$$

$$C_{k,p} = \begin{cases} \dfrac{(N|\boldsymbol{\mu}|/p)_{(N)} - (N\mu_k/p)_{(N)}}{(N\mu_k/p)_{(N)}} & \text{if} \quad p > 0 \\ \dfrac{|\boldsymbol{\mu}|^N - \mu_k^N}{\mu_k^N} & \text{if} \quad p = 0. \end{cases}$$

Proof of Theorem 1.6 can be found in Section 4.2.

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. [DS81, Ald83, AD86]). A Markov chain presents a cutoff if it exhibits a sharp 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 [LP17, Chapter 18] and in the exhaustive work of Chen, Saloff-Coste et al. [SC97, Che06, CSC08, CSC10, CHS17].

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, model which is usually known as *product chain*. Ycart [Yca99] studied the cutoff in total variation for N independent particles driven by a diagonalisable rate matrix. Later, Barrera et al. [BLY06] and Connor [Con10] studied the cutoff on this model according to other notions of distance. See also [Lac15] [LP17, Chapter 20], [CHS17] and [CK18] 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.

As the number of individuals varies we obtain an infinite family of continuous-time finite Markov chains $\{(\mathcal{E}_{K,N}, \mathcal{L}_{N,p}, \nu_{N,p}), N \ge 2\}$. For each $N \ge 2$ let us denote by $\chi^2_{N\mathbf{e}_k}(t)$ (resp. $\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t)$) 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 $\chi^2_{N\mathbf{e}_k}(0) \to \infty$ and $\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(0) \to 1$, when $N \to \infty$.

We say that $\{\chi^2_{N\mathbf{e}_k}(t), N \ge 2\}$ exhibits a (t_N, b_N) chi-square cutoff if $t_N \ge 0$, $b_N \ge 0$, $b_N = o(t_N)$ and

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

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

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

See Definition 2.1 and Remark 2.1 in [CSC08].

The sequences $(t_N)_{N\ge 2}$ and $(b_N)_{N\ge 2}$ are called *cutoff* and *window sequences*, respectively. The cutoff phenomenon describes a sharp 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 $(t_N)_{N\ge 2}$.

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

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

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 [CSC08]. See Sections 2.1 and 2.2 of [CSC08] and Chapter 2 in [Che06] for more details about the definition of (t_N, b_N) cutoff and window optimality.

The next result establishes a strongly optimal cutoff in the chi-square distance 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 $k \in [K]$. Moreover, we prove the Gaussian profile, when $N \to \infty$, of the total variation distance to stationarity of the mutation process driven by $\mathcal{L}_N$, i.e. for $p = 0$, when all the individuals are initially of the same type.

Theorem 1.7 (Strongly optimal cutoffs when $N \to \infty$)**.** *For $k \in [K]$, with $K \geq 2$, $p \geq 0$ and every $c \in \mathbb{R}$, we have*

$$\lim_{N\to\infty} \chi^2_{N\mathbf{e}_k}(t_{N,c}) = \exp\{K_{k,p}\mathrm{e}^{-c}\} - 1, \tag{1.18}$$

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 $\left(\frac{\ln(N)}{2|\boldsymbol{\mu}|}, 1\right)$ chi-square cutoff when $N \to \infty$.

Moreover, when $p = 0$ we get the following result for the total variation distance to stationarity

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

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

Proof of Theorem 1.7 will be given in Section 4.2.

Figures 1 and 2 illustrate the convergence of $\chi^2_{N\mathbf{e}_k}(t_{N,c})$ and $\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t_{N,c})$, for $t_{N,c} = \frac{\ln(N)+c}{2|\boldsymbol{\mu}|}$, to $G_k(c) = \exp\{K_{k,p}\mathrm{e}^{-c}\} - 1$ and $H_k(c) = 2\Phi\left(\frac{1}{2}\sqrt{K_{k,0}\mathrm{e}^{-c}}\right) - 1$, respectively, when $N \to \infty$.

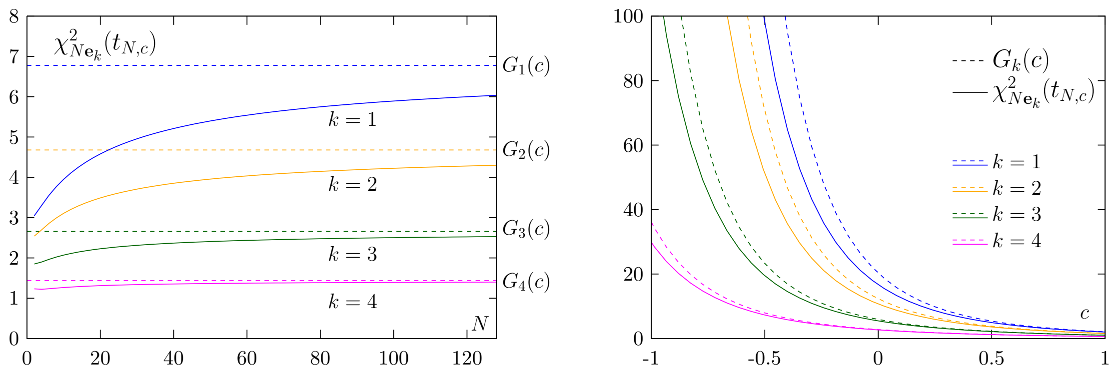


FIGURE 1. For parameters $K = 4$, $k \in [K]$, $\mu = (0.7, 0.8, 1.0, 1.3)$, $p = 1.7$: left panel shows $\chi^2_{N\mathbf{e}_k}(t_{N,c})$ as a function of N, $2 \leq N \leq 128$ for $c = 0.4$, and right panel shows $G_k(c)$ and $\chi^2_{N\mathbf{e}_k}(t_{N,c})$ as functions of c, with $-1 \leq c \leq 1$, for $N = 150$.

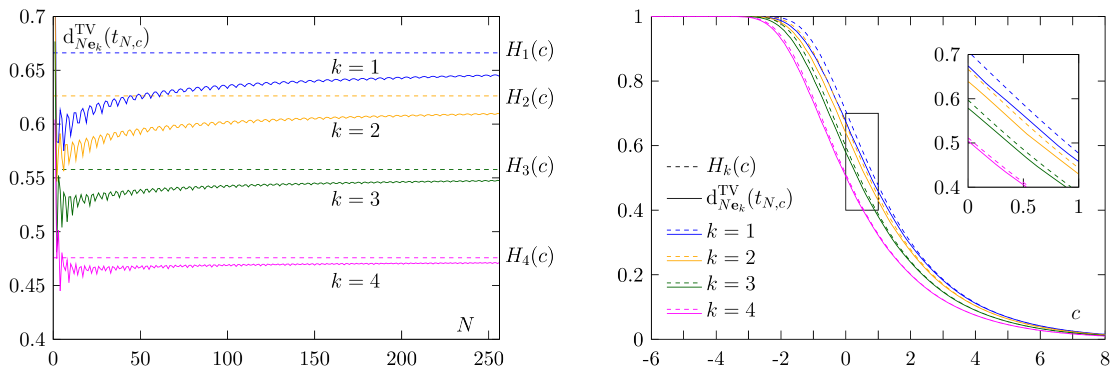


FIGURE 2. For parameters $K = 4$, $k \in [K]$, $\mu = (0.7, 0.8, 1.0, 1.3)$, $p = 0$: left panel shows $\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t_{N,c})$ as a function of N, $2 \leq N \leq 256$ for $c = 0.17$ and right panel shows $H_k(c)$ and $\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t_{N,c})$ as functions of c, with $-6 \leq c \leq 8$, for $N = 100$.

During the proof of Theorem 1.7, we prove the following result which is of independent interest.

Corollary 1.8 (Law of the process driven $\mathcal{L}_N$)**.** *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 $\mathcal{M}\left(\cdot \mid N, \frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|}(1 - \mathrm{e}^{-|\boldsymbol{\mu}|t}) + \mathrm{e}^{-|\boldsymbol{\mu}|t}\mathbf{e}_k\right)$.*

Some authors have studied the existence of a cutoff in Moran type models. For instance, Donelly and Rodrigues [DR00] 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 [KZ09] 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 [DG19] studied the existence of a chi-square and total variation cutoffs for a discrete-time analogous of the mutation process generated by $\mathcal{L}_N$. Our methods for proving Theorem 1.7 could be used to improve the results related to the existence of cutoffs in [KZ09] and [DG19], in order to obtain the explicit profiles for the limits of the chi-square distance, for $p \geq 0$, and the total variation distance, for $p = 0$, when $N \to \infty$.

Links with other models. Moran type models are fundamental in population genetics and other branches of applied mathematics [Dur08], [Eth11]. Simpler than the Wright – Fisher model, the Moran model is more tractable mathematically and several quantities of interest can be explicitly computed. There is a rich literature on Moran models in population genetics and other fields, since the seminal work of Moran [Mor58]. In particular, the study of spectral properties of the generator of a Markov process is an interesting and active topic of research in population genetics. See e.g. [KZ09], [ZL09], [MP14], [Möh18], [Möh19] and the references therein.

We want to remark that the utility of Moran processes is behind population genetics. For instance, the mutation process driven by $\mathcal{Q}_N$ is a particular case of the *zero range process*, where the *kinetics*, i.e. the rate at which the particles are expelled from one state, is proportional to the number of particles occupying that state. Moreover, the mutation process driven by $\mathcal{L}_N$ corresponds to the *mean-field* version of the zero range process. The very recent paper of Hermon and Salez [HS19] shows that the *Dirichlet form* of a zero range process can be controlled in terms of the *Dirichlet form* of a single particle. We believe that the methods in [HS19] could be very useful for the further study of the ergodicity of the Moran process driven by $\mathcal{Q}_{N,p}$, for $p \geq 0$, by controlling its *Dirichlet form*.

Consider a Markov process in $\mathcal{E}_{K,N}$ with generator $\mathcal{F}$ acting on a real function f on $\mathcal{E}_{K,N}$ as follows

$$(\mathcal{F}f)(\eta) = \sum_{i,j \in [K]} \eta(i) \left[f(\eta - \mathbf{e}_i + \mathbf{e}_j) - f(\eta) \right] \left(\mu_{i,j} + \frac{p_i}{N-1} \eta(j) \right), \tag{1.19}$$

for every $\eta \in \mathcal{E}_{K,N}$, where $p_i \geq 0$, for all $i \in [K]$. The process driven by $\mathcal{F}$ is a particular case of the countable state space continuous-time Markov processes introduced by Ferrari and Marić [FM07] to approximate the *quasi stationary distribution* (QSD) of an absorbing Markov chain on a countable space. Ferrari and Marić called these Markov chains *Fleming – Viot particle processes.* The random empirical distribution associated to the process driven by $\mathcal{F}$ has been proved to approximate the QSD of an absorbing Markov process driven by an irreducible rate matrix Q on $[K]$ which jumps, with rate p_i, from i to a fictitious absorbing state [AFG11]. This kind of N particle interacting process was originally introduced independently and simultaneously by Burdzy et. al. [BHM00] and Del Moral and Miclo [DMM00] [1] in the continuous state space settings. The study of the evolution of the proportion of particles in each state for a Moran-type particle system driven by $\mathcal{F}$ is an active topic of research. In particular, many papers have been focused on the convergence and the speed of convergence of the proportion of particles in each state when the time and the number of particles tend toward infinity. See e.g. [FM07], [AFG11], [CT16a], [CT16b], [Vil20] and the references therein.

The generator $\mathcal{F}$ is also interesting in population genetics. From this point of view, it models the evolution of a population with an irreducible mutation process driven by $Q = (\mu_{i,j})_{i,j=1}^K$ and *selection at death* (cf. [MW09]). Unlike the other type of selection that has been mostly considered in population genetics, which is the *selection at reproduction* (cf. [Dur08], [MW09] and [Eth11]), which assumes that the rates p_i in the definition (1.19) do not depend on i but on j, i.e. on the type of the individual that is going to reproduce.

