Uniform in time propagation of chaos for a Moran model

Bertrand Cloez\textsuperscript{a}, Josué Corujo\textsuperscript{b,c,d,∗}

\textsuperscript{a}MISTEA, Univ Montpellier, INRAE, Institut Agro, Montpellier, 34060, France
\textsuperscript{b}IRMA, Université de Strasbourg, Strasbourg, 67084, France
\textsuperscript{c}CEREMADE, Université Paris-Dauphine, Université PSL, CNRS, Paris, 75016, France
\textsuperscript{d}Institut de Mathématiques de Toulouse, Université de Toulouse, Institut National des Sciences Appliquées, Toulouse, 31077, France

Abstract
This article studies the limit of the empirical distribution induced by a mutation-selection multi-allelic Moran model. Our results include a uniform in time bound for the propagation of chaos in \(L^p\) of order \(\sqrt{N}\), and the proof of the asymptotic normality with zero mean and explicit variance, when the number of individuals tend towards infinity, for the approximation error between the empirical distribution and its limit. Additionally, we explore the interpretation of this Moran model as a particle process whose empirical probability measure approximates a quasi-stationary distribution, in the same spirit as the Fleming – Viot particle systems.

Keywords:
multi-allelic Moran model, Feynman – Kac formulae, propagation of chaos, quasi-stationary distribution, Fleming – Viot particle system, asymptotic normality

2020 MSC: 60J28, 60F25 (Primary), 92D10, 62L20 (Secondary)

1. Introduction and main results

This paper is devoted to the study of a mutation-selection multi-allelic Moran model consisting on \(N\) individuals, which can be of different allelic types belonging to a countable set \(E\), equipped with the discrete metric. The state space of the Moran model is the \(N\) simplex

\[
\mathcal{E}_N := \left\{ \eta : E \to \mathbb{N} \middle| \sum_{x \in E} \eta(x) = N \right\}.
\]

The empirical distribution induced by \(\eta \in \mathcal{E}_N\) is defined by

\[
m(\eta) = \sum_{x \in E} \frac{\eta(x)}{N} \delta_x \in \mathcal{M}_1(E),
\]

where \(\mathcal{M}_1(E)\) is the set of probability measures on \(E\). Let \(Q\) be the generator of a continuous-time, non-explosive, irreducible Markov chain, and consider some rates \(V_\mu(x,y) \geq 0\), for every \(x \neq y \in E\) and \(\mu \in \mathcal{M}_1(E)\).

The multi-allelic Moran model is a continuous-time Markov chain evolving on \(\mathcal{E}_N\). The process is at \(\eta \in \mathcal{E}_N\) if there are \(\eta(x)\) individuals of type \(x\), for all \(x \in E\). Between reproduction events, the \(N\) individuals evolve as independent copies of the mutation process generated by \(Q = (Q_{x,y})_{x,y \in E}\). In this sense we call \(Q_{x,y}\), for \(x,y \in E\), the mutation rates. Reproduction events consist of the death of an individual of type \(x\), which is then removed from the population, and the reproduction of an individual of type \(y\), which add \(y\) individual to the population. This happens at the rate \(\eta(y)/N \cdot V_{\mu(y)}(x,y)\). Hence, the transition rate from \(\eta \in \mathcal{E}_N\), with \(\eta(x) > 0\), to \(\eta - e_x + e_y\) is

\[
\eta(x) \left( Q_{x,y} + \frac{\eta(y)}{N} V_{\mu(y)}(x,y) \right),
\]

for every \(x \neq y \in E\), where \(\eta - e_x + e_y\) is the element in \(\mathcal{E}_N\) satisfying

\[
(\eta - e_x + e_y)(z) = \begin{cases} 
\eta(z) & \text{if } z \notin \{x,y\}, \\
\eta(x) - 1 & \text{if } z = x, \\
\eta(y) + 1 & \text{if } z = y.
\end{cases}
\]

∗Corresponding author.

Preprint submitted to Stochastic Processes and their Applications September 28, 2022
We will further detail particular examples but, for the moment, let us see that when $V_{mut}(x, y)$ is constant, each individual dies at the same rate and the parent is chosen uniformly at random over the individuals that are present in the population (this explains the term $\eta(y)/N$ in the transition rate). We can also interpret this rate by the opposed point of view: each individual reproduces at a constant rate, and the dying individual is chosen uniformly at random. This is often called neutral selection in ecology literature, but our models allow choosing various non-constant $V_{mut}(x, y)$. In this sense, we call the rates $V_{f}(x, y)$, for $x, y \in E$ and $\mu \in M_1(E)$, the selection rates.

Note that the reproduction dynamics depends in general on both the types of the parent and the offspring, and may also depend on the empirical distribution induced by the configuration of the population at the current time, in a sense that we will clarify further along in Assumptions (G1) and (C1). The generator of the Moran model is denoted $Q := Q^{\text{mut}} + Q^{\text{sel}}$, where $Q^{\text{mut}}$ and $Q^{\text{sel}}$ act on every function $f \in \mathcal{B}_b(\mathcal{E}_N)$ as follows

\[
(Q^{\text{mut}} f)(\eta) = \sum_{x, y \in E} \eta(x) Q_{x,y} [f(\eta - e_x + e_y) - f(\eta)],
\]

\[
(Q^{\text{sel}} f)(\eta) = \frac{1}{N} \sum_{x, y \in E} \eta(x) \eta(y) V_{f}(x, y)[f(\eta - e_x + e_y) - f(\eta)],
\]

for every $\eta \in \mathcal{E}_N$. Throughout the paper, the following boundedness condition holds:

\[
||V|| := \sup_{\mu \in M_1(E)} \sup_{x, y \in E} V_{\mu}(x, y) < \infty. 
\]

Note that the non-explosion of the process generated by $Q$ and the bound condition (1) let out the possibility of an infinite number of jumps in finite time. Thus, the process generated by $Q$ is well-defined for all $t \geq 0$.

This Moran model is an extension, for $K > 2$, of the model studied by [1]. In general, when generalising the Moran model for more than two allelic types, the selection rates are taken depending only on the children type, i.e. $V_{f}(x, y) = V^{f}(y)$, for all $x, y \in E$ and $\mu \in M_1(E)$, which is called selection at birth or fecundity selection [2, 3, 4]. Moreover, in biological applications it has been also considered models with selection at death or viability selection, when the selection rates only depend on the parent type, i.e. $V_{f}(x, y) = V^{f}(x)$ [3], for all $x, y \in E$ and $\mu \in M_1(E)$. However, the importance of this last model is beyond its biological interpretations: this process is also called Fleming–Viot particle process, which is an interacting particle process intended for the approximation of a quasi-stationary distribution (QSD) of an absorbing Markov chain conditioned on non-absorption. These particle processes have attracted lots of attention in recent years. See for instance [5, 6, 7, 8] for general state spaces, [9, 10, 11, 12] for countable state spaces, and even [13, 14] for finite state spaces.

Structure of the paper

The rest of the paper is organised as follows. In Section 1.1 we state the main results of the paper and comment some of their consequences. In Section 1.2 we consider several examples of mutation and selection rates satisfying Assumptions (G1) and (C1), and such that the exponential uniform in time convergence of the normalised semigroups, namely Assumption (C2), also hold. We end this section with a discussion about some of the possible extensions. Finally, in Section 2, we prove the main results, and other results and proofs are deferred to the appendices.

1.1. Main results

Let us get some insight into the limit of the empirical measure induced by this particle process when the number of particles tends towards infinity. Let us denote by $(\eta_t)_{t \geq 0}$ the continuous-time Markov chain on $\mathcal{E}_N$, generated by $Q$. Although the process generated by $Q$ clearly depends on $N$ and a better notation would be $(\eta_t^{(N)})_{t \geq 0}$, we keep this dependence implicit for the sake of simplicity. By the Kolmogorov equation we know that $\partial_t \mathbb{E}_\eta [m_x(\eta_t)] = \mathbb{E}_\eta [(Qm_x)(\eta_t)]$, where $m_x$ stands for the empirical distribution induced by $\eta$ on the point $x \in E$, i.e. $m_x : \eta \mapsto \eta(x)/N$. Let us thus compute $Qm_x$. On the one hand, it is easy to get

\[
(Q^{\text{mut}} m_x)(\eta) = \sum_{y \in E} Q_{x,y} m_y(\eta),
\]

for every $x \in E$, for all $\eta \in \mathcal{E}_N$. On the other hand,

\[
(Q^{\text{sel}} m_x)(\eta) = -m_x(\eta) \sum_{y \in E} m_y(\eta) [V_{mut}(x, y) - V_{mut}(y, x)].
\]

2
Finally, we get
\[ \partial_t [\mathbb{E}_t [m_i (\eta_1)]] = \sum_{y \in E} Q_{x,y} \mathbb{E}_t [m_i (\eta_1)] - \sum_{y \in E} [V_{m_i}(x, y) - V_{m_i}(y, x)] \mathbb{E}_t [m_i (\eta_1) m_i (\eta_1)]. \]

When the number of individuals \( N \) is large, we expect the Moran process to exhibit a *propagation of chaos phenomenon* and thus the empirical distribution induced by the process approximates the solution of the following nonlinear system of ordinary differential equations:
\[
\partial_t \gamma_t (x) = \sum_{y \in E} Q_{y,x} \gamma_t (y) - \sum_{y \in E} [V_{\gamma_t}(x, y) - V_{\gamma_t}(y, x)] \gamma_t (x) \gamma_t (y),
\]
for all \( x \in E \). For every function \( \phi \) on \( E \) we thus get the nonlinear differential equation
\[
\partial_t \gamma_t (\phi) = \gamma_t (Q_{\gamma_t} \phi),
\]
where \( Q_{\gamma} := Q + \Pi_{\gamma} \) and
\[
\Pi_{\gamma} \phi : x \mapsto \sum_{y \in E} \gamma (y) V_{\gamma}(x, y)[\phi (y) - \phi (x)].
\]
for every probability distribution \( \gamma \) on \( E \).

The main results we provide in this article are related to the speed of convergence of \( (m (\eta_1))_{t \geq 0} \) towards \( (\gamma_t)_{t \geq 0} \) when \( N \to \infty \).

### 1.1.1. Propagation of chaos with general selection rate

Let \( \mathcal{B}_b (E) \) be the set of bounded functions on \( E \) for the uniform norm, which is denoted by \( \| \cdot \| \) and defined by \( \| \phi \| := \sup_{x \in E} | \phi (x) | \). Let us also denote \( \mathcal{B}_i (E) := \{ \phi : E \to \mathbb{R} : \| \phi \| \leq 1 \} \). For two probability distributions \( \mu_1, \mu_2 \in \mathcal{M}_i (E) \) the total variation distance is defined as follows:
\[
\| \mu_1 - \mu_2 \|_{TV} := \sup_{A \subset E} \| \mu_1 (A) - \mu_2 (A) \| = \frac{1}{2} \| \mu_1 (B) - \mu_2 (B) \| = \frac{1}{2} \sum_{x \in E} | \mu_1 (x) - \mu_2 (x) |.
\]
where \( \mu (\phi) \) stands for the mean of \( \phi \) with respect to \( \mu \in \mathcal{M}_i (E) \).

First, we consider the case where the selection rates satisfy the following hypothesis.

**Assumption (G1) (General selection rate).** There exist \( V^d_i, V^p_i \) for \( i \geq 1 \), and a continuous, nonnegative function \( \mu \mapsto V^d_\mu \) from \( \mathcal{M}_i (E), \| \cdot \|_{TV} \) to \( \mathcal{B}_b (E \times E), \| \cdot \| \) such that \( V^d_\mu \) is symmetric in \( E \times E \) for every \( \mu \), and
\[
V^d_\mu (x, y) = \sum_{i \geq 1} V^d_i (x) V^p_i (y) + V^p_i (x, y),
\]
for all \( \mu \in \mathcal{M}_i (E) \) and \( x, y \in E \). In addition, \( V^d_\mu - V_\mu^p \in \mathcal{B}_b (E \times E) \) and
\[
\left\{ \sum_{i \geq 1} | V^d_i - V^p_i |, \sup_{i \geq 1} V^d_i, \sup_{i \geq 1} V^p_i \right\} \subset \mathcal{B}_b (E).
\]

**Remark 1.1 (Decomposition of \( V \in \mathcal{B}_b (E \times E) \)).** Since the state space \( E \) is countable, any function \( V \in \mathcal{B}_b (E \times E) \) can be written as follows
\[
V (x, y) = \sum_{z \in E} I_{\{z \}} (x) V (z, y) = \sum_{z \in E} I_{\{z \}} (y) V (x, z).
\]
Thus, it is always possible to decompose \( V \) in a countable sum as in (6). In particular, Assumption (G1) is always satisfied when \( E \) is finite. However, when \( E \) is infinite such decomposition does not necessarily satisfy the boundedness conditions in Assumption (G1), namely it does not necessarily satisfy
\[
\sum_{z \in E} | V (z, \cdot) | \in \mathcal{B}_b (E) \text{ or } \sum_{z \in E} | V (\cdot, z) | \in \mathcal{B}_b (E).
\]
Consider for example the case where \( E = \mathbb{N} \) and \( V^d_i (x, y) = a_i b_i \), where \( (a_i)_{i \in \mathbb{N}} \) and \( (b_i)_{i \in \mathbb{N}} \) are bounded and are the general terms of two divergent series. Taking \( V^d_0 \equiv 0 \), \( V^d_i : x \mapsto 1_{\{i \}} (x) \) and \( V^p_i : y \mapsto a_i b_i \) (or \( V^d_i : x \mapsto a_i b_i \) and \( V^p_i : y \mapsto 1_{\{i \}} (y) \)), then condition (7) is not satisfied. However, taking \( V^d_i : x \mapsto a_i \), \( V^p_i : y \mapsto b_i \) and \( V^d_i V^p_i \equiv 0 \), for all \( i \geq 2 \), one easily checks that Assumption (G1) is satisfied.
Remark 1.2 (Rates giving the same mean-field limit). Note that, because of (4), the mean-field limit is invariant as soon as the same the terms $V_p(x, y) - V_p(y, x)$, for all $x, y$ $\in$ $E$, do not change. In particular, the mean-field limit does not depend on the symmetric term in Assumption (C1).