[1] Actually, Del Moral and Miclo called this process *Moran-type particle system*, which is maybe a more accurate name in the discrete state space setting, in order to avoid confusion with the existence of a measure-valued process related to the Moran process in population genetics introduced by Fleming and Viot [FV79] and named *Fleming – Viot process.* See [FV79, Appendix B] for a discussion on the relationship between Fleming –Viot process (in the sense of population genetics) and the multi-allelic Moran model, see also [Fen10, Section 6.2].

Note that when $p_i = p$, for all $i \in [K]$, the generator $\mathcal{F}$ reduces to $\mathcal{Q}_{N,p}$. Theorem 1.3 thus provides an explicit description for the eigenvalues of the Fleming–Viot (or Moran type) particle process with irreducible mutation rate matrix Q and the transition rate to the absorbing state is uniform on $[K]$, which is known in the theory of QSD as *uniform killing* [MV12, Section 2.3]. This is, for example, the case of the complete graph process studied by Cloez and Thai [CT16a] and the neutral Moran model process with circulant mutation rate matrix considered in [Cor20].

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, 1.2 and 1.3, respectively. In Section 3.4 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. We also consider several examples of neutral multi-allelic Moran processes with diagonalisable and non-diagonalisable mutation rate matrices, throughout the paper. In Section 4 we consider the neutral multi-allelic Moran process with parent independent mutation and provide a complete description of its eigenvalues and eigenfunctions. Using spectral techniques, we prove Theorem 1.6 which provides detailed estimates of the rate of convergence of the process to its stationary distribution. We also prove Theorem 1.7 about the existence of a strongly optimal cutoff phenomenon in the chi-square distance for $p \geq 0$, and in the total variation distance, for $p = 0$, 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 in [FM07] and [Eth11]. 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 [CT16b] and [EG09], 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

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

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

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

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 $P(x_1, x_2, \dots, x_K) \in H_{K,L}$ with

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

We will say that the *degree of homogeneity* of an homogeneous polynomial P is L, if P is the sum of monomials $\mathbf{x}^\alpha = x_1^{\alpha_1} x_2^{\alpha_2} \dots x_K^{\alpha_K}$ 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 $P(x_1, \dots, x_{K-1}, N - \sum_{k=1}^{K-1} x_k)$ of maximum total degree to define P_L.

The following interpolation result shows that $\mathbb{R}^{\mathcal{E}_{K,N}}$ is actually equal to $H_{K,N}$, the space of homogeneous polynomials of degree N.

Lemma 2.1 (Interpolation on $\mathcal{E}_{K,N}$). *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 2.1 (Dimension of $H_{K,N}$). *As a consequence of Lemma 2.1-(b) we get* $\dim(H_{K,N}) = \binom{K+N-1}{N}$.

Example 1 (Interpolation polynomials on $\mathcal{E}_{3,4}$). *For $K = 3$, $N = 4$, one may check that*

$$\begin{aligned}
P_{[4,0,0]} : (x_1, x_2, x_3) &\mapsto \frac{1}{768} x_1(3x_1 - x_2 - x_3)(x_1 - x_2 - x_3)(x_1 - 3x_2 - 3x_3),\\
P_{[3,1,0]} : (x_1, x_2, x_3) &\mapsto \frac{1}{48} x_1 x_2 (x_1 - x_2 - x_3)(3x_1 - x_2 - x_3),\\
P_{[2,2,0]} : (x_1, x_2, x_3) &\mapsto -\frac{1}{64} x_1 x_2 (3x_1 - x_2 - x_3)(x_1 - 3x_2 + x_3),\\
P_{[2,1,1]} : (x_1, x_2, x_3) &\mapsto \frac{1}{8} x_1 x_2 x_3 (3x_1 - x_2 - x_3).
\end{aligned}$$

All the other P_η, for $\eta \in \mathcal{E}_{3,4}$, can be deduced using the permutations of $\{1, 2, 3\}$.

A natural link between the two state spaces is $\phi_{K,N} : [K]^N \to \mathcal{E}_{K,N}$, defined by

$$\phi_{K,N} : (k_1, k_2, \dots, k_N) \mapsto (\eta(1), \eta(2), \dots, \eta(K)), \tag{2.2}$$

where $\eta(k) = \mathrm{Card}(\{n,\ 1 \leq n \leq N,\ k_n = k\})$, for all $k \in [K]$. The function $\phi_{K,N}$ is obtained by *forgetting 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. $\phi_{K,N} \circ \psi_{K,N} = \mathrm{Id}_{\mathcal{E}_{K,N}}$, where $\mathrm{Id}_{\mathcal{E}_{K,N}}$ 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 projector onto the subspace of symmetric functions, denoted $\mathrm{Sym}(\mathbb{R}^{[K]^N})$. Indeed, note that the space of symmetric function is a linear subspace of $\mathbb{R}^{\mathcal{E}_{K,N}}$. In addition, for every function h symmetric we have $\mathrm{Sym}\, h = h$, thus $\mathrm{Sym}\,(\mathrm{Sym}\, f) = \mathrm{Sym}\, f$, for each $f \in \mathbb{R}^{[K]^N}$, i.e. $\mathrm{Sym}^2 = \mathrm{Sym}$. See [Pea65, Chapter XI] for more details about projectors.

Note that $\phi_{K,N}$ is a symmetric function on $[K]^N$. Furthermore, the equality $\phi_{K,N}(\mathbf{x}) = \phi_{K,N}(\mathbf{y})$ 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 2.2 (Link between $\mathbb{R}^{\mathcal{E}_{K,N}}$ and $\mathrm{Sym}(\mathbb{R}^{[K]^N})$). *The linear operator*

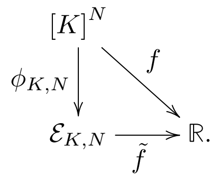
$$\Phi_{K,N} : f \in \mathrm{Sym}\left(\mathbb{R}^{[K]^N}\right) \mapsto f \circ \psi_{K,N} \in \mathbb{R}^{\mathcal{E}_{K,N}}, \tag{2.3}$$

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

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

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. □

Lemma 2.2 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:



We denote by U_0 the K-dimensional all-one vector. This vector 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 $V_1, V_2, \dots, V_L$ 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,

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

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 $s_\sigma : n \in [L] \mapsto \sigma(n) \in \{\sigma(1), \dots, \sigma(L)\}$ is an injective map in $\mathcal{I}_{L,N}$ and σ is completely determined by this function s_σ and a bijective application $\beta : (L+1, \dots, N) \to [N] \setminus s_\sigma([L])$. For each s_σ, there are $(N-L)!$ such applications β. Thus, using (2.4) we obtain

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

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

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

Note that $\xi(V_1, V_2, \dots, V_L) = \frac{N!}{(N-L)!} f$. Since $\xi(V_1, V_2, \dots, V_L)$ is symmetric, using Lemma 2.2, there exists a unique function $\tilde{\xi}(V_1, V_2, \dots, V_L)$ on $\mathcal{E}_{K,N}$ given by

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

The following two equalities are thus satisfied:

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

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

The next result provides recursive expressions for the functions $\xi(V_1, \dots, V_L)$ and $\tilde{\xi}(V_1, \dots, V_L)$, 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 2.3. *The following properties are verified:*

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

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

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

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

$$\begin{aligned} \xi(V_1, \dots, V_L) \quad &= \quad \xi(V_1, \dots, V_{L-1}) \xi(V_L) \\ &\quad - \sum_{i=1}^{L-1} \xi(V_1, \dots, V_{i-1}, V_{i,L}, V_{i+1}, \dots, V_{L-1}), \end{aligned}$$

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

where $V_{i,L}$ is the vector defined by $V_{i,L}(k) = V_i(k)V_L(k)$, for all $1 \le k \le K$.

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

$$\tilde{\xi}(V_1,V_2) : (\eta(1),\eta(2),\ldots,\eta(K)) \mapsto (a_1\eta(1)+\cdots+a_K\eta(K))(b_1\eta(1)+\cdots+b_K\eta(K)) - (a_1b_1\eta(1)+\cdots+a_Kb_K\eta(K)).$$

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

$$\tilde{\xi}(V_1,V_2,\ldots,V_L)(\eta) = \prod_{i=1}^{L}(a_{i,1}\eta(1)+a_{i,2}\eta(2)+\cdots+a_{i,K}\eta(K)) + q(\eta), \tag{2.8}$$

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 2.3 can be found in Appendix A.

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

Proposition 2.4. *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 hold:

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*

$$\tilde{\mathcal{S}}^N := \{\underbrace{U_0 \otimes \cdots \otimes U_0}_{K\ times}\} \cup \bigcup_{L=1}^{N}\{\tilde{V}_\eta, \eta \in \mathcal{E}_{K-1,L}\}, \tag{2.9}$$

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

The proof of Proposition 2.4 is deferred to Appendix A.

3. Spectrum of the neutral multi-allelic Moran process

The main goal of this section is to prove Theorem 1.3 and study some of its consequences. 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 1.2. Using the results in these two sections we prove Theorem 1.3 in Section 3.3. Section 3.4 is devoted to some applications of Theorem 1.3 to the study of the exponential ergodicity of the process driven by $\mathcal{Q}_{N,p}$ in total variation, using spectral properties of Q.

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

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

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

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

where $QV_n(k) := \sum_{r=1}^{K} \mu_{k,r} V_n(r) = \sum_{r=1}^{K} \mu_{k,r}(V_n(r) - V_n(k))$, 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 [Pea65, Chapter XIV] *and* [Dav79, Section 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 eigenbasis of Q form an eigenbasis of $\mathcal{D}_N$, as in Proposition 2.4-(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

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

See Sections 12.4 and 20.4 in [LP17] 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.

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

for all $\eta \in \mathcal{E}_{K,N}$ and for every function f on $\mathcal{E}_{K,N}$. Zhou and Lange [ZL09] 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 [ZL09]: 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 $\mathcal{D}_N$ and $\mathcal{Q}_N$ through the operator $\Phi_{K,N}$.

Lemma 3.1 (Link between the generators of the distinguishable and indistinguishable cases)**.** *For any symmetric function ξ on $[K]^N$, the function $\mathcal{D}_N\, \xi$ is also symmetric. In addition,*

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

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

$$\begin{aligned}
(\mathcal{D}_N\, \xi \circ \psi_{K,N})(\eta) &= \sum_{i=1}^{N} \sum_{k=1}^{K} \mu_{k_i,k} [\xi(k_1, \ldots, k_{i-1}, k, k_{i+1}, \ldots, k_N) - \xi(\psi_{K,N}(\eta))] \\
&= \sum_{k=1}^{K} \sum_{r=1}^{K} \sum_{i:\, k_i = r} \mu_{k_i,k} [\xi(k_1, \ldots, k_{i-1}, k, k_{i+1}, \ldots, k_N) - \xi(\psi_{K,N}(\eta))].
\end{aligned}$$

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,

$$\begin{aligned}(\mathcal{D}_N\,\xi\circ\psi_{K,N})(\eta) &= \sum_{k=1}^{K}\sum_{r=1}^{K}\sum_{i:\,k_i=r}\mu_{k_i,k}[\xi(\psi_{K,N}(\eta-\mathbf{e}_r+\mathbf{e}_k))-\xi(\psi_{K,N}(\eta))]\\ &= \sum_{k=1}^{K}\sum_{r=1}^{K}\eta(r)\mu_{r,k}[\xi(\psi_{K,N}(\eta-\mathbf{e}_r+\mathbf{e}_k))-\xi(\psi_{K,N}(\eta))]\\ &= (\mathcal{Q}_N\,\xi\circ\psi_{K,N})(\eta),\end{aligned}$$

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

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

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

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 $\eta\in\bigcup_{L=1}^{N}\mathcal{E}_{K-1,L}$.*

(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 $Q\,U_k=\lambda_k\,U_k$, $1\le k\le K-1$, 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

$$\mathcal{D}_N\,V_\eta=\lambda_\eta V_\eta, \tag{3.2}$$

where V_η is defined as in (1.8).

Applying $\psi_{K,N}$ to both members of (3.2) we obtain

$$(\mathcal{D}_N V_\eta)\circ\psi_{K,N}=\lambda_\eta V_\eta\circ\psi_{K,N}.$$

Using Lemma (3.1), 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 (2.9) is a basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$, due to Proposition 2.4-(c). Therefore, all the eigenvalues of $\mathcal{Q}_N$ are those described in part (b) and $\mathcal{Q}_N$ is diagonalisable. □

Remark 3.2. *Note that the results in Lemma 3.2 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 3.2 provides all the eigenvalues and right eigenvectors of $\mathcal{Q}_N$ when Q is diagonalisable. However, there exist ergodic non-diagonalisable rate matrices as the following example shows.

Example 2 (Non-diagonalisable rate matrix of an ergodic Markov chain)**.** *Consider the infinitesimal rate matrix Q given by*

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

The eigenvalues of Q are 0 and -14 as a double eigenvalue. However, the right eigenspace associated to -14 is the set of columns vectors proportional to $[1,-1,1]^T$, which has dimension 1. Therefore, Q is not diagonalisable. However, there exists an invertible matrix W such that $Q=WJW^{-1}$, where J is the Jordan form of Q. Indeed, if we define

$$W=\begin{pmatrix}3/14&2&11/14\\3/14&-2&-3/14\\3/14&2&-3/14\end{pmatrix},\quad W^{-1}=\begin{pmatrix}1&7/3&4/3\\0&-1/4&1/4\\1&0&-1\end{pmatrix},$$

then $Q = WJW^{-1}$ *holds, where*

$$J = W^{-1}QW = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -14 & 1 \\ 0 & 0 & -14 \end{pmatrix}.$$

Note that the unique stationary distribution of the process driven by Q *is proportional to the first row of* W^{-1}*, and is thus equal to* $(3/14, 1/2, 2/7)$.

Now our purpose is to extend the Lemma 3.2 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 [Ser10, Corollary 5.1], for instance, proves this result as a consequence of the Schur's Theorem. 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 the Schur's Theorem [Ser10, Theorem 5.1] 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 $\mathrm{D}(\mathrm{Spec}(M), \mathrm{Spec}(N)) < \epsilon$.

See e.g. [HM87] and [Ser10, Theorem 5.2] for a proof of Fact 2.

Proof of Theorem 1.1. From Lemma 3.2 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 $\mathrm{Spec}(Q) = \{0, \lambda_1, \dots, \lambda_{K-1}\}$, 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. $\mathrm{D}(\mathrm{Spec}(\mathcal{Q}_N), \sigma_N(Q)) = 0$.

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 3.2, Lemma 3.2 remains valid and it ensures us that $\mathrm{Spec}(\bar{\mathcal{Q}}_N) = \sigma_N(\bar{Q})$. Thus, using the triangular inequality we get

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

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 3.2. 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 3.2. □

Remark 3.3 (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* [Ser10, Proposition 3.20] *and* [CEZ11] *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 3.2.*

3.2. **Proof of Theorem 1.2.** 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

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

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

Remark 3.4 (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 1.2 we first formally describe the preserving degree polynomial property satisfied by $\mathcal{A}_N$.

Lemma 3.3 ($\mathcal{A}_N$ 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 3.3 is technical and it is deferred to Appendix B. We proceed to prove Theorem 1.2.

Proof of Theorem 1.2. (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

$$\mathcal{B}_0 := \{1\}, \;\; \mathcal{B}_1 := \{x_1, x_2, \ldots, x_{K-1}\}, \;\; \mathcal{B}_L := \{\mathbf{x}^\alpha, \; \alpha \in \mathcal{E}_{K-1,L}\}, \;\; 2 \leq L \leq N,$$

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 3.3-(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 $2 \le L \le N$. 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 $2 \le L \le N$.

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 $2 \le L \le N$. 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. ☐

Remark 3.5 (On the right eigenfunctions of $\mathcal{A}_N$)**.** *Theorem 1.2 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* [Wat61] *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* [Zho08, Section 2.4.4] *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, it is possible to describe the eigenspaces associated to the first three eigenvalues of $\mathcal{A}_N$. As we commented in Remark 3.4, 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 $\{x_k x_r,\ 1 \le k < r \le K\}$. 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 $\{x_k^2 x_r - x_k x_r^2,\ 1 \le k < r \le K\} \cup \{x_k x_r x_s,\ 1 \le k < r < s \le K\}$. The complete characterisation of the eigenvectors of $\mathcal{A}_N$, for $K \ge 3$, is a topic of further research.

3.3. **Proof of Theorem 1.3.** This section is devoted to the proof of Theorem 1.3 providing a description of the spectrum of the neutral multi-allelic Moran process with generator $\mathcal{Q}_{N,p}$, defined by (1.2) as

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

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

Assume $K \ge 2$, $N \ge 2$ 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 $\lambda_{\eta,p}$:

$$\lambda_{\eta,p} = -L(L-1)\frac{p}{N} + \sum_{k=1}^{K-1} \eta(k)\lambda_k. \tag{3.3}$$

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 1.3. It is straightforward to remark 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 3.2, 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 2.4-(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 $1 \le L \le N$ and $\eta \in \mathcal{E}_{K-1,L}$, the expression (1.9) and Lemma 2.3-(c) ensure that $\tilde{V}_\eta$ is a polynomial of total degree equal to L. Using now Theorem 1.2-(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 1.2-(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 $2 \le L \le N$. The first diagonal block of size K is a null matrix. It follows that the matrix

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

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 $\mathrm{Spec}(Q) = \{0, \lambda_1, \dots, \lambda_{K-1}\}$. 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 (3.3) 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,

$$\begin{aligned} \|\mathcal{Q}_{N,p} - \bar{\mathcal{Q}}_{N,p}\|_{\mathrm{Unif}} &= \|\mathcal{Q}_N - \bar{\mathcal{Q}}_N\|_{\mathrm{Unif}}, \\ \mathrm{D}\left(\mathrm{Spec}(\bar{\mathcal{Q}}_{N,p}), \sigma_{N,p}(Q)\right) &= \mathrm{D}\left(\mathrm{Spec}(\bar{\mathcal{Q}}_{K,N}), \sigma_N(Q)\right). \end{aligned}$$

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 3.6 (Alternative proof of Theorem 1.3)**.** *Another proof of Theorem 1.3 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.*

3.4. **Applications to the exponential ergodicity of neutral multi-allelic Moran processes.** We will now study some consequences of Theorem 1.3. Next result establishes that the Jordan form of Q is a diagonal block in the Jordan form of $\mathcal{Q}_{N,p}$. The book of Axler [Axl15] is a good reference for generalised eigenvectors and Jordan forms.

Corollary 3.4 (Jordan forms of Q and $\mathcal{Q}_{N,p}$)**.** *Consider $K \ge 2$, $N \ge 2$ and $p \ge 0$. 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, Q and $\mathcal{Q}_{N,p}$ have that same SLEM.

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 3.1 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 $\mathcal{U} = \{U_0, \dots, U_{K-1}\}$ 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 $\{\tilde{\xi}(U_0), \dots, \tilde{\xi}(U_{K-1})\}$ is a system of linearly independent generalised eigenvectors of $\mathcal{Q}_{N,p}$. 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 $\{\lambda_0, \lambda_1, \dots, \lambda_{K-1}\}$ are those eigenvalues of $\mathcal{Q}_{N,p}$ of smallest modulus. We thus get that Q and $\mathcal{Q}_{N,p}$ have the same SLEM. □

The total variation distance to stationarity at time t of the process driven by Q, with initial distribution μ, is given by $\mathrm{d}^{\mathrm{TV}}(\mu\,\mathrm{e}^{tQ}, \pi)$, where d^{TV} denotes the total variance distance defined by (1.14), μ is the initial distribution on $[K]$ and π is the stationary distribution of the process driven by Q. 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 Q, denoted $\mathrm{D}_Q^{\mathrm{TV}}$, as follows:

$$\mathrm{D}_Q^{\mathrm{TV}}(t) := \max_{\mu} \mathrm{d}^{\mathrm{TV}}(\mu\,\mathrm{e}^{tQ}, \pi),$$

where the maximum runs over all possible initial distributions on $[K]$. Using the convexity of d^{TV}, we can prove that $\mathrm{D}_Q^{\mathrm{TV}}(t) = \frac{1}{2}\|\mathrm{e}^{tQ} - \Pi\|_\infty$, where Π stands for the K-dimensional matrix with every row equal to π, and $\|\cdot\|_\infty$ denotes the infinity norm of matrices (cf. [LP17, Chapter 4]).

Every irreducible finite Markov chain convergences exponentially to stationarity, see e.g. [LP17, Theorem 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.

For a real positive function f we denote by $\mathcal{O}(f)$ another real positive function such that

$$C_1 f(t) \le \mathcal{O}(f)(t) \le C_2 f(t),$$

for two constants $0 < C_1 \le C_2 < \infty$ and for all $t \ge T$, for $T > 0$ large enough.

Let ρ be the SLEM of Q and s the largest multiplicity in the minimal polynomial of Q of all the eigenvalues with modulus ρ, or equivalently, the size of the largest Jordan block associated to eigenvalues with modulus ρ. Then,

$$\mathrm{D}_Q^{\mathrm{TV}}(t) = \mathcal{O}(t^{s-1}\mathrm{e}^{-\rho t}), \tag{3.4}$$

see e.g. [SRW15, Theorem 3.2]. The following result is a consequence of Corollary 3.4 and (3.4).

Corollary 3.5 (Asymptotic exponential ergodicity in total variation). *Let us denote by ρ the* SLEM *of Q and by $s \in \mathbb{N}$ the largest multiplicity in the minimal polynomial of Q of all the eigenvalues with modulus ρ. Then,*

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

The following two examples use Corollary 3.5 to provide the rates for the exponential convergence to stationarity of the neutral multi-allelic Moran (Fleming – Viot particle) process considered in [Cor20] and for the non-diagonalisable mutation rate matrix previously considered in Example 2.

Example 3 (Circulant mutation rate matrix). *Consider the following mutation rate matrix*

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

where $\theta \ge 0$. 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 [Cor20]. *Since Q_θ is circulant, it is possible to explicitly diagonalise it using the Fourier matrix (cf.* [Dav79]*). The eigenvalues of Q_θ are*

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

for $0 \le k \le K-1$ (see e.g. [Cor20]*). Thus, the* SLEM *of Q_θ is $2(1+\theta)\sin^2\left(\frac{\pi}{K}\right)$, 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 Q_θ is $\lambda_1 = 4\sin^2\left(\frac{\pi}{K}\right)$ 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,*

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

Example 4 (Convergence rate for a process with non-diagonalisable mutation rate matrix). *Consider Q as in Example 2 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, λ_L 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 3.5 we get $\rho = 14$ and $s = 2$. Then,

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

Note that for this simple example it is possible to explicitly compute $\mathrm{D}_Q^{\mathrm{TV}}(t)$*. Indeed, denoting* J *and* W *as in Example 2 we get*

$$\mathrm{e}^{tJ} = \exp\left\{ t \begin{pmatrix} 0 & 0 & 0 \\ 0 & -14 & 1 \\ 0 & 0 & -14 \end{pmatrix} \right\} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \mathrm{e}^{-14t} & t\,\mathrm{e}^{-14t} \\ 0 & 0 & \mathrm{e}^{-14t} \end{pmatrix},$$

for all $t \geq 0$*. Then,*

$$\mathrm{D}_Q^{\mathrm{TV}}(t) = \frac{1}{2} \left\| W \begin{pmatrix} 0 & 0 & 0 \\ 0 & \mathrm{e}^{-14t} & t\,\mathrm{e}^{-14t} \\ 0 & 0 & \mathrm{e}^{-14t} \end{pmatrix} W^{-1} \right\|_\infty = \left(2t + \frac{11}{14}\right) \mathrm{e}^{-14t}, \quad t \geq 0.$$

4. Spectral decomposition of the neutral multi-allelic Moran type process with parent independent mutation

In this section we discuss some applications of the Theorem 1.3 and its consequences to the neutral multi-allelic Moran model with parent independent mutation scheme. We start recalling some well-known results on finite state reversible Markov chains and their convergence to stationarity (cf. [SC97], [Bré20] and [LP17]). Later, we will focus on the case where the Moran process has *parent independent mutation* [Eth11]. In this case, the Moran process is reversible. In fact, as we claimed in Theorem 1.5, 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 1.6 and 1.7.