Now, for every $V$ $\in$ $\mathcal{B}_b(E \times E)$ and $y_0$ $\in$ $M_1(E)$, let us define the “implicit Feynman–Kac” flow:

$$\gamma_t^Y (\phi) := E_{y_0} \left[ \phi(X_t) \exp \left\{ \int_0^t \sum_{x \in E} [V(X_s, x) - V(x, X_s)] \gamma_s^Y (x) ds \right\} \right]. \quad (8)$$

Then, $(\gamma_t^Y)_{t \geq 0}$ is a solution of the Cauchy problem

$$\partial_t \mu_t (\phi) = \mu_t (\tilde{Q}_t \phi), \quad \text{with} \quad \mu_0 (\phi) = \gamma_0 (\phi), \quad (9)$$

where

$$\tilde{Q}_t \phi : x \mapsto (Q \phi) (x) + \sum_{y \in E} \mu (y) V(x, y) [\phi(y) - \phi(x)].$$

See [15, p. 25] and the references therein for more details on the implicit Feynman–Kac. Besides, as suggested in the above-mentioned reference, following the martingale arguments in [16], one can check that $(\gamma_t^Y)_{t \geq 0}$ is in fact the unique solution of (9). According to these references, we claim the following result.

Lemma 1.1 (Existence of the solution of (9)). Assume that Assumption (G1) is satisfied. For every $\mu_0$ $\in$ $M_1(E)$, there is a unique solution $(\mu_t)_{t \geq 0}$ of the differential equation (9), and it is given by (8) taking $V = V_p - V_p^p$.

Consider also the following assumption.

Assumption (I) (Initial condition). The empirical measure induced by the particle process at $t = 0$ converges towards the initial distribution $\mu_0$ $\in$ $M_1(E)$ in $L_p$, for every $p \geq 1$. More precisely, for every $p \geq 1$, there exists a constant $C_p > 0$ such that

$$\sup_{\phi \in \mathcal{B}_b(E)} E \left[ |m(\eta_0) (\phi) - \mu_0 (\phi)|^p \right] \leq \frac{C_p}{N^{p/2}}.$$

Note that Assumption (I) is verified when initially all the particles are sampled independently with distribution $\mu_0$, as the next lemma shows.

Lemma 1.2 (Control of the initial error). Assume that initially the $N$ particles are sampled independently according to $\mu_0$ $\in$ $M_1(E)$. Then, Assumption (I) is verified.

The proof of Lemma 1.2 is deferred to Appendix A. We include Assumption (I) in order to be able to apply our results to a wider class of situations, that than described in Lemma 1.2.

Theorem 1.3 (Propagation of chaos). Suppose that Assumptions (G1), and (I) are verified. Then, for every $T \geq 0$ and $p \geq 1$, there exist two positive constants $\alpha_p$ and $\beta_p$, possibly depending on $p$, such that

$$\sup_{\phi \in \mathcal{B}_b(E)} E \left[ \sup_{[0,T]} |m(\eta_t) (\phi) - \mu_t (\phi)|^p \right]^{1/p} \leq \alpha_p \frac{\sqrt{1 + T}}{\sqrt{N}} e^{\beta_p T},$$

where $(\mu_t)_{t \geq 0}$ is as in Lemma 1.1, with initial condition $\mu_0$ $\in$ $M_1(E)$ as in Assumption (I).

The proof of Theorem 1.3 is deferred to Section 2.2.

Let $(x_n)_{n \geq 1}$ be an enumeration of the elements in $E$. We define the following distance in $M_1(E)$:

$$\|\mu_1 - \mu_2\|_w := \sum_{k \geq 1} 2^{-k} |\mu_1 (x_k) - \mu_2 (x_k)|.$$

Note that the space $M_1(E)$ with the convergence in law (the weak topology) is metrisable with this distance. As a consequence of Theorem 1.3 we get the following result.

Corollary 1.4 (Convergence of the empirical measure). Suppose that Assumptions (G1), and (I) are verified. Then, for every $T \geq 0$ and $p \geq 1$, there exist two positive constants $\alpha_p$ and $\beta_p$, such that

$$E \left[ \left( \sup_{[0,T]} \|m(\eta_t) - \mu_t\|_w \right)^p \right]^{1/p} \leq \alpha_p \frac{\sqrt{1 + T}}{\sqrt{N}} e^{\beta_p T},$$

where $(\mu_t)_{t \geq 0}$ is as in Lemma 1.1, with initial condition $\mu_0$ $\in$ $M_1(E)$ as in Assumption (I).
Corollary 1.4 is proved in Section 2.2.

Note that this ensures a functional convergence in $L^p(C([0, T], M_1(E)))$:

$$m(\eta) \xrightarrow{L^p} \mu,$$

with an estimation of the speed of convergence. Furthermore, Theorem 1.3, for $p = 4$, and a Borel–Cantelli argument imply the convergence $m(\eta) \xrightarrow{L^\infty} \mu$ in the weak sense, where c.c. denotes the complete (or universal) convergence (cf. [17, Def. 1.6]). In particular, this implies $m(\eta) \xrightarrow{L^\infty} \mu$, when $N \to \infty$, in the weak sense, no matter in which space the random variables are coupled.

Theorem 1.3 is a generalisation for multi-allelic Moran models with more than two allelic types, of Proposition 3.1 in [1], where the uniform convergence on compacts time intervals in probability is proved. The speed of convergence in Theorem 1.3 can also be related to existing results that ensure the convergence of the empirical measure induced by a Moran type (or Fleming – Viot) particle process towards the law of an absorbing process conditioned to non-absorption. See for instance [18, Prop. 3.5] [19, Lemma 3.1], [6, Thm. 2.2], [20, Thm. 1.1], and [21, Thm. 5.10 and Cor. 5.12]. See also [22, Thm. 3.1 and Rmk. 3.2] where the almost sure convergence (and also the complete convergence) is proved when the state space is finite. As far as we know, Theorem 1.3 and Corollary 1.4 are the first results ensuring the convergence uniformly on compacts in $L^p$, for all $p \geq 1$, with speed of convergence of order $1/\sqrt{N}$ for multi-allelic Moran models with general selection rates in the sense of Assumption (G1), in discrete countable state spaces, not necessarily finite. The idea behind the proof is closed to the methods in [23]: it consists in finding a martingale indexed by the interval $[0, T]$, whose terminal value at time $T$ is precisely $m(\eta_T)(\phi) - \mu_T(\phi)$ plus a term whose $L^p$ norm can be controlled, for any $\phi \in B_0(E)$. Thereafter, the final result comes by a Grönwall type argument, similarly to the proof of Proposition 1 in [24]. Nevertheless, [23] does not contain any uniform bound as in Theorem 1.3.

1.1.2. Uniform in time propagation of chaos for additive selection rates

Under a more specific expression for the selection rates, we prove a uniform in time bound for the convergence of $(m(\eta_t))_{t \geq 0}$ towards $(\mu_t)_{t \geq 0}$, when $N \to \infty$. Consider the following kind of selection rates that we call additive selection.

**Assumption (C1) (Additive selection).** The selection rates are uniformly bounded as in (1). Moreover, there exist two continuous nonnegative functions $\mu \mapsto V^d_\mu$ and $\mu \mapsto V^b_\mu$ from $(\mathcal{M}_1(E), \| \cdot \|_{TV})$ to $(B_0(E), \| \cdot \|)$; and a continuous, nonnegative function $\mu \mapsto V^s_\mu$ from $(\mathcal{M}_1(E), \| \cdot \|_{TV})$ to $(B_0(E \times E), \| \cdot \|)$ such that $V^s_\mu$ is symmetric on $E \times E$, for every $\mu \in \mathcal{M}_1(E)$ and

$$V_\mu(x, y) = V^d_\mu(x) + V^b_\mu(y) + V^s_\mu(x, y),$$

for all $x, y \in E$ and $\mu \in \mathcal{M}_1(E)$. Furthermore, there exists a function $\Lambda \in B_0(E)$ such that

$$\Lambda(x) = V^d_\mu(x) - V^d_\mu(x), \quad (10)$$

for every $x \in E$.

**Remark 1.3 (Selection rates independent on $\mu$).** When the selection rates do not depend on $\mu$, Assumption (C1) reduces to the existence of $V^d, V^b \in B_0(E)$ and a symmetric $V^s \in B_0(E \times E)$ such that

$$V(x, y) = V^d(x) + V^b(y) + V^s(x, y).$$

Let $\Lambda \in B_0(E)$ be a fixed function. Typical examples of functions $V^b$ and $V^d$ satisfying this condition are

$$V^b = (\Lambda - c)^+ \text{ and } V^d = (\Lambda - c)^-,$$

for a fixed constant $c \in \mathbb{R}$, where we use the standard notation $(x)^+ := \max\{x, 0\}$ and $(x)^- := -\min\{x, 0\}$. These are in fact the selection rates considered by Angeli et al. [25, § 3.3] in the context of cloning algorithms. Moreover, the case $c = 0$ is considered in Example 3.1-(2) in [23]. Note that in this case, Assumption (G1) is also verified.

From a biological point of view, the parameter $c \in \mathbb{R}$ can be seen as a fitness parameter. Let us assume that $V^d_\mu$ is null for simplicity, and denote by $\xi_t^{(i)}$ the type of the $i$-th individual, for $1 \leq i \leq N$, at time $t \geq 0$. Then, if $\Lambda(\xi_t^{(i)}) \leq c$, the $i$-th individual dies and another randomly chosen individual reproduces. On the other hand, if $\Lambda(\xi_t^{(i)}) \geq c$ a random chosen individual dies and the $i$-th reproduces with rate $\Lambda(\xi_t^{(i)}) - c$.

Another example of particular interest is when $V^b = 0$. Notice that the Moran process with these selection rates is in fact a Fleming – Viot particle process (cf. [9]).
Remark 1.4 (Selection rates depending on \( \mu \)). Consider a fixed function \( \Lambda \in \mathcal{B}_0(E) \). Typical examples of functions \( V^b_\mu \) and \( V^d_\mu \) are:

\[
V^b_\mu = (\Lambda - \mu(\Lambda))^+ \quad \text{and} \quad V^d_\mu = (\Lambda - \mu(\Lambda))^-.
\]

These are the selection rates considered in [15, § 1.5.2, p. 35], see also Example 3.1-(3) in [23]. In this case, the biological interpretation of \( \mu(\Lambda) \) is similar to that of the parameter \( c \) in the previous remark. Indeed, the fitness coefficient evolves in time according to the evolution of the population.

Consider that Assumption (C1) is satisfied. Then, from (5) we can recover the nonlinear differential equation

\[
\partial_t \gamma(t, \phi) = \gamma(t, (Q + \Lambda)\phi - \gamma(t, \Lambda)\phi),
\]

where \( \Lambda \) is as defined in (10). Consider the Feynman – Kac semigroup \( (P^t_\Lambda)_{t \geq 0} \), where

\[
P^t_\Lambda(\phi) : x \mapsto \mathbb{E}_x \left[ \phi(X_t) \exp \left\{ \int_0^t \Lambda(X_s)ds \right\} \right]
\]

whose generator is \( Q + \Lambda \). Let us define the normalised version of this semigroup as follows

\[
\mu_t(\phi) := \frac{\mu_0 P^t_\Lambda(\phi)}{\mu_0 P^t_\Lambda(1)},
\]

where 1 denotes the all-one function on \( E \). Then, \( (\mu_t)_{t \geq 0} \) is the solution of the nonlinear differential equation (11) with initial value \( \mu_0(\phi) \) for \( t = 0 \) [15, Eq. (1.17)].