4.1. **Background on reversible finite continuous-time Markov chains.** Let us consider $\mathcal{G}$, the infinitesimal generator of an irreducible reversible continuous-time Markov chain on a finite space Ω, such that $\mathrm{Card}(\Omega) = M < \infty$. Let us denote by $\pi = (\pi_\omega)_{\omega \in \Omega}$ the stationary distribution on Ω of the process driven by $\mathcal{G}$. For every measure m on Ω, we consider $l^2(\mathbb{R}^\Omega, m)$ the Hilbert space with inner product

$$\langle \cdot, \cdot \rangle_m : (f, g) \mapsto \sum_{\omega \in \Omega} f(\omega) g(\omega) m(\omega),$$

induced by m, and we denote by $\| \cdot \|_m$ its induced norm. In the sequel we are mainly interested in the Hilbert spaces $l^2(\mathbb{R}^\Omega, \pi)$ and $l^2(\mathbb{R}^\Omega, \frac{1}{\pi})$, where $1/\pi$ denotes the measure $k \mapsto 1/\pi_k$. We also denote by $\langle \cdot, \cdot \rangle$ the usual inner product in $\mathbb{R}^\Omega$ and by $\| \cdot \|$ its associated norm.

Reversibility is equivalent to the fact that $\mathcal{G}$ is self-adjoint in $l^2(\mathbb{R}^\Omega, \pi)$, i.e. $\langle \mathcal{G} f, g \rangle_\pi = \langle f, \mathcal{G} g \rangle_\pi$, for every functions f and g in $\mathbb{R}^\Omega$. Thus, using the spectral decomposition for self-adjoint operators, we obtain an orthonormal basis of $l^2(\mathbb{R}^\Omega, \pi)$, denoted $\{f_k\}_{k=1}^M$ such that $\mathcal{G} f_k = \lambda_k f_k$, for $k \in [M]$. We assume that the M eigenvalues of $\mathcal{G}$ are ordered such that $0 = \lambda_1 < \lambda_2 \leq \cdots \leq \lambda_M$ and $f_1 \equiv 1$.

Remark 4.1 (Left and right eigenfunctions of reversible processes)**.** *Let* $\mathcal{G}$ *be the infinitesimal generator of an irreducible Markov chain which is reversible for* π*. Let* f *be a right eigenfunction of* $\mathcal{G}$ *associated to the eigenvalue* λ*, then* $\phi = \pi f$ *is a left eigenfunction of* $\mathcal{G}$*, also with eigenvalue* λ*. Indeed, for every function* h *we get*

$$\phi\, \mathcal{G}(h) \;=\; \langle \phi, \mathcal{G} h \rangle = \langle \pi f, \mathcal{G} h \rangle = \langle f, \mathcal{G} h \rangle_\pi = \langle \mathcal{G} f, h \rangle_\pi = \lambda \langle f, h \rangle_\pi = \lambda \phi(h).$$

Thus, $\phi\, \mathcal{G} = \lambda \phi$*. Analogously, if* ϕ *is a left eigenvector for* $\mathcal{G}$*, then* $f = \phi/\pi$ *is a right eigenvector for* $\mathcal{G}$ *with the same eigenvalue.*

Then, $\{\phi_k\}_{k=1}^M$, where $\phi_k = \pi f_k$, is a set of left eigenfunctions of $\mathcal{G}$ forming an orthonormal basis of $l^2(\mathbb{R}^\Omega, \frac{1}{\pi})$, and the sets $\{\phi_k\}_{k=1}^M$ and $\{f_k\}_{k=1}^M$ are biorthogonal with respect to the usual inner product, i.e. $\langle f_k, \phi_n \rangle = \delta_{k,n}$, for all $k, n \in [M]$, where $\delta_{k,n}$ stands for the Kronecker delta function.

For every ϕ in $\mathbb{R}^\Omega$ we have

$$\phi = \sum_{k=1}^M \langle \phi, \phi_k \rangle_{\frac{1}{\pi}} \phi_k,$$

and consequently

$$\phi\, \mathcal{G}^n = \sum_{k=1}^M \langle \phi, \phi_k \rangle_{\frac{1}{\pi}}\, \phi_k\, \mathcal{G}^n = \sum_{k=1}^M \lambda_k^n \langle \phi, \phi_k \rangle_{\frac{1}{\pi}} \phi_k,$$

for all $n \geq 1$. Thus,

$$\phi\, \mathrm{e}^{t\mathcal{G}} \quad = \quad |\phi|\, \pi + \sum_{k=2}^{M} \mathrm{e}^{t\lambda_k} \langle \phi, \phi_k \rangle_{\frac{1}{\pi}} \phi_k,$$

where the last inequality comes from $\lambda_1 = 0$, $\phi_1 = \pi$ and $|\phi| = \sum_{k=1}^{M} \phi(k)$. In particular, one can explicitly obtain the transition function for the process driven by $\mathcal{G}$ with initial distribution concentrated at $\omega \in \Omega$:

$$\delta_\omega \mathrm{e}^{tQ} - \pi = \sum_{k=2}^{M} \mathrm{e}^{\lambda_k t} \langle \delta_\omega, \phi_k \rangle_{\frac{1}{\pi}} \phi_k = \sum_{k=2}^{M} \mathrm{e}^{\lambda_k t} \frac{\phi_k(\omega)}{\pi(\omega)} \phi_k = \sum_{k=2}^{M} \mathrm{e}^{t\lambda_k} f_k(\omega)\, \phi_k, \tag{4.1}$$

The chi-square distance, as defined in (1.15), to stationarity can be characterised using the distance induced by the inner product in $l^2(\mathbb{R}^\Omega, \frac{1}{\pi})$ as $\chi^2_\omega(t) := \chi^2(\delta_\omega\, \mathrm{e}^{t\mathcal{G}} \mid \pi) = \|\delta_\omega\, \mathrm{e}^{t\mathcal{G}} - \pi\|^2_{\frac{1}{\pi}}$, for every $\omega \in \Omega$. Then, using (4.1) and the orthogonality of $\{\phi_k\}_{k=1}^M$, we deduce

$$\chi^2_\omega(t) = \sum_{k=2}^{M} \mathrm{e}^{2\lambda_k t} f_k(\omega)^2. \tag{4.2}$$

Let $\rho = -\lambda_2 > 0$ be the SLEM of $\mathcal{G}$. Note that for reversible Markov chain this value is equal to the *spectral gap* of $\mathcal{G}$. Using (4.2) we get the following upper bound for the convergence to stationarity of the Markov process driven by $\mathcal{G}$ when $t \to \infty$ in the chi-square distance:

$$\chi^2_\omega(t) \leq \|\delta_\omega - \pi\|^2_{\frac{1}{\pi}} \mathrm{e}^{-2\rho t} = \frac{1 - \pi(\omega)}{\pi(\omega)} \mathrm{e}^{-2\rho t}, \tag{4.3}$$

for every $\omega \in \Omega$.

Hölder's inequality implies the following identity:

$$\|\mu\|_{\frac{1}{\pi}} = \max_{\|f\|_{\frac{1}{\pi}} \leq 1} |\langle \mu, f \rangle|\,,$$

where $\langle \mu, f \rangle := \sum_{\omega \in \Omega} f(\omega)\mu(\omega)$, see e.g. Proposition 6.13 in [Fol99] for a proof. Thus, taking the right eigenfunction f_2 associated to $\lambda_2 = -\rho$ as a test function, we get

$$\chi^2_\omega(t) \geq \frac{|\langle \delta_\omega\, \mathrm{e}^{tQ}, f_2 \rangle - \langle \pi, f_2 \rangle|}{\|f_2\|_{\frac{1}{\pi}}} = \frac{|f_2(\omega)|}{\|f_2\|_{\frac{1}{\pi}}} \mathrm{e}^{-\rho t}. \tag{4.4}$$

Note that (4.3) and (4.4) imply that the exponential rate of convergence to stationarity in the chi-square distance is $-\rho$. The function f_2 has norm one in $l^2(\mathbb{R}^\Omega, \pi)$ but in general the value of its norm in $l^2(\mathbb{R}^\Omega, 1/\pi)$ is unknown. So, it is usually difficult to get an explicit expression for the lower bound (4.4) even when we know the complete spectral decomposition of the generator $\mathcal{G}$. A rougher lower bound can be deduced as follows. Note that

$$\|f\|^2_{\frac{1}{\pi}} = \sum_{\omega \in \Omega} \frac{f^2(\omega)}{\pi_\omega} \leq \frac{1}{\pi_\star^2} \|f\|^2_\pi,$$

for every real function f on Ω, where $\pi_\star = \min_{\omega \in \Omega} \pi_\omega$. Thus, we get $\chi^2_\omega(t) \geq \pi_\star\, |f_2(\omega)| \mathrm{e}^{-\rho t}$. On the other hand, the *Cauchy-Schwartz inequality* in $l^2(\mathbb{R}^\Omega, \frac{1}{\pi})$ implies

$$\mathrm{d}^{\mathrm{TV}}\left(\delta_\omega\, \mathrm{e}^{t\mathcal{G}}, \pi\right)^2 \leq \frac{1}{4} \chi^2_\omega(t) \leq \frac{1 - \pi(\omega)}{4\, \pi(\omega)} \mathrm{e}^{-2\rho t},$$

for every $\omega \in \Omega$, where d^{TV} is the total variation distance defined by (1.14). Additionally, the usual 1-norm in $\mathbb{R}^M$, denoted $\|\cdot\|_1$, allows the next characterisation:

$$\|\mu\|_1 = \max_{\|f\|_\infty = 1} |\langle \mu, f \rangle|\,.$$

Thus, using f_2 as a test function, we get the following lower bound for the convergence in total variation

$$\mathrm{d}^{\mathrm{TV}}\left(\delta_\omega \mathrm{e}^{t\mathcal{G}}, \pi\right) \geq \frac{|f_2(\omega)|}{\|f_2\|_\infty} \mathrm{e}^{-\rho t}. \tag{4.5}$$

Unlike (4.4), it should be not so hard in general to compute an explicit expression for $\|f_2\|_\infty$ and thus the lower bound (4.5). Therefore, we have obtained the following inequalities:

$$\frac{|f_2(\omega)|}{\|f_2\|_\infty} \mathrm{e}^{-\rho t} \leq \mathrm{d}^{\mathrm{TV}}\left(\delta_\omega \mathrm{e}^{t\mathcal{G}}, \pi\right) \leq \sqrt{\chi^2_\omega(t)} \leq \sqrt{\frac{1 - \pi(\omega)}{\pi(\omega)}} \mathrm{e}^{-\rho t}, \tag{4.6}$$

for every $\omega \in \Omega$ and any right eigenfunction f_2 of $\mathcal{G}$ with eigenvalue $-\rho$. Inequalities in (4.6) control the exponential decay in total variation and chi-square distances to stationarity of a reversible Markov chain.

4.2. **Proof of Theorems 1.6 and 1.7.** Let us recall that the generator of the neutral multi-allelic Moran process with parent independent mutation defined by (4.2), 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 $\eta \in \mathcal{E}_{K,N}$. We next prove Corollary 1.4, which provides the spectrum of $\mathcal{L}_{N,p}$, for all $p \geq 0$.

Proof of Corollary 1.4. Since $Q_{\boldsymbol{\mu}}$ is an infinitesimal matrix, zero is one of its eigenvalues with right eigenfunction f_1, the K-dimensional all-one vector. Note that $\boldsymbol{\pi} := \boldsymbol{\mu}/|\boldsymbol{\mu}| = (\mu_1/|\boldsymbol{\mu}|, \ldots, \mu_K/|\boldsymbol{\mu}|)$ is the unique stationary distribution of $Q_{\boldsymbol{\mu}}$, which is also reversible. Moreover, note that $Q_{\boldsymbol{\mu}}f = -|\boldsymbol{\mu}|(f + \langle f, \boldsymbol{\pi}\rangle)$, for every $f \in \mathbb{R}^{\mathcal{E}_{K,N}}$. Thus, every function f satisfying $\langle f, \boldsymbol{\pi}\rangle = 0$ is a right eigenfunction of $Q_{\boldsymbol{\mu}}$, i.e. the eigenspace associated to $-|\boldsymbol{\mu}|$ is the space of the orthogonal functions to $\boldsymbol{\pi}$, which has dimension $K-1$. The expression (1.12) for the eigenvalues comes from Theorem 1.3. Since the process is reversible we obtain that $\mathcal{L}_{N,p}$ is diagonalisable. The spectral gap is obtained for $L = 1$. □

Remark 4.2 (First degree eigenvectors of $\mathcal{L}_{N,p}$)**.** *Every function f in $\mathbb{R}^{[K]}$ satisfying $\langle f, \boldsymbol{\pi}\rangle = 0$ is a right eigenfunction of $Q_{\boldsymbol{\mu}}$. Thus, Remark 3.5 and Lemma 2.3-(a) imply that the linear function:*

$$\tilde{\xi}(f) : \eta \mapsto f(1)\eta(1) + \cdots + f(K)\eta(K)$$

is a right eigenvector of $\mathcal{L}_{N,p}$ associated to the eigenvalue $-|\boldsymbol{\mu}|$, for every f satisfying $\langle f, \boldsymbol{\pi}\rangle = 0$.

Using the classic results on reversible Markov chains reviewed in Section 4.1, we can ensure that the convergence to stationarity of the process driven by $\mathcal{L}_{N,p}$ is exponential with rate $-|\boldsymbol{\mu}|$. Next, we prove Theorem 1.6 providing explicit non-asymptotic bounds for the convergence for the process driven by $\mathcal{L}_{N,p}$ to stationarity in the chi-square and total variation distances.

Proof of Theorem 1.6. To prove the first inequality in (1.16) just note that, as we commented in Remark 4.2, for any function f in $\mathbb{R}^{[K]}$ such that $\langle f, \boldsymbol{\mu}\rangle = 0$, the linear function $\tilde{\xi}(f)$ is a right eigenfunction of $\mathcal{L}_{N,p}$ with eigenvalue $-|\boldsymbol{\mu}|$. Moreover, $\|\tilde{\xi}(f)\|_\infty = N\|f\|_\infty$, where, abusing notation, the infinity norms are defined on $\mathbb{R}^{\mathcal{E}_{K,N}}$ and $\mathbb{R}^{[K]}$, respectively. Then, using $\tilde{\xi}(f)$ as a test function in (4.5) we get

$$\mathrm{d}^{\mathrm{TV}}_\eta(t) \geq \frac{|\tilde{\xi}(f)(\eta)|}{2N\|f\|_\infty}\mathrm{e}^{-|\boldsymbol{\mu}|t}.$$

The optimal bound is obtained taking the maximum over all the functions orthogonal to $\boldsymbol{\mu}$. The remaining inequalities in (1.16) are a straightforward consequence of the classical bounds for chi-square and total variation distances on finite reversible Markov chains, summarised in (4.6).

To prove the first inequality in (1.17), note that

$$\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t) \geq \frac{|\phi(N\mathbf{e}_k)|}{2N\|\phi\|_\infty}\mathrm{e}^{-|\boldsymbol{\mu}|t},$$

where $\phi : \mathcal{E}_{K,N} \to \mathbb{R}$ is any right eigenfunction of $\mathcal{L}_{N,p}$ associated to $-|\boldsymbol{\mu}|$. Then, from Remark 4.2 the linear function $\phi : \eta \mapsto \eta(k)/\mu_k - \eta(r)/\mu_r$, where $r \neq k$, is a right eigenfunction of $\mathcal{L}_{N,p}$ associated to $-|\boldsymbol{\mu}|$. Then, $\|Q_2\|_\infty = N/(\mu_k \wedge \mu_r)$ and

$$\mathrm{d}^{\mathrm{TV}}_{N\mathbf{e}_k}(t) \geq \frac{\mu_k \wedge \mu_r}{\mu_k}\mathrm{e}^{-|\boldsymbol{\mu}|t}.$$

It is straightforward to check that the optimal constant is κ_k, as defined in Theorem 1.6.

To prove the last inequality in (1.17), simply note that the definition (1.13) of $\nu_{N,p}$ yields

$$\frac{1 - \nu_{N,p}(N\mathbf{e}_k)}{\nu_{N,p}(N\mathbf{e}_k)} = \begin{cases} \dfrac{(N|\boldsymbol{\mu}|/p)_{(N)} - (N\mu_k/p)_{(N)}}{(N\mu_k/p)_{(N)}} & \text{if} \quad p > 0 \\[2ex] \dfrac{|\boldsymbol{\mu}|^N - \mu_k^N}{\mu_k^N} & \text{if} \quad p = 0. \end{cases}$$

□

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 1.7. 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

$$H_\eta(\mathbf{x}; N, \boldsymbol{\alpha}) \quad := \quad \frac{1}{(N)_{[|\eta|]}} \prod_{k=1}^{K-1} (-N + |\mathbf{x}_{k-1}| + |\eta^{k+1}|)_{(\eta(k))} H_{\eta(k)}(x_k; M_k, \alpha_k, \gamma_k) \tag{4.7}$$

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

$$\begin{aligned} H_n(x; N, \beta, \gamma) \quad &:= \quad {}_3F_2\left(\begin{matrix} -n, & n+\beta+\gamma-1, -x \\ \beta, & -N \end{matrix} \middle| 1 \right) \\ &= \quad \sum_{j=0}^{n} \frac{(-n)_{(j)}(n+\beta+\gamma-1)_{(j)}(-x)_{(j)}}{\beta_{(j)}(-N)_{(j)}} \frac{1}{j!}. \end{aligned} \tag{4.8}$$

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 no 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:

$$K_\eta(\mathbf{x}; N, \mathbf{q}) \quad := \quad \frac{1}{(N)_{[|\eta|]}} \prod_{k=1}^{K-1} (-N + |\mathbf{x}_{k-1}| + |\eta^{k+1}|)_{(\eta(k))} K_{\eta(k)}(x_k; M_k, q_k) \tag{4.9}$$

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

$$\begin{aligned} K_n(x; N, q) \quad &:= \quad {}_2F_1\left(\begin{matrix} -n, & -x \\ & -N \end{matrix} \middle| \frac{1}{q} \right) \\ &= \quad \sum_{j=0}^{n} \frac{(-n)_{(j)}(-x)_{(j)}}{(-N)_{(j)}} \frac{1}{j!q^j}. \end{aligned} \tag{4.10}$$

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 [Ism05, Chapter 6] and [KLS10, Chapter 9] for more details about the univariate Hahn and Krawtchouk polynomials. We define the univariate Hahn and Krawtchouk polynomials in (4.8) and (4.10), respectively, using the hypergeometric functions notation which could be very useful for algebraic manipulations (cf. [KLS10, Chapter 10]). For instance, consider $\boldsymbol{\alpha} = N\boldsymbol{\mu}/p$ in the definition of Hahn polynomials, then

$$\begin{aligned} \lim_{p\to 0^+} H_{\eta(k)}(x_k; M_k, \alpha_k, |\boldsymbol{\alpha}^{k+1}| + 2|\eta^{k+1}|) &= \lim_{p\to 0^+} H_{\eta(k)}\left(x_k; M_k, \frac{N\mu_k}{p}, \frac{N|\boldsymbol{\mu}^{k+1}|}{p} + 2|\eta^{k+1}|\right) \\ &= \lim_{p\to 0^+} {}_3F_2\left(\begin{matrix} -\eta(k), & \eta(k) + N\mu_k/p + N|\boldsymbol{\mu}^{k+1}|/p + 2|\eta^{k+1}| - 1, -x_k \\ N\mu_k/p, & -M_k \end{matrix} \middle| 1 \right) \\ &= {}_2F_1\left(\begin{matrix} -\eta(k), -x_k \\ -M_k \end{matrix} \middle| \frac{\mu_k + |\boldsymbol{\mu}^{k+1}|}{\mu_k} \right) = {}_2F_1\left(\begin{matrix} -\eta(k), -x_k \\ -M_k \end{matrix} \middle| \frac{|\boldsymbol{\mu}^{k}|}{\mu_k} \right) \\ &= K_{\eta(k)}\left(x_k; N, \frac{|\boldsymbol{\mu}^k|}{\mu_k}\right), \end{aligned}$$

for every $k \in [K]$, where the calculation of the limit in the third equation follows from Equation (1.4.5) in [KLS10] and the last inequality follows from the definition of univariate Krawtchouk polynomials in

(4.10). Now, using the previous limit and the definitions (4.7) and (4.9) 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

$$Q_\eta(\mathbf{x}; N, \boldsymbol{\mu}, p) := \begin{cases} H_\eta\left(\mathbf{x}; N, \frac{N\boldsymbol{\mu}}{p}\right) & \text{if} \quad p > 0 \\ K_\eta\left(\mathbf{x}; N, \frac{\boldsymbol{\mu}}{|\boldsymbol{\mu}|}\right) & \text{if} \quad p = 0, \end{cases} \tag{4.11}$$

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 4.1 (Orthogonality of the Hahn and Krawtchouk polynomials)**.** *The multivariate polynomials Q_η defined by (4.11) 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.*

$$\begin{aligned} \mathbb{E}_{\nu_{N,p}}\left[Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p) Q_{\eta'}(\cdot\,; N, \boldsymbol{\mu}, p)\right] &= \sum_{\xi\in\mathcal{E}_{K,N}} Q_\eta(\xi\,; N, \boldsymbol{\mu}, p) Q_{\eta'}(\xi\,; N, \boldsymbol{\mu}, p)\nu_{N,p}(\xi) \\ &= d^2_{\eta,p}\,\delta_{\eta,\eta'}, \end{aligned}$$

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

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

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

See Theorem 5.4 in [IX07] and Proposition 2.1, also Remark 2.2, in [KZ09] for the proofs of these results on multivariate Hahn polynomials. See Theorem 6.2 in [IX07] and Proposition 2.4 in [KZ09] 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 [DG14].

Kernel polynomials for Dirichlet multinomial and multinomial distributions. Consider ν a multivariate distribution on $\mathcal{E}_{K,N}$ and $\{Q^0_\eta\}$ 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

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

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 do below in Proposition 4.2.

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,

$$h_n(\mathbf{x}, \mathbf{y}; p) = (|\boldsymbol{\alpha}| + 2n - 1)\frac{(|\boldsymbol{\alpha}| + N)_{(n)}}{N_{[n]}} \sum_{m=0}^{n} (-1)^{n-m} \frac{(|\boldsymbol{\alpha}| + m)_{(n-1)}}{m!(n-m)!} \psi_m(\mathbf{x}, \mathbf{y}; p), \tag{4.12}$$

for all $n \in [N]_0$, where

$$\psi_m(\mathbf{x},\mathbf{y};p) = \sum_{|\mathbf{z}|=m} \binom{N}{\mathbf{z}} \frac{|\boldsymbol{\alpha}|_{(m)}}{\prod_{i=1}^K (\alpha_i)_{(z_i)}} \frac{\prod_{i=1}^K (\alpha_i + x_i)_{(z_i)} (\alpha_i + y_i)_{(z_i)}}{(|\boldsymbol{\alpha}| + N)_m (|\boldsymbol{\alpha}| + N)_m},$$

for all $p > 0$, where $\alpha = N\boldsymbol{\mu}/p$. This expression becomes much simpler for $\mathbf{x} = \mathbf{y} = N\mathbf{e}_k$:

$$h_n(N\mathbf{e}_k, N\mathbf{e}_k; p) = \binom{N}{L} \frac{(|\boldsymbol{\alpha}| + 2n - 1)(|\boldsymbol{\alpha}|)_{(n-1)} (|\boldsymbol{\alpha}| - \alpha_k)_{(n)}}{(|\boldsymbol{\alpha}| + N)_{(n)} (\alpha_k)_{(n)}}, \tag{4.13}$$

for all $p > 0$. For $p = 0$, $\nu_{N,0}$ follows a $\mathcal{M}(\cdot \mid N, \boldsymbol{\mu}/|\boldsymbol{\mu}|)$ distribution and the n-th kernel polynomial can be written as

$$h_n(\mathbf{x},\mathbf{y};0) = \sum_{m=0}^{n} \binom{N}{n} \binom{N-m}{n-m} (-1)^{n-m} \zeta_m(\mathbf{x},\mathbf{y};0), \tag{4.14}$$

for all $n \in [N]_0$, where

$$\zeta_m(\mathbf{x},\mathbf{y};0) = \sum_{|\mathbf{z}|=m} \binom{m}{\mathbf{z}} \frac{\prod_{i=1}^K (x_i)_{[z_i]} (y_i)_{[z_i]}}{N_{[m]} N_{[m]}} \left(\frac{\mu_k}{|\boldsymbol{\mu}|} \right)^{-z_i}.$$

In particular,

$$\begin{aligned} h_n(\mathbf{x}, N\mathbf{e}_k; 0) &= \sum_{m=0}^{n} \binom{N}{m} \binom{N-m}{n-m} (-1)^{n-m} \frac{(x_k)_{[m]}}{N_{[m]}} \left(\frac{\mu_k}{|\boldsymbol{\mu}|} \right)^{-m}, & (4.15)\\ h_n(N\mathbf{e}_k, N\mathbf{e}_k; 0) &= \binom{N}{n} \left(\frac{|\boldsymbol{\mu}|}{\mu_k} - 1 \right)^n. & (4.16) \end{aligned}$$

For more details on the kernel polynomials for the multinomial distribution see e.g. [KZ09, Proposition 2.8] and [DG14]. Also, for more details on the kernel polynomials for the Dirichlet multinomial distribution see e.g. [KZ09, Proposition 2.6] and [GS13].