Furthermore, the definition of \( (\mu_t) \) in (12) is invariant by translation of the function \( \Lambda \), i.e. for every real \( \beta \) we have that \( \mu_t(\phi) = \frac{\mu_0 P^t_\Lambda(\phi)}{\mu_0 P^t_\Lambda(1)} \). In particular, one can always interpret \( (\mu_t)_{t \geq 0} \) as the distribution of an absorbed Markov chain conditioned to non-absorption up to time \( t \) with killing rate \( \kappa = \sup \Lambda - \Lambda \). This naturally relates the study of the behaviour of \( (\mu_t)_{t \geq 0} \) when \( t \to \infty \), to the theory of quasi-stationary distributions (QSD), see e.g. [26, § 7] and [27, § 1.2].

Consider the following assumptions related to the control in the norm \( L^p \) of the initial error and the exponential convergence of \( (\mu_t)_{t \geq 0} \), as defined by (12), towards a unique limit, for every initial distribution on \( \mu_0 \in \mathcal{M}_1(E) \).

Assumption (C2) (Uniform exponential ergodicity of the normalised semigroup). There exist a distribution \( \mu_\infty \in \mathcal{M}_1(E) \) and \( C, \gamma > 0 \), such that for every initial distribution \( \mu_0 \in \mathcal{M}_1(E) \) and for all \( t \geq 0 \):

\[
\|\mu_t - \mu_\infty\|_{\text{TV}} \leq Ce^{-\gamma t},
\]

where \( (\mu_t)_{t \geq 0} \) is defined as in (12).

Assumption (C2) is always satisfied when \( E \) is finite. In this case, inequality (13) was proved by Darroch and Seneta [28] and the result comes as a consequence of the Perron–Frobenius Theorem (see [29, Thm. 8] for the specific context of quasi-stationary distributions). The case where \( E \) is countable is more delicate and has attracted lots of attention and several methods have been applied. See for example the work of Del Moral and Miclo [30], the articles of Champagnat and Villemonais on exponential convergence towards the quasi-stationary distribution, specifically [31, 32], and also the works of Bansaye et al. [33, 34]. See also the recent work of Del Moral et al. [35] on the stability of positive semigroups. In Section 1.2, we provide several examples of process where Assumption (C2) holds.

We are now in a position to state our main results for the multi-allelic Moran model with additive selection.

Theorem 1.5 (Uniform in time propagation of chaos). Under Assumptions (I), (C1) and (C2), for every \( p \geq 1 \), there exists a constant \( C_p \), such that

\[
\sup_{\phi \in \mathcal{B}_1(E)} \sup_{t \geq 0} \mathbb{E}[|m(\eta_t)(\phi) - \mu_t(\phi)|^p]^{1/p} \leq \frac{C_p}{\sqrt{N}}.
\]

The proof of Theorem 1.5 is based on the methods developed by Rousset [23] and it is deferred to Appendix C.

Corollary 1.6 (Convergence of the empirical measure). Suppose that Assumptions (I), (C1) and (C2) are verified. Then, for every \( p \geq 1 \), there exists a constant \( C_p > 0 \), such that

\[
\sup_{t \geq 0} \mathbb{E}[|m(\eta_t) - \mu_t|_{\text{TV}}^p]^{1/p} \leq \frac{C_p}{\sqrt{N}}.
\]
The proof of Corollary 1.6 is analogous to that of Corollary 1.4, and we skip it for the seek of brevity.

Remark 1.5 (Almost sure convergence). Corollary 1.6, for $p = 4$, and a Borel–Cantelli argument imply the convergence $m(\eta_T) \overset{c.c.}{\rightarrow} \mu_T$ in $M_1(E)$, when $N \rightarrow \infty$, for every $T \geq 0$, where c.c. denotes the complete (or universal) convergence. In particular, this implies $m(\eta_T) \overset{a.s.}{\rightarrow} \mu_T$, when $N \rightarrow \infty$, for every $T \geq 0$. Note that in contrast with Corollary 1.4, Corollary 1.6 does not ensure the convergence in $C([0,T], M_1(E))$.

As a consequence of Theorem 1.5, it is also possible to control the bias of one particle, following a similar method of that in [23, Thm. 4.2]. Thus, since these results are somehow expected, we have decided, for the sake of brevity, to place the statement (Theorem C.6) and its proof in the appendices.

The following result ensures the exponential ergodicity of the unnormalised semigroup.

**Lemma 1.7** (Exponential ergodicity of the unnormalised semigroup). Suppose that Assumptions (C1) and (C2) are verified. Then, there exists a unique triplet $(\mu_\infty, h, \lambda) \in M_1(E) \times \mathcal{B}_b(E) \times \mathbb{R}$, of eigenelements of $Q + \Lambda$ such that $h$ is strictly positive, $\mu_\infty(h) = 1$,

$$
\mu_\infty P_t^\Lambda = e^{t} \mu_\infty \text{ and } P_t^\Lambda(h) = e^{t}h.
$$

Moreover, there exist $C, \gamma > 0$ such that for all $t \geq 0$:

$$
\sup_{(\mu_\infty, h, \lambda) \in M_1(E)} ||e^{-t} \mu_0 P_t^\Lambda - \mu_0(h)\mu_\infty||_{TV} \leq C e^{-\gamma t}.
$$

Furthermore, $\lambda \leq 0$ whether $\Lambda \leq 0$.

This result is basically a consequence of Theorem 2.1 of [32]. Lemma 1.7 establishes an exponential control on the speed of convergence of the unnormalised semigroup. A similar estimate is stated by Angeli et al. [25, Assumption 2.2] as hypothesis. However, their assumption implies that the eigenfunction $h$ is constant, which in practice makes their assumption only valid when $\Lambda$ ($\mathcal{V}$ in their notation) is constant.

Let us define

$$
S_\mu(\phi) := \sum_{x,y \in E} (\phi(x) - \phi(y))^2 V_\mu(x,y) \mu(x) \mu(y),
$$

for every $\phi \in \mathcal{B}_b(E)$, and the operator $W_{t,T}$ for $t \leq T$ as follows

$$
W_{t,T} : \phi \mapsto \frac{P^\Lambda_{t,T}(\phi)}{\mu_1(P^\Lambda_{t,T}(1))},
$$

Our last two results are addressed to the study of the asymptotic square error of the approximation of $\mu_T$ by $m(\eta_T)$ when $T, N \rightarrow \infty$. These results are particularly important when the Moran process is used for approximating a quasi-stationary distribution. Let us define the asymptotic quadratic errors:

$$
\sigma^2_1(\phi) := \lim_{N \rightarrow \infty} N^{-1} \mathbb{E} \left[ \left( m(\eta_T)(\phi) - \mu_T(\phi) \right)^2 \right],
$$

$$
\sigma^2_\infty(\phi) := \lim_{t \rightarrow \infty} \sigma^2_1(\phi),
$$

for every $\phi \in \mathcal{B}_b(E)$. First, we prove the asymptotic normality of the bias and we provide explicit expressions for $\sigma^2_1(\phi)$ and $\sigma^2_\infty(\phi)$. Then, we use this expression to show how to define another Moran process approaching the same distribution $\mu_\infty$, with smaller or equal asymptotic square error.

In order to prove the asymptotic normality of the statistic $\sqrt{N}(m(\eta_T)(\phi) - \mu_T(\phi))$, for every $T \geq 0$, we naturally need to ask, in addition to the law of large numbers established by Assumption (I), for the existence of a central limit theorem on the initial empirical distribution, as stated in the following hypothesis.

**Assumption (I')** (Asymptotic normality for initial empirical distribution). For every $\phi \in \mathcal{B}_b(E)$, the empirical measure induced by the particle process at $t = 0$ satisfy the following condition: $\sqrt{N}(m(\eta_0)(\phi) - \mu_0(\phi))$ converges in law towards a centred Gaussian distribution of variance $\mu_0(\phi^2)$, when $N \rightarrow \infty$.

Analogously to Lemma 1.2, we have that Assumption (I') is verified when initially the $N$ particles are sampled independently according to $\mu_0 \in M_1(E)$. The proof of this result is a consequence of the classical functional central limit theorem for martingales.
Theorem 1.8 (Asymptotic normality). Suppose that Assumptions (I), (I'), (C1) and (C2) are verified. Then, for every \( \phi \in \mathcal{B}_b(E) \) and \( T \geq 0 \), we have that \( \sqrt{N}(m(t\eta_T)(\phi) - \mu_T(\phi)) \) converges in law, when \( N \) goes to infinity, towards a Gaussian centered random variable of variance

\[
\sigma^2_T(\phi) = \text{Var}_\mu(\phi) + \int_0^T S_\mu(W_{s,T}(\phi_T))ds + 2 \int_0^T \mu_s \left(W_{s,T}(\phi_T)^2 \left(V_{\mu_s}^b + \mu_s(V_{\mu_s}^d)\right)\right)ds,
\]

where \( \text{Var}_\mu(\phi) \) stands for the variance with respect to \( \mu \), \( \phi_T := \phi - \mu_T(\phi) \) and \( S_\mu \) and \( W_{s,T} \) are as defined in (15) and (16), respectively. Moreover,

\[
\sigma^2_\infty(\phi) = \text{Var}_\mu(\phi) + \int_0^\infty \alpha_s \left(P_s^\mu(\phi)\right)ds + 2 \int_0^\infty \lambda_s \left(P_s^\mu(\phi)\right)^2 \left(V_{\mu_s}^b + \mu_s(V_{\mu_s}^d)\right)ds,
\]

where \( \text{Var}_\mu(\phi) \) stands for the variance with respect to \( \mu_\infty \), also \( \phi_\infty := \phi - \mu_\infty(\phi) \) and \( \lambda \) is the eigenvalue in the statement of Lemma 1.7.

The proof of Theorem 1.8 can be found in Section 2.3. Note that the two integrals in the expression of \( \sigma^2_T(\phi) \) in Theorem 1.8 converge as a consequence of Lemma 1.7.

Let us mention the relation between Theorem 1.8 and some existing results in the literature. When \( V = \phi \) is null. and the selection rates do not depend on \( \mu \), our result is related to Proposition 3.7 in [5]. Indeed, for the particular choice of the parameters of the model in [5] as follows: \( V = 2V^b \), \( V' = 2V^d \) and \( \rho = 1/2 \), Theorem 1.8 can be obtained from Proposition 3.7 in [5]. See also the multivariate fluctuation study in [18, § 3.3.2].

Moreover, when \( V^b_\mu = 0 \) and \( V^d_\mu = 0 \) and \( \lambda = -V^d \leq 0 \), we get

\[
\sigma^2_\infty(\phi) = \text{Var}_\mu(\phi) - 2\lambda \int_0^\infty \alpha_s \left(P_s^\mu(\phi)\right)ds.
\]

When the process \( (\eta_t)_{t\geq 0} \) is ergodic and converges in law to some random variable \( \eta_\infty \), when \( t \to \infty \), Theorem 1.8 states that \( \sqrt{N}(m(\eta_T)(\phi) - \mu_\infty(\phi)) \) converges to a centered Gaussian law of variance \( \sigma^2_\infty(\phi) \), when \( N \to \infty \). Indeed, recall that a Gaussian sequence converges in law if their first two moments converge. In particular, we recover (and extend) the result of Lellièvre et al. [14, Thm. 2.4] for finite state spaces. Note that the negative constant \( \lambda \) in the previous expression for \( \sigma^2_\infty(\phi) \) is the opposite of that in [14, Thm. 2.4].

The expression for \( \sigma^2_\infty(\phi) \), when \( V^d_\mu = 0 \), is also similar to the expression for the asymptotic square error in Theorem 4.4 in [23]. However, the results in [23] do not include the asymptotic normality we prove in Theorem 1.8. See also Corollary 2.7 and Remark 2.8 [7] for a central limit theorem for the empirical measure induced by Fleming – Viot particle systems.

Note that the three summands in the expression of \( \sigma^2_\infty(\phi) \) in Theorem 1.8 are positive, for every \( T \geq 0 \). Moreover, the limit \( (\mu_s)_{s\geq 0} \) is invariant by the choice of the symmetric component \( V^b_\mu \) in Assumption (C1). As a consequence, for a given selection rate \( V_\mu \) we can obtain another Moran process approaching the same limit distribution taking the selection rate \( V_\mu - \Sigma_\mu \geq 0 \), where \( \Sigma_\mu \) is a symmetric function in \( \mathcal{B}_b(E \times E) \). We thus get the following result.

Corollary 1.9 (Moran process with smaller asymptotic square error). Suppose that Assumptions (I), (I'), (C1) and (C2) are verified. Let \( (\eta^b_t)_{t\geq 0} \) and \( (\eta^d_t)_{t\geq 0} \) be the Moran processes with the same mutation rates and selection rates given by \( V^b_\mu \) and \( V^d_\mu \), respectively, where

\[
\Sigma_\mu(x,y) := \min \left\{ V^b_\mu(x), V^d_\mu(x) \right\} 1_A + \min \left\{ V^b_\mu(y), V^d_\mu(y) \right\} 1_{\bar{A}} + V^b_\mu(x,y),
\]

where \( 1_A \) stands for the indicator function on \( A \subset E \). Then,

\[
\lim_{N\to\infty} N E \left[ (m(\eta_T)(\phi) - \mu_T(\phi))^2 \right] \leq \lim_{N\to\infty} N E \left[ (m(\eta_T)(\phi) - \mu_T(\phi))^2 \right],
\]

for all \( T \geq 0 \).