The following proposition proves that the right eigenfunctions of $\mathcal{L}_{N,p}$ are given by multivariate orthogonal polynomials defined by (4.11).

Proposition 4.2 (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*

$$\left\{ Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p), \eta \in \bigcup_{L=0}^{N} \mathcal{E}_{K-1,L} \right\} \tag{4.17}$$

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:

$$(\mathrm{e}^{t\mathcal{L}_{N,p}} \delta_\xi)(\eta) = \nu_{N,p}(\xi) \left(1 + \sum_{L=1}^{N} \mathrm{e}^{\lambda_{L,p} t} h_L(\eta, \xi; p) \right), \tag{4.18}$$

where $h_L(\eta, \xi; p)$ is defined by (4.12) and (4.14) for $p > 0$ and $p = 0$, respectively.

Remark 4.3. *Griffiths and Spanò [GS13] give the expression (4.18) 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. However, to the best of our knowledge, we have not found a proof of this result in the literature. For the sake of completeness we next provide a proof, which is similar to the proofs of Propositions 4.7 and 4.10 in [KZ09].*

Proof of Proposition 4.2. We know from Proposition 4.1-(b) that the multivariate polynomials $Q_\eta(\cdot; N, \boldsymbol{\mu}, p)$ defined by (4.11) are orthogonal in $l^2(\mathbb{R}^{\mathcal{E}_{K,N}}, \nu_{N,p})$, for every $p \geq 0$. Let us prove that they are the right eigenfunctions of $\mathcal{L}_{N,p}$. Indeed, since the generator $\mathcal{L}_{N,p}$ is reversible for the stationary distribution $\nu_{N,p}$, we have

$$\langle \mathcal{L}_{N,p} P_\eta, P_{\eta'} \rangle_{\nu_{N,p}} = \langle P_\eta, \mathcal{L}_{N,p} P_{\eta'} \rangle_{\nu_{N,p}},$$

for every pair of polynomials P_η and $P_{\eta'}$ on $\mathcal{E}_{K,N}$. Consider $\eta \in \mathcal{E}_{K-1,L}$, with $L \in [N]_0$, and $Q_\eta(\mathbf{x}; N, \boldsymbol{\mu}, p)$ the multivariate polynomial defined by (4.11) indexed by η. From Lemma 3.2, Theorem 1.2 and Proposition 4.1-(a) we know that for every $\eta \in \mathcal{E}_{K-1,L}$, $L \in [N]_0$, we obtain

$$\mathcal{L}_{N,p} Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p) = \lambda_{L,p} Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p) + \sum_{n=0}^{L-1} \sum_{\xi \in \mathcal{E}_{K-1,n}} a_\xi \, Q_\xi(\cdot\,; N, \boldsymbol{\mu}, p),$$

where $\lambda_{L,p}$ is given by (1.12) and a_ξ are real constants. Using Proposition 4.1-(b) and the fact that $\mathcal{L}_{N,p}$ is self-adjoint in $l^2(\nu_{N,p})$ we obtain

$$a_\xi = \langle \mathcal{L}_{N,p} Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p), Q_\xi(\cdot\,; N, \boldsymbol{\mu}, p) \rangle_{\nu_{N,p}} = \langle Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p), \mathcal{L}_{N,p} Q_\xi(\cdot\,; N, \boldsymbol{\mu}, p) \rangle_{\nu_{N,p}} = 0.$$

Note that the last equality holds because $\mathcal{L}_{N,p} Q_\xi(\cdot\,; N, \boldsymbol{\mu}, p)$ is a polynomial with total degree strictly less that $|\xi| < L$, and thus orthogonal to $Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p)$ in $l^2(\nu_{N,p})$. So, every polynomial $Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p)$ satisfies $\mathcal{L}_{N,p} Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p) = -\lambda_{L,p} Q_\eta(\cdot\,; N, \boldsymbol{\mu}, p)$, for $\eta \in \mathcal{E}_{K-1,L}$, and $L \in [N]_0$, proving that they are right eigenfunctions of $\mathcal{L}_{N,p}$. Since the system (4.17) forms an orthogonal basis of $\mathbb{R}^{\mathcal{E}_{K,N}}$, it forms an eigenbasis of right eigenfunctions of $\mathbb{R}^{\mathcal{E}_{K,N}}$.

The expression for the left eigenfunctions of $\mathcal{L}_{N,p}$ simply comes from Remark 4.1. Finally, the decomposition (4.18) for the transition kernel holds due to (4.1). □

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 η and its stationary distribution at a given time t.

Corollary 4.3 (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 -square distance between the distribution of the reversible process driven by $\mathcal{L}_{N,p}$ at time t and its stationary distribution:*

$$\chi^2_\eta(t) = \sum_{L=1}^{N} \mathrm{e}^{2\lambda_{L,p} t} h_L(\eta, \eta; p), \tag{4.19}$$

where $\lambda_{L,p}$ is given by (1.12) and $h_L(\eta, \eta; p)$ is defined by (4.12) and (4.14), for $p > 0$ and $p = 0$, respectively. In particular, when $\eta = N\mathbf{e}_k$, for $k \in [K]$, the chi-square distance to stationarity reduces to

$$\chi^2_{N\mathbf{e}_k}(t) = \sum_{L=1}^{N} \mathrm{e}^{2\lambda_{L,p} t} \binom{N}{L} \frac{(|\boldsymbol{\alpha}| + 2L - 1)(|\boldsymbol{\alpha}|)_{(L-1)}(|\boldsymbol{\alpha}| - \alpha_k)_{(L)}}{(|\boldsymbol{\alpha}| + N)_{(L)}(\alpha_k)_{(L)}}. \tag{4.20}$$

Proof. Using Corollary 1.6 and Proposition 4.2 we obtain the following equality for the chi-square distance:

$$\chi^2_\eta(t) = \sum_{L=1}^{N} \mathrm{e}^{2\lambda_L t} \sum_{\xi \in \mathcal{E}_{K-1,L}} \frac{Q_\xi(\eta; N, \boldsymbol{\mu}, p)^2}{d^2_{\alpha,p}},$$

where $d^2_{\alpha,p}$ is given as in Proposition 4.1. Note that $\sum_{\xi \in \mathcal{E}_{K-1,L}} Q_\xi(\eta; N, \boldsymbol{\alpha})^2 / d^2_{\alpha,p}$ is the *kernel polynomial* associated to $\nu_{N,p}$. Thus, (4.19) and (4.20) simply comes from the expressions for the kernel polynomials associated to $\nu_{N,p}$ in (4.12) and (4.14), respectively. □

We now take advantage of the explicit expression in (4.20) 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 $N \to \infty$. Let us first establish a result that will be very useful during the proof of Theorem 1.7.

Lemma 4.4 (Lemma A.2 in [NT20])**.** *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*

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

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 of the very recent work of Nestoridi and Thomas [NT20].

Proof of Theorem 1.7. Let us first prove the existence of the chi-square cutoff. When $p = 0$, we simply use (4.16) to deduce that

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

Now, for $t_{N,c} = \dfrac{\ln N + c}{2|\boldsymbol{\mu}|}$ we obtain

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

Now, since $K_{k,0} = |\boldsymbol{\mu}|/\mu_k - 1$, we have proved the existence of the limit (1.18) for $p = 0$.

Now, for $p > 0$ let us focus on expression (4.20). 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

$$\begin{aligned}\phi_{L,k}(N) &:= \frac{|\boldsymbol{\alpha}|+2L-1}{|\boldsymbol{\alpha}|+L-1}\,\frac{\prod_{r=0}^{L-1}(|\boldsymbol{\alpha}|+r)(|\boldsymbol{\alpha}|-\alpha_k+r)}{\prod_{r=0}^{L-1}(|\boldsymbol{\alpha}|+N+r)(\alpha_k+r)} \\ &= \frac{N|\boldsymbol{\mu}|/p+2L-1}{N|\boldsymbol{\mu}|/p+L-1}\,\frac{\prod_{r=0}^{L-1}\left(\frac{N|\boldsymbol{\mu}|}{p}+r\right)\left(\frac{N(|\boldsymbol{\mu}|-\mu_k)}{p}+r\right)}{\prod_{r=0}^{L-1}\left(\frac{N|\boldsymbol{\mu}|}{p}+N+r\right)\left(\frac{N\mu_k}{p}+r\right)} \\ &= \frac{N|\boldsymbol{\mu}|/p+2L-1}{N|\boldsymbol{\mu}|/p+L-1}\left[\frac{|\boldsymbol{\mu}|(|\boldsymbol{\mu}|-\mu_k)}{\mu_k(|\boldsymbol{\mu}|+p)}\right]^L \frac{\prod_{r=0}^{L-1}\left(1+\frac{p}{N|\boldsymbol{\mu}|}r\right)\left(1+\frac{p}{N(|\boldsymbol{\mu}|-\mu_k)}r\right)}{\prod_{r=0}^{L-1}\left(1+\frac{p}{N(|\boldsymbol{\mu}|+p)}r\right)\left(1+\frac{p}{N\mu_k}r\right)}.\end{aligned}$$

Thus, 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,

$$\begin{aligned}\binom{N}{L} &= \frac{\prod_{r=0}^{L-1}(N-r)}{L!} = \frac{N^L}{L!}\prod_{r=0}^{L-1}\left(1-\frac{r}{N}\right) \underset{N}{\sim} \frac{N^L}{L!}, \text{ and} \\ \mathrm{e}^{2\lambda_L t_N} &= \exp\left\{-2\left(|\boldsymbol{\mu}|L+\frac{p}{N}L(L-1)\right)\left(\frac{\ln(N)+c}{2|\boldsymbol{\mu}|}\right)\right\} \\ &\underset{N}{\sim} \exp\left\{-(\ln(N)+c)L\right\} = \frac{(\mathrm{e}^{-c})^L}{N^L},\end{aligned}$$

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 (4.13) 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 (4.20) yields