Note that the selection rate \( V_\mu - \Sigma_\mu \) in the statement of Corollary 1.9 satisfies Assumption (C1). The proof of the previous result is thus a simple consequence of Theorem 1.8. In Example 1, we discuss the application of this result to the simple case of the bi-allelic Moran model, that is, when the cardinality of \( E \) is 2.
1.2. Examples

In this section we consider several examples where Assumption (C2) holds, for the process with additive selection satisfying (C1) and (G1).

The first example we consider is precisely when $E = \{1, 2\}$. This example offers us the opportunity to compare our result with the existing results on bi-allelic Moran models and the Fleming – Viot particle process approximating the QSD of an absorbing Markov chain with two transient states.

**Example 1 (Two-allelic Moran model).** Consider the two-allelic Moran model on $E = \{1, 2\}$ with mutation rate matrix

$$Q = \begin{pmatrix} -a & a \\ b & -b \end{pmatrix}$$

and selection rates $V_{1,2} = p$ and $V_{2,1} = q$, with $a, b > 0$ and $p, q \geq 0$. Let us assume, without loss of generality, that $p \leq q$.

The empirical probability measure induced by this Moran process approaches the QSD of the absorbing Markov chain on $E \cup \{\hat{d}\}$, where $\hat{d}$ is an absorbing state, with infinitesimal generator

$$
\begin{pmatrix} -(a + p) & a & p \\ b & -(b + q) & q \\ 0 & 0 & 0 \end{pmatrix}.
$$

See [1] and [20, § 3] for a deeper treatment of this model and the limit behaviour of the interacting particle process approximating its QSD.

Theorem 1.3 applied in this case improves Proposition 3.1 in [1] and Theorem 3.1 (see also Remark 3.2) in [22]. Furthermore, Theorem 1.5, and also (C8), improve the control of the speed of convergence to stationarity of the bounds obtained in [12, Cor. 1.5] and [8, Thm. 2.4].

Likewise, as a consequence of Corollary 1.9, we have that the Moran model with the same mutation rate matrix $Q$ and with selection rates $V_{1,2} = 0$ and $V_{2,1} = q - p$, approaches the same QSD but with smaller asymptotic square error.

Consider now the case when $p = q$. Then, the empirical distribution induced by the particle system approaches the stationary distribution of the process generated by $Q$. When $E$ is finite, the results about the spectrum of the generator $Q$ in [36] imply that the asymptotic ergodicity is independent of the value of $p$. Besides, Corollary 1.9 implies that a minimal asymptotic variance is obtained when there is no selection, that is, when the particle system is simply driven by $N$ independent particles, where each of them is driven by $Q$.

We now focus on the classical birth and death Markov chain. The existence and uniqueness of QSD for these models have been well understood. We rely on existing results to find explicit conditions on the parameters of the birth and death chain that are equivalent to the existence of a unique QSD and the uniform exponential convergence.

**Example 2 (Birth and death chain).** Consider two positive sequences $(b_i)_{i \geq 1}$ and $(d_i)_{i \geq 1}$ and the Markov chain on $\mathbb{N}$ with rate matrix

$$Q_{x,y} := \begin{cases} b_x & \text{if } x \geq 1 \text{ and } y = x + 1 \\ d_x & \text{if } x \geq 2 \text{ and } y = x - 1 \\ 0 & \text{otherwise}, \end{cases}$$

and $\Lambda := d_1 I_{x=1}$. Note that 0 is an absorbing state and $\mathbb{N}$ is a transient class. Van Doorn [37, Thm. 3.2] has found an explicit condition characterising the three possible cases: there is no QSD, there exists a unique QSD or there exists an infinite continuum of QSDs. See also [29, § 4]. Furthermore, Martínez et al. [38, Thm. 2] have proved that the existence of a unique QSD is in fact equivalent to the uniform exponential convergence of the law of the conditioned process to its QSD, i.e. Assumption (C2). In addition, this occurs if and only if

$$\sum_{k \geq 2} \frac{1}{d_k \alpha_k} \sum_{r \geq k} \alpha_r < \infty, \quad (17)$$

where $\alpha_r := \prod_{i=1}^{r-1} b_i \prod_{i=1}^{r} d_i$. We refer also to [31, § 4.1], where the uniform exponential convergence is ensured for some generalisations of the classical birth and death chain.

We end this section presenting two quantitative criteria on the transition rates and on the spectral elements, respectively, ensuring the uniform exponential convergence in (C2).
**Example 3** (A criterion on the mutation and selection rates). We next describe a criterion on the transition rates, which is Theorem 3 in [38]. Assume that (C1) is verified, and the following condition holds: there exists a finite subset \( K \subset E \) such that

\[
\inf_{y \in E \backslash K} \left( \Lambda(y) + \sum_{x \in K} Q_{y,x} \right) > \sup_{y \in E} \Lambda(y).
\]

Then, (C2) holds. This provides an easy condition on the mutation rates and \( \Lambda \) to verify Assumption (C2), which is applicable to a wide range of Moran processes with discrete countable state space. See also [12, Thm. 1.1], where a stronger condition is asked in order to provide, via a coupling technique, explicit constants for the upper bound in (C2).

**Example 4** (A spectral criterion). Assume there exists a triplet \( (\mu_{\infty}, h, \lambda) \in M_1(E) \times B_b(E) \times \mathbb{R} \), of eigenelements of \( Q + \Lambda \) such that \( \lambda \) is an eigenvalue of \( Q + \Lambda \), \( h \) is strictly positive, \( \mu_{\infty}(h) = 1 \), and

\[
\mu_{\infty} P_t^h = e^{\lambda t} \mu_{\infty} \quad \text{and} \quad P_t^h(h) = e^{\lambda t}.
\]

Note that these are the eigenelements in the statement of Lemma 1.7. Let us also assume \( \|h^{-1}\| \leq \infty \), which is always true if \( E \) is finite, and furthermore, there exists \( \epsilon > 0 \) such that the set

\[
K_\epsilon := \{ x \in E : \Lambda(x) \geq \lambda - \epsilon \}
\]

is finite. Then, (C2) is verified.

The proof is based on the methods in [39], and is very similar to the proofs of Proposition 3.2 in [23] and Proposition A.5 in [25]. Consider the Doob’s \( h \)-transform

\[
P_t^{h,h} := \frac{1}{h} e^{-h t} P_t^h(h),
\]

which is the semigroup associated to an irreducible continuous-time Markov chain on \( E \) with generator \( Q^h \) acting on every \( \phi \in M_1(E) \) as follows

\[
Q^h(\phi) = \frac{1}{h} (Q + \Lambda - \lambda)(h\phi).
\]

Furthermore, the process driven by \( Q^h \) has stationary distribution \( \mu_{\infty}^h \in M_1(E) \), satisfying \( \mu_{\infty}^h(\phi) = \mu_{\infty}(h\phi) \), for every \( \phi \in B_b(E) \). Now, note that \( h^{-1} \) is bounded on \( K_\epsilon \), and consequently there exists \( \beta > 0 \) such that

\[
Q^h(h^{-1}) = \frac{\Lambda - \lambda}{h} \leq -\epsilon h^{-1} + \beta 1_{K_\epsilon}.
\]

Thus, condition (\( \tilde{D} \)) in [39] is verified and using their Theorem 5.2-(c) we get the \( h^{-1} \)-uniform exponential ergodicity of \( P_t^{h,h} \) as follows

\[
\sup_{|t| \leq h^{-1}} \left| \frac{1}{h(x)} e^{-\delta t} P_t^h(hg) - \mu_{\infty}(hg) \right| \leq \frac{C}{h(x)} e^{\beta}.\]

Multiplying by \( h(x) \) the previous inequality and taking \( \phi = hg \in B_1(E) \), we get the uniform exponential ergodicity (14). Finally, it is not difficult to verify that Assumption (C2) also holds, using the exponential ergodicity (14) and the fact that \( \|h^{-1}\| < \infty \).

**Remark 1.6.** In [25, Appendix], the authors state a similar result, but they do not include the fact that \( \|h^{-1}\| \) is bounded in their hypothesis. We have not found or understood the arguments making the authors claim that \( \|h^{-1}\| \) is bounded when the state space is locally compact [25, p. 150]. We next provide an example of a birth and death chain whose generator allows an unbounded eigenfunction associated to its greatest eigenvalue. Indeed, let us consider the following parameters for the birth and death chain in Example 2: \( b_i = b, \) and \( d_i = d, \) for all \( i \geq 2, \) with \( b < d. \) Moreover, take \( b_1 > b \) and \( d_1 = d(e-1). \) Hence, taking \( h : n \in \mathbb{N} \mapsto e^{-n} \) we get \( (Q + \Lambda)(h) = \lambda h, \) for \( \lambda = b(e^{-1} - 1) + d(e-1) > 0. \) Moreover, \( K_\epsilon = \{ 1 \} \) is finite (compact), but \( \|r^{-1}\| = \infty \). In fact, the infinite sum (17) diverges, thus this birth and death chain allows an infinite number of QSDs (cf. [37, Thm. 3.2]).

**1.3. Final remarks**

The fact that the state space \( E \) is countable and discrete is not necessary for our proofs. Therefore, we expect to be able to extend all our results to more general Markov processes following the same methods.

There are lots of possible directions to continue this research. Maybe, the more natural is to weaken the condition (C2) and consider the case where there exists a minimal QSD but the exponential convergence is not
uniform on $M_1(E)$. Lots of research have been done for controlling the domains of attraction of the minimal QSD. See for example the works of Champagnat and Villemonais [40, 41, 42, 43, 8] and also the related works of Bansaye et al. [33, 34], and the references therein. Another interesting research direction is to improve the upper bound constants in Theorem 1.3. In this sense, the results of Arnaudon and Del Moral [21, Thm. 5.10 and Cor. 5.12] suggest that a bound of type $C_T$ could hold. Hence, a future research direction would be to combine the approach of [21] and this paper to improve the upper bound in Theorem 1.3. Moreover, the results in [21, § 5] could also be useful to obtain exponential concentration inequalities, which is a natural continuation of the research on the long time behaviour of the empirical measure induced by Moran type particle processes.

2. Proof of the main results

This section contains the proofs of Theorems 1.3 and 1.8, which are, in our opinion, the main results in the present article with original proof methods. In Section 2.1 we briefly study the martingale problem associated to the generator $Q_N$. Then, in Sections 2.2 and 2.3 we prove Theorems 1.3 and 1.8, respectively. Other comments, results and proofs are deferred to the appendices.

2.1. The associated martingale problem

For a Markovian generator $L$, its associated “carré-du-champ” operator, denoted $\Gamma_L$, is defined by

$$\Gamma_L : \phi \mapsto L(\phi^2) - 2\phi L\phi.$$ 

See, for example, [44, Def. 2.5.1] for more details on the theory related to this operator.

It is not difficult to prove that $\Gamma_L$ satisfies

$$\Gamma_L(\phi)(\eta) = \sum_{x \in E} \eta(x) \sum_{y \in E} \left( Q_{xy} + V_{\mu}(x,y) \frac{\eta(y)}{N} \right) [\psi(\eta - e_x + e_y) - \psi(\eta)]^2,$$

where for every $\eta \in E_N$. We recall that $m(\eta)$ denotes the empirical distribution induced by $\eta \in E_N$. Moreover, $m(\eta)(\phi)$ stands for the mean of $\phi$ with respect to $m(\eta)$, for every $\phi \in B_b(E)$. Suppose that one of Assumptions (G1) or (C1) is verified. In either case, let us denote $V^\mu := V_\mu - V^\mu_\eta$.

**Lemma 2.1.** Suppose that one of Assumptions (G1) or (C1) is verified. We have

$$Q(m(\cdot)(\phi)) = m(\cdot) \left( \widetilde{Q}_m(\phi) \right),$$

$$\Gamma_Q(m(\cdot)(\phi)) = \frac{1}{N} m(\cdot) \left( \Gamma_{\widetilde{Q}_m}(\phi) \right),$$

where

$$\widetilde{Q}_\mu \phi : x \mapsto (Q\phi)(x) + \sum_{y \in E} \mu(y) \widetilde{V}_\mu(x,y)[\phi(y) - \phi(x)],$$

$$Q\mu \phi : x \mapsto (Q\phi)(x) + \sum_{y \in E} \mu(y) V_\mu(x,y)[\phi(y) - \phi(x)],$$

for every $\phi \in B_b(E)$ and all $x \in E$.

The proof of Lemma 2.1 is deferred to Appendix B.1.

Using this result, we can study the martingale problem associated to the process $(m(\eta_t)(\psi_t))_{t \geq 0}$.

**Proposition 2.2** (Martingale decomposition). Let $\psi$ be a function on $E \times \mathbb{R}_+$ such that $\psi(\cdot, x)$ is continuously differentiable in $\mathbb{R}_+$, for every $x \in E$ and $\psi(\cdot) \in B_b(E)$, for every $t \in \mathbb{R}_+$. Then, the process $(M_t(\psi))_{t \geq 0}$ such that

$$M_t(\psi) := m(\eta_t)(\psi_t) - m(\eta_0)(\psi_0) - \int_0^t m(\eta_s) \left( \partial_s \psi_s + \widetilde{Q}_m(\psi_s) \right) ds,$$

where $\widetilde{Q}_\mu$ is defined as in Lemma 2.1, is a local martingale, with predictable quadratic variation given by

$$\langle M(\psi) \rangle_t = \frac{1}{N} \int_0^t m(\eta_s) \left( \Gamma_{\widetilde{Q}_m}(\psi_s) \right) ds.$$

Moreover,

$$|\Delta M_t(\psi_s)| \leq \frac{2 ||\psi||}{N}.$$
The proof of Proposition 2.2 is deferred to Appendix B.2.

Now, for a function \( \psi \) on \( E \times \mathbb{R}_+ \), continuously differentiable in \( \mathbb{R}_+ \), we get

\[
dm(\eta)(\psi) = dM_1(\psi) + m(\eta)\partial \psi + \tilde{Q}_{m_0}(\psi)dt.
\]

Thus, the empirical measure induced by the particle process is a perturbation of the dynamic given by (5), by a martingale whose jumps and predictable quadratic variation are of order \( 1/N \).

2.2. Proof of Theorem 1.3

Throughout this section, we will suppose that the expression for the selection rates in Assumption (G1) is verified. We will denote \( \tilde{Q}_\mu = Q + \Pi_\mu \) as in Lemma 2.1, namely

\[
\tilde{\Pi}_\mu \phi : x \mapsto \sum_{y \in \mathcal{E}} \mu(y) V(x, y)[\phi(y) - \phi(x)],
\]

where \( V = V_\mu - V_\mu^* \), which is independent of \( \mu \in \mathcal{M}_1(E) \).

The family of generators \( (\tilde{Q}_\mu)_{\mu \geq 0} \) defines an inhomogeneous-time Markov chain, which is associated to a map \( (s, t) \mapsto P(s, t) \), for all \( s \leq t \) such that \( P(s, s) = I \), for all \( s \geq 0 \) and satisfies the forward and backward Kolmogorov equations:

\[
\begin{align*}
\partial_t P(s, t) &= P(s, t) \tilde{Q}_\mu, & \text{for } t \geq s, \\
\partial_s P(s, t) &= -\tilde{Q}_\mu P(s, t), & \text{for } s \leq t.
\end{align*}
\]

See [45] and the references therein. Besides, using the forward Kolmogorov equation (18), we get that \( (\mu_t)_{t \geq 0} \) as in Lemma 1.1, satisfies the propagation equation \( \mu_T = \mu P(t, T) \). Note that since \( P(t, T) \) is the propagator of an inhomogeneous Markov chain, we get \( \|P(t, T)\| \leq 1 \) for all \( t \in [0, T] \), which implies

\[
\int_0^T \|P(s, T)(\phi)\|^p ds \leq T.
\]

Let us now study the control in \( \mathbb{L}^p \) norm for the martingales that are obtained taking the functions \( t \in [0, T] \mapsto P(\cdot, T)(\phi) \) and \( t \in [0, T] \mapsto P(\cdot, T)(\phi)^2 \) in Proposition 2.2. Note that,

\[
\partial_t \{P(t, T)(\phi)\} = -\tilde{Q}_\mu \{P(t, T)(\phi)\}.
\]

From Proposition 2.2, we get the following local martingale for \( t \in [0, T] \):

\[
\mathcal{M}_t\{P(\cdot, T)(\phi)\} := m(\eta_t)\{P(t, T)(\phi)\} - m(\eta_0)\{P(0, T)(\phi)\}
\-
\int_0^t m(\eta_s)\{\tilde{Q}_{m_0}(P(s, T)(\phi)) - \tilde{Q}_\mu(P(s, T)(\phi))\} ds.
\]

Similarly, we get

\[
\mathcal{M}_t\{P(\cdot, T)(\phi)^2\} := m(\eta_t)\{P(t, T)(\phi)^2\} - m(\eta_0)\{P(0, T)(\phi)^2\}
\-
\int_0^t m(\eta_s)\{\tilde{Q}_{m_0}(P(s, T)(\phi)^2) - \tilde{Q}_\mu(P(s, T)(\phi)^2)\} ds
\]

Moreover, by definition, \( P(T, T)(\phi) = \phi \).

Lemma 2.3 (Control of the predictable quadratic variation). Assume that Assumption (G1) is verified. For every test function \( \phi \in \mathcal{B}_1(E) \), we have

\[
\mathcal{N}\{\mathcal{M}(P(\cdot, T)(\phi)^2)\}_t \leq C(t + 1) - \mathcal{M}_t\{P(\cdot, T)(\phi)^2\}, \text{ for all } t \in [0, T].
\]

The proof of Lemma 2.3 is postponed to Appendix B.3.

The following lemma is a generalisation of the classical Burkholder–Davis–Gundy (BDG) inequality [46, Thm. 20.12]. The lower bound is obtained from the classical BDG inequality. The proof of the upper bound can be found in [23, Lemma 6.2].
Lemma 2.4 (BDG inequalities). Let \((M_t)_{t \geq 0}\) be a quasi-left-continuous (i.e. with continuous predictable increasing process) locally square-integrable martingale with \(M_0 = 0\) and bounded jumps

\[
\sup_{0 \leq t \leq T} |\Delta M_t| \leq a < +\infty.
\]

Then, there exists a constant \(C\), possibly depending on \(q\), such that

\[
E \left( \sup_{t \in [0,T]} |M_t|^q \right) \leq C E \left( \langle M \rangle_T \right)^{q/2} \leq C \sum_{k=0}^{q} a^{2q - 2k+1} E \left( \langle M \rangle_T \right)^{2k}.
\]

We are now in position to establish a control of the quadratic variation of the martingale \(\{M_t(P(\cdot, T)(\phi))\}_{t \in [0,T]}\).

Lemma 2.5 (Control of the quadratic variation). Assume that Assumption (G1) is verified. For all \(p > 0\) and all test function \(\phi \in B_1(E)\), there exists a positive \(C_p\) (possibly depending on \(p\)) such that

\[
E \left[ \left( \left( M \left( P(\cdot, T)(\phi) \right) \right) \right)^p \right] \leq \frac{C_p(t + 1)^p}{N^p}, \quad \text{for all } t \in [0, T]
\]

The proof of this result is inspired by the proof of Theorem 5.4 in [23], and it is deferred to Appendix B.4.

Proof of Theorem 1.3. Let us denote \(\psi_{s,T} := P(s,T)(\tilde{\phi}_T)\), which satisfies the backward Kolmogorov equation (18). We have that \((M_t(\psi_{s,T}))_{t \in [0,T]}\), defined as in Proposition 2.2, is a local martingale. Moreover,

\[
M_t(\psi_{s,T}) = m(\eta_T)(\psi_{s,T}) - m(\eta_0)(\psi_0,T) - \int_0^T m(\eta_s)(-\tilde{\theta}_{\eta_s}(\psi_{s,T}) + \tilde{\theta}_{\theta_0}(\psi_{s,T})) ds
\]

\[
= m(\eta_T)(\phi) - \mu_T(\phi) - m(\eta_0)(\psi_0,T) - \int_0^T m(\eta_s)(\tilde{\Pi}_{m(\eta_s)}(\psi_{s,T}) - \tilde{\Pi}_s(\psi_{s,T})) ds. \tag{21}
\]

Note that for any two probability measures \(\lambda\) and \(\mu\) on \(E\) and for every function \(\psi\) on \(E\) we have

\[
\lambda(\tilde{\Pi}_s(\psi)) = -\mu(\tilde{\Pi}_s(\psi)), \tag{22}
\]

where \(\tilde{\Pi}_s\) acts on a test function \(\phi\) as follows:

\[
\tilde{\Pi}_s(\psi) : x \mapsto \sum_{y \in E} \lambda(y)V(y,x)[\psi(y) - \psi(x)].
\]

Now, using (21) and (22) we get

\[
M_T(\psi_{s,T}) = m(\eta_T)(\phi) - \mu_T(\phi) - m(\eta_0)(\psi_0,T) + \int_0^T (m(\eta_s) - \mu_s)(\tilde{\Pi}_{m(\eta_s)}(\psi_{s,T})) ds, \tag{23}
\]

where \(\psi_{s,T} = P(s,T)(\tilde{\phi}_T)\).

Hence, using (23), we can ensure the existence of a positive constant \(C_p > 0\), depending on \(p\), such that

\[
\sup_{t \in [0,T]} |m(\eta_t)(\phi) - \mu_t(\phi)|^p \leq C_p \left( |m(\eta_0)(\psi_0,T) - \mu_0(\psi_0,T)|^p + \sup_{t \in [0,T]} |M_t(\psi_{s,T})|^p + R_p(T) \right),
\]

where

\[
R_p(T) = \int_0^T \left[ \left| (m(\eta_s) - \mu_s)(\tilde{\Pi}_{m(\eta_s)}(\psi_{s,T})) \right| \right]^p ds.
\]

The initial error can be controlled using Assumption (I). Indeed, there exists \(C_p^{(1)} > 0\) such that

\[
E \left[ |m(\eta_0)(\psi_0,T) - \mu_0(\psi_0,T)|^p \right] \leq \frac{C_p^{(1)}}{N^{p/2}}.
\]

Furthermore, using Lemma 2.5 and BDG inequality, we can ensure the existence of a positive constant \(C_p^{(2)}\) such that

\[
E \left[ \sup_{t \in [0,T]} |M_t(P(\cdot, T)(\tilde{\phi}_T))|^p \right] \leq \frac{C_p^{(2)} (T + 1)^{p/2}}{N^{p/2}}.
\]
for all $p \geq 1$. Let us denote by $\lambda_s$, the (random) signed measure $\lambda_s := m(\eta_s) - \mu_s$.

We have
\[ R_p(T) \leq C_p \left( \int_0^T \left| \lambda_s \left( \hat{\Pi}_{\psi, T}(x) \right) \right|^p \, ds + \int_0^T \left| \lambda_s \left( \Pi_{\psi, T}(x) \right) \right|^p \, ds \right). \]

The first term in the last expression can be controlled, since $\hat{\Pi}_{\psi, T}$ is not random. Indeed,
\[ I_1(T) := \int_0^T \left( m(\eta_s) - \mu_s \left( \hat{\Pi}_{\psi, T}(x) \right) \right)^p \, ds \leq 2^p \| V \|^p \int_0^T \left( m(\eta_s) - \mu_s \left( \Pi_{\psi, T}(x) \right) \right)^p \, ds. \]

For the second term, note that
\[ I_2(T) := \int_0^T \left| \sum_{i \geq 1} \sum_{j \in \mathcal{E}} \lambda_i(x) \lambda_j(y) V^i(y) \left( \psi_{s, T}(y) - \psi_{s, T}(x) \right) \right|^p \, ds \]
\[ = \int_0^T \left| \sum_{i \geq 1} \lambda_i (V^i - V^i_1) \lambda_i (V^p_1 \Phi \psi_{s, T} + \lambda_i (V^p - V^p_1) \psi_{s, T}) \right|^p \, ds. \]

Hence, using the Assumption (G1), we conclude that there exists a positive constant $C_p^{(3)}$, depending on $p$, such that
\[ I_2(T) \leq C_p^{(3)} \int_0^T \left( \sum_{i \geq 1} \left| V^i - V^i_1 \right| \right)^p \, ds, \]
where $\kappa = \left\| \sum_{i \geq 1} \left| V^i - V^i_1 \right| \right\| < \infty$.