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

Moreover,

$$\begin{aligned} e^{2\lambda_L t_N} h_L(N\mathbf{e}_k, N\mathbf{e}_k; p) &\leq e^{-L(c+\ln(N))} h_L(N\mathbf{e}_k, N\mathbf{e}_k; p) \\ &= \frac{e^{-cL}}{N^L}\binom{N}{L}\frac{|\boldsymbol{\alpha}|+2L-1}{\boldsymbol{\alpha}|+L-1}\frac{|\boldsymbol{\alpha}|_{(L)}(|\boldsymbol{\alpha}|-\alpha_k)_{(L)}}{(|\alpha|+N)_{(L)}(\alpha_k)_{(L)}} \\ &= \frac{e^{-cL}}{L!}\frac{|\boldsymbol{\alpha}|+2L-1}{\boldsymbol{\alpha}|+L-1}\prod_{r=0}^{L-1}\left[\frac{N-r}{N}\frac{|\boldsymbol{\alpha}|+r}{|\boldsymbol{\alpha}|+N+r}\frac{|\boldsymbol{\alpha}|-\alpha_k+r}{\alpha_k+r}\right] \\ &\leq 3\frac{(\gamma e^{-c})^L}{L!}, \end{aligned}$$

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

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

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

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

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

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

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

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

Let us now prove the existence of a total variation cutoff for the process driven by $\mathcal{L}_N$. Using (4.15) and (4.18) we get

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

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 1.8. Moreover,

$$\begin{aligned} d^{\mathrm{TV}}_{N\mathbf{e}_k}(t) &= \frac{1}{2}\sum_{\xi\in\mathcal{E}_{K,N}}\left|(e^{t\mathcal{L}_N}\delta_\xi)(N\mathbf{e}_k) - \nu_N(\xi)\right| \\ &= \frac{1}{2}\sum_{\xi\in\mathcal{E}_{K,N}}\nu_N(\xi)\left|(1-e^{-|\boldsymbol{\mu}|t})^{N-\xi_k}\left[1-e^{-|\boldsymbol{\mu}|t}+\frac{|\boldsymbol{\mu}|e^{-|\boldsymbol{\mu}|t}}{\mu_k}\right]^{\xi_k} - 1\right| \end{aligned}$$

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

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

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

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

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

Then, using Lemma 4.4 we obtain

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

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

□

5. Discussion and open problems

In this paper we have studied the spectral properties of the neutral multi-allelic Moran model with irreducible mutation rate matrix Q. We have explicitly described the spectrum of the neutral multi-allelic Moran process using the spectrum of Q. The fact that diagonalisable matrices are dense in the space of real matrices allowed us to extend our results about the spectrum of $\mathcal{Q}_{N,p}$, for $p \geq 0$, to processes with irreducible rate matrices non necessarily diagonalisable. Moreover, when Q is diagonalisable we have proved that the composition chain driven by $\mathcal{Q}_N$ is also diagonalisable and we have described the set of right eigenvalues using the symmetrised tensor products of the right eigenvalues of Q.

The neutral multi-allelic Moran process was proved to be reversible if and only if the mutation scheme is parent independent. In this case, we have studied the spectral properties of $\mathcal{L}_{N,p}$, which allowed us to provide non-asymptotic bounds for the speed of convergence to stationarity and to prove the existence of a strongly optimal cutoff in chi-square and total variation distances when $p \geq 0$ and $p = 0$, respectively.

There are lots of future directions to explore for the further study 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. [SH17]). 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 [Zho08, Section 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 [ZL09], 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. Two possible generalisations of Theorem 1.7 could be to prove the existence of a total variation cutoff phenomenon for the Moran process with general mutation rate matrix and studying the existence of cutoff for the Moran process with parent independent mutation, when initially all the individuals are not of the same type.

A further study of the exponential ergodicity of the Moran process driven by $\mathcal{Q}_{N,p}$ could be carried out by studying the *Dirichlet form* of the Moran process. The results in this direction of the recent paper of Hermon and Salez [HS19] for the zero range process could be applied to the mutation process driven by $\mathcal{Q}_N$, in order to bound its *Poincaré's constant*. It would be very interesting to obtain a result in the spirit of *Aldous' conjecture* to control the spectral of $\mathcal{Q}_{N,p}$ using the spectral gap of Q, i.e. something similar to Theorem 2 in [HS19], but for the neutral multi-allelic Moran process driven by $Q_{K,N,p}$.

Another interesting problem to address is the study of the spectrum of the multi-allelic Moran process with selection. Under selection at birth 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, i.e. with generator given by (1.19), seems more complicated from the spectral point of view because it is non-reversible. However, this process is very interesting in population genetics and also for its interpretation as Fleming – Viot particle system, which approximates the quasi-stationary distribution of a continuous-time Markov chain. We believe that the exact results on the neutral Moran process will offer clues on the study of the spectrum of the more complicated Moran processes with selection.

Appendix A. Proofs of Lemmas 2.1 and 2.3, and Proposition 2.4

This section is devoted to the proofs of Lemmas 2.1 and 2.3, and Proposition 2.4.

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

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

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

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

which finishes the proof of part (a).

(b) From (2.5), we get

$$\begin{aligned}
\xi(V_1, V_2, \dots, V_L) &= \sum_{s \in \mathcal{I}_{L-1,N}} V_1(k_{s(1)}) V_2(k_{s(2)}) \dots V_{L-1}(k_{s(L-1)}) \sum_{i \in [N] \setminus s([L-1])} V_L(k_i) \\
&= \sum_{s \in \mathcal{I}_{L-1,N}} V_1(k_{s(1)}) V_2(k_{s(2)}) \dots V_{L-1}(k_{s(L-1)}) \left(\sum_{i=1}^{N} V_L(k_i) - \sum_{i=1}^{L-1} V_L(k_{s(i)}) \right) \\
&= \xi(V_1, V_2, \dots, V_{L-1})(k_1, k_2, \dots, k_N) \xi(V_L)(k_1, k_2, \dots, k_N) \\
&\quad - \sum_{i=1}^{L-1} \xi(V_1, \dots, V_{i,L}, \dots, V_{L-1})(k_1, k_2, \dots, k_N).
\end{aligned}$$

Using (2.7) we obtain the result for $\tilde{\xi}(V_1, V_2, \dots, V_L)$. The particular case $L = 2$ comes from part (a).

(c) We can prove equation (2.8) by induction on L. For $L = 1$ the result easily comes by (a). If we suppose that (2.8) is satisfied for L, for $2 \leq L < N - 1$, then, using (b) and (a), we can check that (2.8) holds for $L + 1$. □

Proof of Proposition 2.4. 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. Lemma 12.12 in [LP17]). 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 $V_\eta \neq V_\alpha$, for $\eta \neq \alpha$. However, $\mathcal{S}^N$ is a generator system of $\mathrm{Sym}(\mathbb{R}^{[K]^N})$ satisfying

$$\begin{aligned}\mathrm{Card}\left(\mathcal{S}^N\right) &\leq 1 + \sum_{L=1}^{N} \mathrm{Card}(\mathcal{E}_{K-1,L}) \\ &= \sum_{L=0}^{N} \binom{K-2+L}{L} = \binom{K-1+N}{N},\end{aligned}$$

where the last equality is the well-known *Hockey–Stick identity* in combinatorics, see e.g. [LPV03]. 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 $\mathcal{S}^N$. □

Appendix B. Proof of Lemma 3.3

Proof of Lemma 3.3. Let us first prove that if m is a monomial on $\mathcal{E}_{K,N}$ of total degree $|\alpha| = L$ with $0 \leq L \leq N$. Then,

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

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

As we commented in Remark 3.4, the result is true for $L = 1$. Let us assume $L \geq 2$ 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

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

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 $\alpha_k \geq 2$ 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 $\alpha_r \geq 2$ 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

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

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) may be written

$$\sum_{k,r:k\neq r}\left(\prod_{s\notin\{k,r\}}\eta(s)^{\alpha_s}\right)\alpha_r\eta(k)^{\alpha_k+1}\eta(r)^{\alpha_r} = p(\eta)\sum_{k,r:k\neq r}(\alpha_r\eta(k) - \alpha_k\eta(r)) = 0,$$

by antisymmetry in k, r of the summand. The third term is

$$-\sum_{k,r:k\neq r}\left(\prod_{s\notin\{k,r\}}\eta(s)^{\alpha_s}\right)\alpha_k\alpha_r\eta(k)^{\alpha_k}\eta(r)^{\alpha_r}=-c_1p(\eta),$$

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

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

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

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

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

Now, a polynomial P of total degree L may be written, in a unique way as $P=P_L+R$, where P_L is a non-null homogeneous polynomial of degree L and R is a polynomial of total degree strictly less than L. Note that $P_L=\sum_{i\in I}m_i$, where I is finite and m_i is a monomial of degree L, for $i\in I$. Thus, (a) implies that $\mathcal{A}_NV_{m_i}=-L(L-1)V_{m_i}+V_{q_i}$, for all $i\in I$ where q_i has a total degree strictly less than L. Also, $\mathcal{A}_NV_R=V_{R'}$ where R' is a polynomial with total degree strictly less than L. Therefore,

$$\begin{aligned}\mathcal{A}_NV_p&=\sum_{i\in I}\mathcal{A}_NV_{m_i}+\mathcal{A}_NV_R\\&=-L(L-1)\sum_{i\in I}V_{m_i}+\sum_{i\in I}V_{q_i}+V_{R'}\\&=-L(L-1)V_P+V_{R''},\end{aligned}$$

where R'' has total degree strictly less than L.

□

Appendix C. Proof of Theorem 1.5

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 reproduction*.

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. [Kel79, Theorem 1.3]. We have

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

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,

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

Therefore, since the stationary probability is reversible, the *Kolmogorov cycle reversibility criterion* [Kel79, Theorem 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 $\boldsymbol{\mu}$. □

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 Theorem 1.5 can be proved, where the values of the parameter p in (4.2) also depend on j, i.e. a model with selection at reproduction or fecundity selection [MW09]. 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

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

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 $(X_1, X_2, \dots, X_K)$ follows a $\mathcal{MD}(\cdot \mid N, N\boldsymbol{\mu})$. Note that the measure defined by (C.2) with the normalisation constant (C.3) is a probability distribution. See [JKB97] and [NRdA06] 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.1. *This result is known for multi-allelic Moran models with parent independent mutation. See e.g.* [EG09, 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 Theorem 1.5.*

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,

$$\begin{aligned} Z\,\mathcal{WDM}(\eta \mid N, \boldsymbol{\mu}, \mathbf{p})\mathcal{L}_{N,\boldsymbol{p}}[\eta, \xi] &= \binom{N}{\eta}\left[\prod_{k=1}^{K} q_k^{\eta(k)} \left(\frac{\mu_k}{q_k}\right)_{(\eta(k))}\right] \eta(i)[\mu_j + \eta(j) q_j] \\ &= \frac{N!}{\prod_{k \notin \{i,j\}} \eta(k)!} \, \frac{1}{\eta(i)!\eta(j)!} \left[\prod_{k=1}^{K} \prod_{l=0}^{\eta(k)-1} (\mu_k + l\, q_k)\right] \eta(i)[\mu_j + \eta(j)\, q_j], \end{aligned} \tag{C.4}$$

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

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

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.

□

Acknowledgement

The author would like to thank his advisors, Djalil Chafaï and Simona Grusea, for their encouragement and many fruitful discussions on this research. The author would also like to extend his gratitude to Didier Pinchon for his valuable help that greatly improved the quality of this manuscript.

References

[AD86] D. Aldous and P. Diaconis, *Shuffling cards and stopping times*, Amer. Math. Monthly **93** (1986), no. 5, 333–348. MR 841111

[AFG11] A. Asselah, P. A. Ferrari, and P. Groisman, *Quasistationary distributions and Fleming – Viot processes in finite spaces*, J. Appl. Probab. **48** (2011), no. 2, 322–332. MR 2840302

[Ald83] D. Aldous, *Random walks on finite groups and rapidly mixing Markov chains*, Seminar on probability, XVII, Lecture Notes in Math., vol. 986, Springer, Berlin, 1983, pp. 243–297. MR 770418

[Axl15] S. Axler, *Linear algebra done right*, third ed., Undergraduate Texts in Mathematics, Springer, Cham, 2015. MR 3308468

[BHM00] K. Burdzy, R. Hołyst, and P. March, *A Fleming – Viot Particle Representation of the Dirichlet Laplacian*, Comm. Math. Phys. **214** (2000), no. 3, 679–703. MR 1800866

[BLY06] J. Barrera, B. Lachaud, and B. Ycart, *Cut-off for n-tuples of exponentially converging processes*, Stochastic Process. Appl. **116** (2006), no. 10, 1433–1446. MR 2260742

[Bré20] P. Brémaud, *Markov chains: Gibbs fields, Monte Carlo simulation, and Queues*, second ed., Texts in Applied Mathematics, vol. 31, Springer, 2020.