Let us define
\[ \Phi_p(T) := \sup_{\phi \in \mathcal{B}_{1}(E)} \mathbb{E} \left[ \sup_{s \leq T} \left| m(\eta_s)(\phi) - \mu_s(\phi) \right|^p \right]. \]

Thus, taking the expectations of $I_1(T)$ and $I_2(T)$, we can ensure the existence of a positive constant $C_p^{(4)}$ such that
\[ \mathbb{E}[R_p(T)] \leq C_p^{(4)} \int_0^T \Phi_p(s) \, ds. \]

Hence, there exists $K_p > 0$ such that
\[ \Phi_p(T) \leq K_p \left( 1 + (1 + T)^{p/2} \right) + N^{p/2} + C_p^{(4)} \int_0^T \Phi_p(s) \, ds, \]
which, using Grönwall inequality, ensures the existence of two constants $\alpha_p$ and $\beta_p$ (possibly depending on $p$) such that
\[ \Phi_p(T)^{1/p} \leq \alpha_p \frac{\sqrt{1 + T}}{\sqrt{N}} e^{\beta_p T}. \]

\[ \square \]

**Proof of Corollary 1.4.** Let $(x_n)_{n \geq 1}$ be the enumeration of the elements in $E$, in the definition of the distance $\| \cdot \|_\infty$. Note that
\[ \mathbb{E} \left[ \sup_{t \in [0, T]} \left( \| m(\eta_t) - \mu_t \|_\infty \right)^p \right]^{1/p} \leq \mathbb{E} \left[ \sup_{t \in [0, T]} \left( \sum_{k \geq 1} 2^{-k} \| m(\eta_t)(x_k) - \mu_t(x_k) \|_\infty \right)^p \right]^{1/p} \]
\[ \leq \sum_{k \geq 1} 2^{-k} \left[ \mathbb{E} \left[ \sup_{t \in [0, T]} \| m(\eta_t)(x_k) - \mu_t(x_k) \|_\infty \right] \right]^{p/2} \]
\[ \leq \alpha_p \frac{\sqrt{1 + T}}{\sqrt{N}} e^{\beta_p T}, \]
where the first inequality is a consequence of the Minkowski’s inequality for infinite sums and the last inequality is a consequence of Theorem 1.3. \[ \square \]
2.3. Proof of Theorem 1.8

In this section, we will consider that Assumption (C1) is verified. Namely, the selection rates can be written as $V_{\mu}(x, y) = V^0_{\mu}(x) + V^1_{\mu}(y) + V^2_{\mu}(x, y)$, where $V^0_{\mu}$ is symmetric. Note that under (C1), equation (5) becomes equivalent to

$$
\partial_t \gamma_t(\phi) = \gamma_t((Q + \Lambda)\phi - \gamma_t(\Lambda)\phi),
$$

(24)

where $\Lambda = V^0_{\mu} - V^1_{\mu}$.

Using (24) we can simplify the expression of the martingales $(M_t(\psi))_{t \geq 0}$ in Proposition 2.2 as follows

$$
M_t(\psi) = m(\eta_t(\psi_t) - m(\eta_0(\psi_0) - \int_0^t m(\eta_s) (\partial_s \psi_s + (Q + \Lambda)(\psi_s) - m(\eta_s(\Lambda) \cdot \psi_s) ds),
$$

(25)

for every bounded function $\psi$ on $E \times \mathbb{R}$, such that $\psi(x)$ is continuously differentiable in $\mathbb{R}$, for every $x \in E$; and $\psi_t(\cdot) \in B_b(E)$, for every $t \in \mathbb{R}$. The previous expression is essential: for a suitable choice of the function $\psi$, we can control the integral part in the expression of $M_t(\psi)$.

Let us study the martingales obtained from Proposition 2.2 using as argument the function

$$
W_{t,T}(\phi) := \frac{P^\Lambda_{T-t}(\phi)}{\mu_t(P^\Lambda_{T-t}(1))},
$$

i.e. where $W_{t,T}$ defined as in (16) and $\phi \in B_b(E)$ and $t \in [0, T]$. Note that $W_{t,T}$ verifies the propagation equation $\mu_t(\phi) = \mu_t(W_{t,T}(\phi))$. Besides,

$$
\partial_t \mu_t(P^\Lambda_{T-t}(1)) = \partial_t \left( \frac{\mu_t P^\Lambda_{T-t}(1)}{\mu_t P^\Lambda_{T-t}(1)} \right) = - \frac{\mu_t P^\Lambda_{T-t}(1)}{\mu_t P^\Lambda_{T-t}(1)} \partial_t \mu_t(P^\Lambda_{T-t}(1)) = - \mu_t(P^\Lambda_{T-t}(1)) \partial_t \mu_t(\Lambda).
$$

Thus,

$$
\partial_t W_{t,T}(\phi) = -((Q + \Lambda)W_{t,T}(\phi) = - \frac{\partial_t \mu_t(P^\Lambda_{T-t}(1)) \partial_t \mu_t(P^\Lambda_{T-t}(1))}{\mu_t(P^\Lambda_{T-t}(1))} = - (Q + \Lambda)W_{t,T}(\phi) + \mu_t(\Lambda)W_{t,T}(\phi).
$$

Hence, $W_{t,T}(\phi)$ is the solution of the Cauchy problem $\partial_t \psi_s = -((Q + \Lambda)\psi_s)$, with condition $\psi_T = \phi$. Let us denote $\psi_{s,T} := W_{t,T}(\phi)$, for any $\phi \in B_b(E)$. Note that,

$$
\partial_t (\psi_{s,T}) = -((Q + \Lambda - \mu_t(\Lambda))\psi_{s,T} \text{ and } \partial_t (\psi^2_{s,T}) = -2\psi_{s,T} \cdot ((Q + \Lambda - \mu_t(\Lambda))\psi_{s,T}).
$$

We are in position to define the martingales $(M_t(\psi_{s,T}))_{t \in [0, T]}$ and $(M_t(\psi_{s,T}^2))_{t \in [0, T]}$, as stated in Proposition 2.2, which under Assumption (C1) can be written as follows

$$
M_t(\psi_{s,T}) = m(\eta_s(\psi_{s,T}) - m(\eta_0(\psi_{s,0})) - \int_0^t m(\eta_s(\psi_{s,T})) [m(\eta_s(\Lambda)) - \mu_s(\Lambda)] ds,
$$

(26)

$$
M_t(\psi_{s,T}^2) = m(\eta_s(\psi_{s,T}^2) - m(\eta_0(\psi_{s,0}))^2 - 2 \int_0^t m(\eta_s(\psi_{s,T}^2)) [\mu_s(\Lambda) - m(\eta_s(\Lambda))] ds - \Psi_t,
$$

(27)

where

$$
\Psi_t := \int_0^t m(\eta_s)((Q + \Lambda - m(\eta_s(\Lambda))(\psi_{s,T}^2) - 2\psi_{s,T} \cdot (Q + \Lambda - m(\eta_s(\Lambda))(\psi_{s,T}))) ds.
$$

Furthermore, note that

$$
\mu(\phi \cdot Q_t \phi) = \mu(\phi(Q + \Lambda - \mu(\Lambda))\phi) + \mu(\phi)\mu(V_{\phi, t}) - \mu(\phi^2)\mu(V_{\phi, t}),
$$

(28)

where $V_{\phi, t} := V^0_{\phi} + V^1_{\phi}$, for every $\mu \in M_t(E)$.

Thus, the predictable quadratic variation of $(M_t(\psi_{s,T}))_{t \in [0, T]}$ satisfies

$$
N(M(\psi_{s,T})_t := \int_0^t m(\eta_s)(\Gamma_{Q_{m(\eta_s)(\psi_{s,T})}} ds
$$

$$
= \int_0^t m(\eta_s)(\tilde{Q}_{m(\eta_s)(\psi_{s,T})} - 2\psi_{s,T} \cdot \tilde{Q}_{m(\eta_s)(\psi_{s,T})}) ds + \int_0^t S_{m(\eta_s)}(\psi_{s,T}) ds,
$$

15
where $S_{\mu}$ is defined as in (15), for every $\mu \in M_{1}(E)$. Now, using (28) we obtain
\[
N\{M(\psi_{s,T})\}_{t} = \Psi_{t} - 2 \int_{0}^{t} \mu(\psi_{s,T})m(\eta_{s})dW_{s,T}(\psi_{s,T})ds + 2 \int_{0}^{t} m(\eta_{s})\psi_{s,T}^{b}dW_{m(\eta_{s})}(\psi_{s,T})ds
+ 2 \int_{0}^{t} m(\eta_{s})\psi_{s,T}^{2}dW_{m(\eta_{s})}(\psi_{s,T})ds + \int_{0}^{t} S_{m(\eta_{s})}(\psi_{s,T})ds.
\]
Then, using (27) we can substitute the value of $\Psi_{t}$ into this last expression and get
\[
N\{M(\psi_{s,T})\}_{t} = \mu(\psi_{s,T})^{2} + \int_{0}^{t} m(\eta_{s})\psi_{s,T}^{2}dW_{m(\eta_{s})}(\psi_{s,T})ds + \int_{0}^{t} S_{m(\eta_{s})}(\psi_{s,T})ds + R_{t},
\]
where
\[
R_{t} := -2 \int_{0}^{t} m(\eta_{s})\psi_{s,T}^{2}dW_{m(\eta_{s})}(\psi_{s,T})ds - 2 \int_{0}^{t} m(\eta_{s})\psi_{s,T}^{2}dW_{m(\eta_{s})}(\psi_{s,T})ds.
\]

The key component in the proof of Theorem 1.8 is a central limit theorem for the martingale $(M_{t}(\psi_{s,T}))_{t \in [0,T]}$. Let us first introduce an auxiliary result.

Consider the process $(\tilde{M}_{t}(W_{s,T}(\phi_{T})))_{t \in [0,T]}$ defined as
\[
\tilde{M}_{t}(W_{s,T}(\phi_{T})) := \sqrt{N}m(\eta_{0}))(W_{0,T}(\phi_{T}))) + \sqrt{N}M_{t}(W_{s,T}(\phi_{T})),
\]
for $t \in [0,T]$. Then, $(\tilde{M}_{t}(W_{s,T}(\phi_{T})))_{t \in [0,T]}$ is a martingale, with initial value
\[
\tilde{M}_{0}(W_{s,T}(\phi_{T})) = \sqrt{N}m(\eta_{0}))(W_{0,T}(\phi_{N})).
\]

**Proposition 2.6** (Central limit theorem). The martingale $(\tilde{M}_{t}(W_{s,T}(\phi_{T})))_{t \in [0,T]}$ converges in law when $N \to \infty$ towards a Gaussian martingale whose variance at time $t \in [0,T]$ is $\sigma_{2}^{2}(\phi)$, defined as
\[
\sigma_{2}^{2}(\phi) := \mu(\psi_{s,T})^{2} + \int_{0}^{t} \mu(\psi_{s,T})m(\eta_{s})dW_{s,T}(\psi_{s,T})ds + \int_{0}^{t} S_{m(\eta_{s})}(\psi_{s,T})ds,
\]
and $\psi_{s,T} = W_{s,T}(\phi_{T})$.

**Proof.** Using Theorem 3.11 in [47, § 8], and arguing as in the proofs of Proposition 3.31 in [18] and Proposition 3.7 in [5], we only need to check that the result holds for the initial value $\tilde{M}_{0}(W_{s,T}(\phi_{T})) = \sqrt{N}m(\eta_{0}))(W_{0,T}(\phi_{T})))$ and that $N(M_{t}(\psi_{s,T}))$ converges in probability to a continuous function, when $N$ goes to infinity. The first point is in fact Assumption (1'). Furthermore, Theorem 1.5 implies, by a Borel–Cantelli argument, the convergence $m(\eta_{s}) \to \mu_{s}$, when $N \to \infty$, for all $s \geq 0$, see Remark 1.5. Now, using the Cauchy–Schwarz inequality and Theorem 1.5 (see also equation (C.6)), we easily prove that $R_{t}$, defined by (30), converges to 0 in probability and that $N(M_{t}(\psi_{s,T}))$ converges to the continuous function $\sigma_{2}^{2}(\phi)$ in probability, when $N \to \infty$, which concludes the proof. \qed

**Proof of Theorem 1.8.** As a consequence of Proposition 2.6 and (26) we have that
\[
\tilde{M}_{T}(W_{s,T}(\phi_{T})) = \sqrt{N}m(\eta_{T}))(\phi_{T}) - \mu_{T}(\phi_{T}) - \sqrt{N} \int_{0}^{T} m(\eta_{s})(\psi_{s,T})ds,
\]
converges to a Gaussian random variable of variance $\sigma_{2}^{2}(\phi)$, when $N \to \infty$. Thus, the first part of Theorem 1.8 comes from the fact that
\[
\sqrt{N} \int_{0}^{T} m(\eta_{s})(\psi_{s,T})ds,
\]
converges to 0 almost surely when $N \to \infty$. Indeed, this is a consequence of the Cauchy–Schwarz inequality and Theorem 1.5 (see also (C.6)). Thus, $m(\eta_{T}))(\phi_{T}) - \mu_{T}(\phi_{T})$ converges in law to a centred Gaussian law with variance
\[
\sigma_{2}^{2}(\phi) = \mu_{T}(\phi_{T})^{2} + \int_{0}^{T} S_{m(\eta_{s})}(W_{s,T}(\phi_{T}))ds + 2 \int_{0}^{T} \mu_{s}(W_{s,T}(\phi_{T}))^{2}dW_{s,T}(\phi_{T})ds,
\]
Consider now the change of variables $u = T - s$ in the last integral of the previous expression, and then take limit when $T \to \infty$. The final result comes due the following convergences:

$$
\mu_{T-s} \xrightarrow{T \to \infty} \mu_\infty,
$$

$$
\phi_T = \phi - \mu_T(\phi) \xrightarrow{T \to \infty} \phi - \mu_\infty(\phi),
$$

$$
W_{T-s,T}(\phi_T) \xrightarrow{T \to \infty} \frac{P_T^\Lambda(\phi_\infty)}{\mu_\infty P_T^\Lambda(1)} = e^{-\lambda s} P_s^\Lambda(\phi_\infty),
$$

where the last identity is a consequence of the identity $\mu_\infty(\Lambda) = \lambda$ and

$$
\mu_t(P_T^\Lambda(1)) = \exp\left\{ \int_0^T \mu_s(\Lambda) \, ds \right\}.
$$

Indeed, (31) is a consequence of the next two identities

$$
\frac{d}{dt} \ln \left( \mu_t P_T^\Lambda(1) \right) = \mu_t(\Lambda) \quad \text{and} \quad \mu_t(P_T^\Lambda(1)) = \frac{\mu_0(P_T^\Lambda(1))}{\mu_0(P_T^\Lambda(1))}.
$$

Acknowledgements

The authors would like to thank Djalil Chafaï, Pierre Del Moral, Simona Grusea, Vlada Limic and Denis Villemonais for their careful reading and helpful comments. The authors would like to extend their gratitude to the anonymous reviewers whose comments and suggestions greatly improved the quality of this manuscript. The work of the first author was partially supported by the ANR NOLO ANR-20-CE40-0015. The work of the second author was partially supported by the ITI IRMIA++.

References

[1] F. Cordero, The deterministic limit of the Moran model: a uniform central limit theorem, Markov Process. Relat. Fields 23 (2) (2017) 313–324.
[2] R. Durrett, Probability models for DNA sequence evolution, 2nd Edition, Probability and its Applications (New York), Springer, New York, 2008. doi:10.1007/978-0-387-78166-6.
[3] C. A. Murhead, J. Wakeley, Modeling multialelic selection using a Moran model, Genetics 182 (4) (2009) 1141–1157. doi:10.1534/genetics.108.089474.
[4] A. Etheridge, Some mathematical models from population genetics, Vol. 2012 of Lecture Notes in Mathematics, Springer, Heidelberg, 2011, lectures from the 39th Probability Summer School held in Saint-Flour, 2009, École d’Eté de Probabilités de Saint-Flour. [Saint-Flour Probability Summer School].
[5] P. Del Moral, L. Miclo, Particle approximations of Lyapunov exponents connected to Schrödinger operators and Feynman–Kac semigroups, ESAIM Probab. Stat. 7 (2003) 171–208. doi:10.1051/ps:2003001.
[6] D. Villemonais, General approximation method for the distribution of Markov processes conditioned not to be killed, ESAIM Probab. Stat. 18 (2014) 441–467. doi:10.1051/ps/2013045.
[7] F. Cérou, B. Delyon, A. Guyader, M. Rousset, A central limit theorem for Fleming–Viot particle systems, Ann. Inst. Henri Poincaré, Probab. Stat. 56 (1) (2020) 637–666. doi:10.1214/19-AIHP976.
[8] N. Champagnat, D. Villemonais, Convergence of the Fleming–Viot process toward the minimal quasi-stationary distribution, ALEA, Lat. Am. J. Probab. Math. Stat. 18 (1) (2021) 1–15. doi:10.30757/ALEA.v18-01.
[9] P. Ferrari, N. Maríć, Quasi Stationary Distributions and Fleming–Viot processes in countable spaces, Electron. J. Probab. 12 (2007) no. 24, 684–702. doi:10.1214/EJP.v12-415.
[10] P. Grossman, M. Jonckheere, Simulation of quasi-stationary distributions on countable spaces, Markov Process. Related Fields 19 (3) (2013) 521–542.
[11] A. Asselah, P. A. Ferrari, P. Groisman, M. Jonckheere, Fleming–Viot selects the minimal quasi-stationary distribution: the Galton–Watson case, Ann. Inst. Henri Poincaré, Probab. Stat. 52 (2) (2016) 647–668. doi:10.1214/14-AIHP635.
[12] B. Cloez, M.-N. Thai, Quantitative results for the Fleming–Viot particle system and quasi–stationary distributions in discrete space, Stochastic Process. Appl. 126 (3) (2016) 680–702. doi:10.1016/j.spa.2015.09.016.
[13] A. Asselah, P. A. Ferrari, P. Grossman, Quasi-stationary distributions and Fleming–Viot processes in finite spaces, J. Appl. Probab. 48 (2) (2011) 322–332. doi:10.1239/jap/1308662630.
[14] T. Lelievre, L. Pillaud-Vivien, J. Reygner, Central limit theorem for stationary Fleming–Viot particle systems in finite spaces, ALEA, Lat. Am. J. Probab. Math. Stat. 15 (2) (2018) 1163–1182. doi:10.30757/ALEA.v15-43.
[15] P. Del Moral, Feynman–Kac formulae, Genealogical and interacting particle systems with applications, New York, NY: Springer, 2004. doi:10.1007/978-1-4684-9393-1.
[16] P. Del Moral, L. Miclo, A Moran particle system approximation of Feynman–Kac formulae, Stochastic Process. Appl. 86 (2) (2000) 193–216. doi:10.1016/S0304-4149(99)00094-0.
A. Proof of Lemma 1.2

Let us first prove the following result, which has an independent interest.

**Lemma A.1** ($L^p$ norm bound for sum of i.i.d. centred r.v.). *Let us consider $Y_1, Y_2, \ldots$ a sequence of independent identically distributed random variables with zero-mean and finite second moment, such that $E[|Y_i|^p] < \infty$, for a given $p \geq 1$. Then, there exists a universal constant $C_p$ such that

$$
\left( E \left( \frac{1}{N} \sum_{i=1}^{N} Y_i \right)^p \right)^{1/p} \leq \frac{C_p}{\sqrt{N}}
$$

*Proof.* First note that for $p \leq 2$ we get the following result as a consequence of Jensen inequality for concave functions:

$$
E \left( \frac{1}{N} \sum_{i=1}^{N} Y_i \right)^p = E \left( \left( \frac{1}{N} \sum_{i=1}^{N} Y_i \right)^2 \right)^{p/2} \leq \left( E \left( \frac{1}{N} \sum_{i=1}^{N} Y_i \right)^2 \right)^{p/2} = \left( \frac{E[Y_1^2]}{N} \right)^{p/2}.
$$

For $p > 2$, the proof follows from Marcinkiewicz–Zygmund inequality, which is a consequence of the BDG inequality for discrete-time martingales. Indeed, the Marcinkiewicz–Zygmund inequality (cf. [48]) ensures us that

$$
E \left( \frac{1}{N} \sum_{i=1}^{N} Y_i \right)^p \leq \frac{K_p}{N^{p/2}} E[|Y_1|^p].
$$

(A.1)

Thus,

$$
E \left( \frac{1}{N} \sum_{i=1}^{N} Y_i \right)^p \leq \frac{C_p}{N^{p/2}},
$$

where $C_p = \begin{cases} \frac{E[Y_1^2]^{p/2}}{N} & \text{if } p \leq 2 \\ \frac{K_p E[|Y_1|^p]}{N} & \text{if } p > 2. \end{cases}$

□

**Remark A.1** (Qualitative results for the Marcinkiewicz–Zygmund constant $K_p$). See the work of Ren and Liang [48] for a qualitative study of the constant $K_p$ in inequality (A.1). They show that $(K_p)^{1/p}$ grows like $\sqrt{p}$, when $p \to \infty$, and give the estimate $K_p \leq (3 \sqrt{2})^p p^{p/2}$.

B. Auxiliary results for proving Theorem 1.3

**Proof of Lemma 2.2.** Note that $m(\eta_0)(\phi) = \frac{1}{N} \sum_{i=1}^{N} \phi(\xi^{(i)}_0)$, where $\xi^{(i)}_0$, for $i = 1, \ldots, N$ are independent random variables. Moreover, $\phi(\xi^{(i)}_0)$ has mean $\mu_0(\phi)$, for all $i = 1, \ldots, N$. Thus,

$$
m(\eta_0)(\phi) - \mu_0(\phi) = \frac{1}{N} \sum_{i=1}^{N} \phi(\xi^{(i)}_0) - N \mu_0(\phi),
$$

can be written as a sum of $N$ zero-mean random variables. The result comes from Lemma A.1. □

**B.1. Proof of Lemma 2.1**

The first identity is simply a consequence of (2) and (3), and the fact that

$$
\mu(\mathcal{Q}_p \phi) = \mu(\tilde{\mathcal{Q}}_p \phi).
$$

(B.1)

Now, to prove the second identity, note that

$$
\Gamma_\phi(m(\cdot)(\phi))(\eta) = \sum_{x \in E} \frac{m(\eta)}{N} \sum_{y \in E} \left( Q_{x,y} + V_{m(\eta)(x,y)} \frac{m(\eta)}{N} \right) \left[ m(\eta) - e_x + e_y \right][\phi(x) - \phi(\eta)]^2
$$

$$
= \frac{1}{N} \sum_{x \in E} \frac{m(\eta)}{N} \left( Q_{x,y} + V_{m(\eta)(x,y)} \frac{m(\eta)}{N} \right) \left[ \phi(x) - \phi(\eta) \right]^2
$$

$$
= \frac{1}{N} \sum_{x \in E} \left( Q_{x,y} + V_{m(\eta)(x,y)m(\eta)} \right) \left[ \phi(x) - \phi(\eta) \right]^2 m_\phi(\eta)
$$

$$
= \frac{1}{N} m(\eta) \left( \Gamma_{\mathcal{Q}_m(\phi)} \right). \square
$$
2.2. Proof of Proposition 2.2

The usual martingale problem associated to \((\eta_t)_{t \geq 0}\) implies that for every function \(\phi\) on \(E\), the process

\[
t \mapsto m(\eta_t)(\phi) - m(\eta_0)(\phi) - \int_0^t Q(m(\eta_s)(\phi))\,ds
\]

is a local martingale. Note that the equality is due to the first identity in Lemma 2.1. Then, for a function \(\psi\) on \(E \times \mathbb{R}_+\), continuously differentiable in \(\mathbb{R}_+\), the Itô formula implies that \((\mathcal{M}_t(\psi))_{t \geq 0}\) is a local martingale, as desired.

The predictable quadratic variation is obtained using the identity

\[
\langle \mathcal{M}(\psi) \rangle_t = \int_0^t \Gamma_2(m(\eta_s)(\psi_s))\,ds,
\]

and the final result comes from the second identity in Lemma 2.1.

Furthermore, the bound for the jump is due to the fact that each jump only concerns one particle that jumps from one position to another. \(\square\)

B.3. Proof of Lemma 2.3

The predictable quadratic variation of the martingale \(\{\mathcal{M}_t(P(\cdot, T)(\phi))\}_{t \in [0,T]}\) satisfies

\[
N\left(\mathcal{M}(P(\cdot, T)(\phi))\right)_t = \int_0^t m(\eta_s)\left(\Gamma_{Q_{\mu(t)}}(P(s, T)(\phi))\right)\,ds
\]

\[
= \int_0^t m(\eta_s)\left(\varphi_{m(\eta_s)}\left(P(s, T)(\phi)^2\right) - 2P(s, T)(\phi)\cdot Q_{m(\eta_s)}(P(s, T)(\phi))\right)\,ds,
\]

where the second equality holds because of the definition of carré-du-champ operator and (B.1).

Thus, using (20) we get

\[
N\left(\mathcal{M}(P(\cdot, T)(\phi))\right)_t = -\mathcal{M}(P(\cdot, T)(\phi)^2) - m(\eta_t)(P(t, T)(\phi)^2) + m(\eta_0)(P(0, T)(\phi)^2)
\]

\[
+ 2 \int_0^t m(\eta_s)(P(s, T)(\phi) \cdot \left[\varphi_{m(\eta_s)} - Q_{m(\eta_s)}\right](P(s, T)(\phi)))\,ds.
\]

Now, because of (19) and the boundedness conditions on \(V_\mu\) in Assumption (G1) we can ensure the existence of a constant \(C > 0\) such that

\[
N\left(\mathcal{M}(P(\cdot, T)(\phi))\right)_t \leq C(t + 1) - \mathcal{M}(P(\cdot, T)(\phi)^2).
\]

B.4. Proof of Lemma 2.5

First, by localisation, we can suppose that the martingales are bounded. Now, we will prove the inequalities for \(p = 2^q\), and then using the Jensen inequality we will extend the result for all \(p \geq 1\). The result for \(p \in (0, 1)\) is simply a consequence of the result for \(p = 1\) and the Jensen inequality for concave functions.

We want to prove the following inequalities:

\[
\mathbb{E}\left[\left(\mathcal{M}(P(\cdot, T)(\phi))\right)^{2^q}\right] \leq \frac{C(t + 1)^{2^q}}{N^{2^q}}, \quad \mathbb{E}\left[\left(\left|\mathcal{M}(P(\cdot, T)(\phi))\right|\right)^{2^q}\right] \leq \frac{C(t + 1)^{2^q}}{N^{2^q}}.
\]

For \(q = 0\), the first inequality is a consequence of Lemma 2.3 and the second one is due to the fact that \(\left\{\mathcal{M}(P(\cdot, T)(\phi))\right\}_{t \in [0,T]}\) is a local martingale.