[CEZ11] D. Couty, J. Esterle, and R. Zarouf, *Décomposition effective de Jordan-Chevalley*, Gaz. Math. (2011), no. 129, 29–49. MR 2850401

[Che06] G.-Y. Chen, *The cutoff phenomenon for finite Markov chains*, Ph.D. thesis, Cornell University, 2006.

[CHS17] G.-Y. Chen, J.-M. Hsu, and Y.-C. Sheu, *The L^2-cutoffs for reversible Markov chains*, Ann. Appl. Probab. **27** (2017), no. 4, 2305–2341. MR 3693527

[CK18] G.-Y. Chen and T. Kumagai, *Cutoffs for product chains*, Stochastic Process. Appl. **128** (2018), no. 11, 3840–3879. MR 3860012

[Con10] S. B. Connor, *Separation and coupling cutoffs for tuples of independent Markov processes*, ALEA Lat. Am. J. Probab. Math. Stat. **7** (2010), 65–77. MR 2644042

[Cor20] J. Corujo, *Dynamics of a Fleming – Viot type particle system on the cycle graph*, arXiv e-prints (2020), arXiv:2001.08000.

[CSC08] G.-Y. Chen and L. Saloff-Coste, *The cutoff phenomenon for ergodic Markov processes*, Electron. J. Probab. **13** (2008), no. 3, 26–78. MR 2375599

[CSC10] ———, *The L^2-cutoff for reversible Markov processes*, J. Funct. Anal. **258** (2010), no. 7, 2246–2315. MR 2584746

[CT16a] B. Cloez and M.-N. Thai, *Fleming – Viot processes: two explicit examples*, ALEA Lat. Am. J. Probab. Math. Stat. **13** (2016), no. 1, 337–356. MR 3487076

[CT16b] ———, *Quantitative results for the Fleming – Viot particle system and quasi – stationary distributions in discrete space*, Stochastic Process. Appl. **126** (2016), no. 3, 680–702. MR 3452809

[Dav79] P. J. Davis, *Circulant matrices*, John Wiley & Sons, New York-Chichester-Brisbane, 1979, A Wiley-Interscience Publication, Pure and Applied Mathematics. MR 543191

[DG14] P. Diaconis and R. Griffiths, *An introduction to multivariate Krawtchouk polynomials and their applications*, J. Statist. Plann. Inference **154** (2014), 39–53. MR 3258404

[DG19] P. Diaconis and R. C. Griffiths, *Reproducing kernel orthogonal polynomials on the multinomial distribution*, J. Approx. Theory **242** (2019), 1–30. MR 3915331

[DMM00] P. Del Moral and L. Miclo, *A Moran particle system approximation of Feynman – Kac formulae*, Stochastic Process. Appl. **86** (2000), no. 2, 193–216. MR 1741805

[DR00] P. Donnelly and E. R. Rodrigues, *Convergence to stationarity in the Moran model*, J. Appl. Probab. **37** (2000), no. 3, 705–717. MR 1782447

[DS81] P. Diaconis and M. Shahshahani, *Generating a random permutation with random transpositions*, Z. Wahrsch. Verw. Gebiete **57** (1981), no. 2, 159–179. MR 626813

[Dur08] R. Durrett, *Probability models for DNA sequence evolution*, second ed., Probability and its Applications (New York), Springer, New York, 2008. MR 2439767

[EG09] A. M. Etheridge and R. C. Griffiths, *A coalescent dual process in a Moran model with genic selection*, Theoretical Population Biology **75** (2009), no. 4, 320–330.

[Eth11] A. Etheridge, *Some mathematical models from population genetics*, Lecture Notes in Mathematics, vol. 2012, Springer, Heidelberg, 2011, Lectures from the 39th Probability Summer School held in Saint-Flour, 2009, École d'Été de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School]. MR 2759587

[Fen10] S. Feng, *The Poisson-Dirichlet distribution and related topics*, Probability and its Applications (New York), Springer, Heidelberg, 2010, Models and asymptotic behaviors. MR 2663265

[FM07] P. Ferrari and N. Marić, *Quasi Stationary Distributions and Fleming – Viot processes in countable spaces*, Electron. J. Probab. **12** (2007), no. 24, 684–702. MR 2318407

[Fol99] G. B. Folland, *Real analysis*, second ed., Pure and Applied Mathematics (New York), John Wiley & Sons, Inc., New York, 1999, Modern techniques and their applications, A Wiley-Interscience Publication. MR 1681462

[FV79] W. H. Fleming and M. Viot, *Some measure – valued Markov processes in population genetics theory*, Indiana Univ. Math. J. **28** (1979), no. 5, 817–843. MR 542340

[GS13] R. C. Griffiths and D. Spanò, *Orthogonal polynomial kernels and canonical correlations for Dirichlet measures*, Bernoulli **19** (2013), no. 2, 548–598. MR 3037164

[HM87] G. Harris and C. Martin, *The roots of a polynomial vary continuously as a function of the coefficients*, Proc. Amer. Math. Soc. **100** (1987), no. 2, 390–392. MR 884486

[HS19] J. Hermon and J. Salez, *A version of Aldous' spectral-gap conjecture for the zero range process*, Ann. Appl. Probab. **29** (2019), no. 4, 2217–2229. MR 3984254

[Ism05] M. E. H. Ismail, *Classical and quantum orthogonal polynomials in one variable*, Encyclopedia of Mathematics and its Applications, vol. 98, Cambridge University Press, Cambridge, 2005, With two chapters by Walter Van Assche, With a foreword by Richard A. Askey. MR 2191786

[IX07] P. Iliev and Y. Xu, *Discrete orthogonal polynomials and difference equations of several variables*, Adv. Math. **212** (2007), no. 1, 1–36. MR 2319761

[JKB97] N. L. Johnson, S. Kotz, and N. Balakrishnan, *Discrete multivariate distributions*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons, Inc., New York, 1997, A Wiley-Interscience Publication. MR 1429617

[JKK05] N. L. Johnson, A. W. Kemp, and S. Kotz, *Univariate discrete distributions*, third ed., Wiley Series in Probability and Statistics, Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2005. MR 2163227

[Kel79] F. P. Kelly, *Reversibility and stochastic networks*, John Wiley & Sons, Ltd., Chichester, 1979, Wiley Series in Probability and Mathematical Statistics. MR 554920

[KLS10] R. Koekoek, P. A. Lesky, and R. F. Swarttouw, *Hypergeometric orthogonal polynomials and their q-analogues*, Springer Monographs in Mathematics, Springer-Verlag, Berlin, 2010, With a foreword by Tom H. Koornwinder. MR 2656096

[KM65] S. Karlin and J. McGregor, *Ehrenfest urn models*, J. Appl. Probability **2** (1965), 352–376. MR 184284

[KM75] ———, *Linear growth models with many types and multidimensional Hahn polynomials*, Theory and application of special functions (Proc. Advanced Sem., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1975), 1975, pp. 261–288. Math. Res. Center, Univ. Wisconsin, Publ. No. 35. MR 0406574

[KZ09] K. Khare and H. Zhou, *Rates of convergence of some multivariate Markov chains with polynomial eigenfunctions*, Ann. Appl. Probab. **19** (2009), no. 2, 737–777. MR 2521887

[Lac15] H. Lacoin, *A product chain without cutoff*, Electron. Commun. Probab. **20** (2015), no. 19, 9. MR 3320407

[Lan02] S. Lang, *Algebra*, third ed., Graduate Texts in Mathematics, vol. 211, Springer-Verlag, New York, 2002. MR 1878556

[LP17] D. A. Levin and Y. Peres, *Markov chains and mixing times*, American Mathematical Society, Providence, RI, 2017, Second edition of [MR2466937], With contributions by E. L. Wilmer, With a chapter on "Coupling from the past" by J. G. Propp and D. B. Wilson. MR 3726904

[LPV03] L. Lovász, J. Pelikán, and K. Vesztergombi, *Discrete mathematics*, Undergraduate Texts in Mathematics, Springer-Verlag, New York, 2003, Elementary and beyond. MR 1952453

[Möh18] M. Möhle, *A spectral decomposition for the block counting process and the fixation line of the* $\mathrm{beta}(3,1)$-*coalescent*, Electron. Commun. Probab. **23** (2018), Paper No. 102, 15. MR 3896840

[Möh19] ———, *A spectral decomposition for a simple mutation model*, Electron. Commun. Probab. **24** (2019), Paper No. 15, 14. MR 3933039

[Mor58] P. A. P. Moran, *Random processes in genetics*, Proc. Cambridge Philos. Soc. **54** (1958), 60–71. MR 127989

[Mos62] J. E. Mosimann, *On the compound multinomial distribution, the multivariate β-distribution, and correlations among proportions*, Biometrika **49** (1962), 65–82. MR 143299

[MP14] M. Möhle and H. Pitters, *A spectral decomposition for the block counting process of the Bolthausen-Sznitman coalescent*, Electron. Commun. Probab. **19** (2014), no. 47, 11. MR 3246966

[MV12] S. Méléard and D. Villemonais, *Quasi-stationary distributions and population processes*, Probab. Surv. **9** (2012), 340–410. MR 2994898

[MW09] C. A. Muirhead and J. Wakeley, *Modeling multiallelic selection using a moran model*, Genetics **182** (2009), no. 4, 1141–1157.

[NN14] F. Nielsen and R. Nock, *On the chi square and higher-order chi distances for approximating f-divergences*, IEEE Signal Processing Letters **21** (2014), no. 1, 10–13.

[NRdA06] J. Navarro, J. M. Ruiz, and Y. del Aguila, *Multivariate weighted distributions: a review and some extensions*, Statistics **40** (2006), no. 1, 51–64. MR 2207404

[NT20] E. Nestoridi and S. Thomas, *Limit Profiles for Markov Chains*, arXiv e-prints (2020), arXiv:2005.13437.

[Pea65] Marshall C. Pease, III., *Methods of matrix algebra*, Mathematics in Science and Engineering. Vol. 16, Academic Press, New York-London, 1965. MR 0207719

[SC97] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413. MR 1490046

[Ser10] D. Serre, *Matrices*, second ed., Graduate Texts in Mathematics, vol. 216, Springer, New York, 2010, Theory and applications. MR 2744852

[SH17] D. Schrempf and A. Hobolth, *An alternative derivation of the stationary distribution of the multivariate neutral wrightfisher model for low mutation rates with a view to mutation rate estimation from site frequency data*, Theoretical Population Biology **114** (2017), 88 – 94.

[SRW15] O. Szehr, D. Reeb, and M. M. Wolf, *Spectral convergence bounds for classical and quantum Markov processes*, Comm. Math. Phys. **333** (2015), no. 2, 565–595. MR 3296158

[Vil20] D. Villemonais, *Lower Bound for the Coarse Ricci Curvature of Continuous-Time Pure-Jump Processes*, J. Theoret. Probab. **33** (2020), no. 2, 954–991. MR 4091580

[Wat61] G. A. Watterson, *Markov chains with absorbing states: A genetic example*, Ann. Math. Statist. **32** (1961), 716–729. MR 125633

[Yca99] B. Ycart, *Cutoff for samples of Markov chains*, ESAIM Probab. Statist. **3** (1999), 89–106. MR 1716128

[Zho08] H. Zhou, *Examples of Multivariate Markov Chains with Orthogonal Polynomial Eigenfunctions*, PhD Thesis, Stanford University, 2008.

[ZL09] H. Zhou and K. Lange, *Composition Markov chains of multinomial type*, Adv. in Appl. Probab. **41** (2009), no. 1, 270–291. MR 2514954

CEREMADE, Université Paris-Dauphine, PSL University and Institut National des Sciences Appliques de Toulouse, Institut de Mathématiques de Toulouse, France
E-mail address: jcorujo@insa-toulouse.fr