We will prove the previous inequalities by induction. Let us assume they are true for \(q\) and lower. Thus, by Lemma 2.3 and the Minkowski inequality, there exists a \(K > 0\) such that

\[
I_p := \mathbb{E}\left[\left(\left|\mathcal{M}(P(\cdot, T)(\phi))\right|\right)^{p}\right]
\]

\[
\leq \mathbb{E}\left[\left[|K(t + 1)| + \left|\mathcal{M}(P(\cdot, T)(\phi)^2)\right|\right]^p\right]
\]

\[
\leq \left(K(t + 1) + \mathbb{E}\left[\left|\left|\mathcal{M}(P(\cdot, T)(\phi)^2)\right|\right|^{1/p}\right]^p\right).
\]
for all \( p \geq 1 \). Using now the BDG inequality, we get

\[
I_{2^{q+1}} \leq \left( K(t+1) + \kappa \left( \mathbb{E} \left[ \left( \mathcal{M}(P(.), T)(\phi)^2 \right)^2 \right] \right)^{1/2^{q+1}} \right)
\leq \left( K(t+1) + \kappa \sqrt{\frac{t+1}{N}} \right)^{2^{q+1}} \leq C'(t+1)^{2^{q+1}},
\]

where the second inequality holds by the induction hypothesis and the last one due to \( N \geq 1 \) and \( t + 1 \geq 1 \). Now, the martingale \( (\mathcal{M}_t(P(\cdot, T)(\phi)))_{t \in [0,T]} \) has jumps verifying

\[
a \leq 2 \frac{\|P(\cdot, T)(\phi)\|}{N} \leq \frac{2}{N}.
\]

Thus, using Lemma 2.4 we get

\[
\mathbb{E} \left[ \left( \mathcal{M}(P(\cdot, T)(\phi)) \right)^{2^{q+1}} \right] \leq C'' \sum_{k=0}^{q+1} \frac{\mathbb{E} \left[ \left( \mathcal{M}(P(\cdot, T)(\phi)) \right)^{2^k} \right]}{N^{2^{q+1}-2^{2^k}}} \leq C'' \sum_{k=0}^{q+1} \frac{(t+1)^{2^k}}{N^{2^{q+1}-2^{2^k} + 2^k}}
= \frac{C''}{N^{2^{q+1}}} \sum_{k=0}^{q+1} N(t+1)^{2^k} = \frac{C''(q+1)}{N^{2^{q+1}}} N^{2^{q+1}-(t+1)^{2^q+1}}
\leq \frac{C'(t+1)^{2^{q+1}}}{N^{2^{q+1}}}.
\]

This concludes the proof for \( p = 2^q \).

Now, for arbitrary \( p \), there exists \( q \) such that \( p \leq 2^q \). Thus, using the Jensen inequality (for the concave function \( x \mapsto x^{p/2^q} \)) we get

\[
\mathbb{E} \left[ \left( \mathcal{M}(P(\cdot, T)(\phi)) \right)^p \right] \leq \left( \mathbb{E} \left[ \left( \mathcal{M}(P(\cdot, T)(\phi)) \right)^{2^q} \right] \right)^{p/2^q} \leq C'(t+1)^{p/2^q} \leq C(\frac{t+1)^{2^q}}{N^{2^q}}) \leq \frac{C'(t+1)^p}{N^p}.
\]

The result for \( \mathbb{E} \left[ \left( \mathcal{M}(P(\cdot, T)(\phi)) \right)^p \right] \) is analogously obtained. \( \square \)

C. Proof of Theorems 1.5 and its corollaries

Obtaining a uniform in time bound as the one provided by Theorem 1.5 is a hard problem and this kind of results are uncommon in the literature. Del Moral and Guionnet in [49, Thm. 3.1] have proved a similar result for a resembling but discrete-time model, where the potential function \( \Lambda \) is assumed uniformly bounded and also bounded away from zero. Moreover, their upper bound for the speed of convergence is of order \( 1/N^\alpha \), with \( \alpha < 1/2 \). The same order of convergence \( 1/N^\alpha \) for a similar model is obtained in Theorem 2.11 in [18] for the control of the convergence in \( \mathcal{F} \)-norms, and not only for a single test function. Also, Theorem 14.3.7 in [26, § 14.3.3] provides uniform estimates for time-discretization models, see [26, § 12.2.3.1]. However, those estimates depend on the time mesh and are not bounded when the size of the time steps tend to zero. Also in the discrete-time settings, see Theorem 5.8 and Corollary 5.12 in [50], where a uniform in time bound with the optimal order \( 1/\sqrt{N} \) is proved. Roussset [23, Thm. 4.1] has proved a uniform in time bound in \( L^p \) with the same speed of convergence as our result. However, the model studied by Roussset is in continuous state space and the diffusion process driving the mutation process is assumed reversible. Similarly, Angeli et al. [25, Thm. 3.2] obtained an equivalent result for jump processes on locally compact spaces in the context of cloning algorithms and for \( p \geq 2 \). See also Theorem 5.10 and Corollary 5.12 in [21], for a related result when \( p = 2 \).

Our model is different, since we consider the case where the state space is countable and discrete, not necessarily finite and in Assumption (C1) we allow the selection rates to depend on the empirical probability measure induced by the particle system, in the same spirit of [23]. Nonetheless, our methods are similar to those of Roussset [23] and Angeli et al. [25] (see also [18, § 3.3.1]): it consists in finding a martingale indexed by the interval \([0,T]\), whose terminal value at time \( T \) is precisely \( m_{TT}(\phi) - \mu_T(\phi) \) plus a term whose \( L^p \) norm can be controlled, for
any $\phi \in B_b(E)$. Thereafter, the final result comes by a control of the quadratic variation of the martingale and an induction principle.

In the rest of this section, we will consider that Assumption (C1) is verified. Namely, the selection rates can be written as follows $V_\mu(x, y) = V^a_\mu(x) + V^b_\mu(y) + V^c_\mu(x, y)$, where $V^a_\mu$ is symmetric.

Let us denote by $\bar{m}(\eta)$ the mean empirical probability measure induced by $\eta$, which is defined as

$$\bar{m}(\eta) := \sum_{x \in E} \frac{\eta(x)}{N} \delta_x \in \mathcal{M}_1(E).$$

When $V^a_\mu$ and $V^b_\mu$ are nulls, and thus $V_\mu$ is symmetric, we obtain the following result as an immediate consequence of (25).

**Corollary C.1** ($V_\mu$ is symmetric). Assume that Assumption (C1) is verified in such a way that $V_\mu = V^c_\mu$ and Assumptions (I) and (C2) are also verified. Then the process

$$\left(m(\eta_t)(e^{(T-t)Q}(\phi)) - m(\bar{\eta}_t)(e^{TQ}(\phi))\right)_{t \in [0,T]},$$

is a local martingale, for every $\phi \in B_b(E)$. In particular, $\bar{m}(\eta_t) = \bar{m}(\eta_t)e^{tQ}$, for all $t \geq 0$.

**Proof of Corollary C.1.** Note that since the selection rates are symmetric, $\Lambda$ as defined in (10) is null. The proof simply follows as a consequence of (25), taking $\psi_t = e^{(T-t)Q}(\phi)$, for all $t \in [0,T]$ and $\phi \in B_b(E)$. □

Let us recall the operator $W_{t,T}$ defined in (16) as $W_{t,T} : \phi \mapsto P_{T-t}^{\Lambda}(\phi)/\mu_t(P_{T-t}^{\Lambda}(1))$. Recall that $\tilde{\phi}_T := \phi - \mu_T(\phi)$. We get $m(\eta_t)(W_{t,T}(\tilde{\phi}_T)) = m(\eta_t)(\phi) - \mu_T(\phi)$, which is the difference we intend to control. The following results establish a control of the uniform norm of $W_{t,T}$.

**Lemma C.2.** The operator $(W_{t,T})_{t \in [0,T]}$ verifies the following properties:

a) Given $p \geq 1$, for any test function $\phi \in B_b(E)$, there exists $C > 0$ such that

$$\|W_{t,T}(\phi)\| \leq C, \quad \text{and} \quad \int_t^T \|W_{s,T}(\phi)\|^p ds \leq C(T - t).$$

b) There exists a constant $\rho \in (0, 1)$, such that

$$\|W_{t,T}(\tilde{\phi}_T)\| \leq C\rho^{T-t}, \quad \text{and} \quad \int_t^T \|W_{s,T}(\tilde{\phi}_T)\|^p ds \leq C.$$

**Proof of Lemma C.2.** The proof of this result is inspired by the proof of Lemma 5.1 in [23], but we do not make any direct assumption of the spectrum of $Q + \Lambda$.

Note that $\mu_t(P_{T-t}^{\Lambda}(1)) = \mu_0(P_{T-t}^{\Lambda}(1))/\mu_0(P_{T-t}^{\Lambda}(1))$. Moreover, using Lemma 1.7, the function $t \mapsto e^{-\lambda t}\mu_0(P_{T-t}^{\Lambda}(1))$ is continuous and positive, going from 1 to $\mu_0(h) > 0$. This proves part a).

To prove part b) of the lemma, note that

$$\mu_T(\phi) = \frac{\mu_t P_{T-t}^{\Lambda}(\phi)}{\mu_t P_{T-t}^{\Lambda}(1)} \quad \text{and} \quad W_{t,T}(\mu_T(\phi)) = \mu_T(\phi)\frac{P_{T-t}^{\Lambda}(1)}{\mu_t P_{T-t}^{\Lambda}(1)},$$

since $\mu_T(\phi)$ is constant. Thus,

$$\|W_{t,T}(\tilde{\phi}_T)\| = \left\|\frac{P_{T-t}^{\Lambda}(\phi)}{\mu_t P_{T-t}^{\Lambda}(1)} - \mu_T(\phi)\frac{P_{T-t}^{\Lambda}(1)}{\mu_t P_{T-t}^{\Lambda}(1)}\right\|$$

$$= \left\|\frac{\mu_t P_{T-t}^{\Lambda}(1)}{\mu_0 P_{T-t}^{\Lambda}(1)} \cdot P_{T-t}^{\Lambda}(\phi) - \mu_T(\phi)\frac{P_{T-t}^{\Lambda}(1)}{\mu_t P_{T-t}^{\Lambda}(1)}\right\| \leq C\rho^{T-t},$$

where the last inequality is a consequence of the fact that the function $t \mapsto e^{-\lambda t}\mu_0(P_{T-t}^{\Lambda}(1))$ is bounded away from zero, and the uniform convergence of $e^{-\lambda t}P_t^{\Lambda}(\phi)$ towards $h_{\mu_0}(\phi)$, when $t \to \infty$, claimed in Lemma 1.7. □
Let $\psi$ be a function on $E \times \mathbb{R}_+$ as in the statement of Proposition 2.2. We denote by $(\mathcal{M}^\tau_t(\psi,\cdot))_{t\in[0,T]}$ the local martingale defined as $\mathcal{M}^\tau_t(\psi,\cdot) := \mathcal{M}_t(\psi,\cdot) - \mathcal{M}_s(\psi,\cdot)$, for all $t \in [0,T]$. We denote by $(\langle \mathcal{M}(\psi) \rangle^\tau_t)_{t\in[0,T]}$, the predictable quadratic variation of the local martingale $(\mathcal{M}^\tau_t(\psi,\cdot))_{t\in[0,T]}$.

Using (29) we can prove the next two results, analogously to Lemmas 2.3 and 2.5, establishing a control on the predictable quadratic variation and the quadratic variation of the martingale $(\mathcal{M}^\tau_t(\psi,\cdot))_{t\in[0,T]}$, respectively.

**Lemma C.3** (Control of the predictable quadratic variation). For every test function $\phi \in \mathcal{B}_1(E)$ and every $t \in [0,T]$ we have

$$N(\mathcal{M}(W,T)(\phi))_t^T \leq C(T-t+1) - \mathcal{M}^\tau_t(W,T)(\phi^2), \text{ for all } t \in [0,T],$$

and for $\hat{\phi}_T = \phi - \mu_T(\phi)$ we have

$$N(\mathcal{M}(W,T)(\hat{\phi}_T))_t^T \leq C - \mathcal{M}^\tau_t(W,T)(\hat{\phi}_T^2), \text{ for all } t \in [0,T].$$

**Lemma C.4** (Control of the quadratic variation). For all $p > 0$ and every test function $\phi \in \mathcal{B}_1(E)$ there exists a positive $C_p$ (possibly depending on $p$) such that

$$\mathbb{E} \left[ \left( \mathcal{M}(W,T)(\phi) \right)_t^T \right]^p \leq C_p(T-t+1)^p \mathbb{E} \left[ \mathcal{M}(W,T)(\hat{\phi}_T) \right]_t^T \leq C_p \left( \mathbb{E} \left[ \mathcal{M}(W,T)(\phi) \right]_t^T \right)^p,$$

and for a centred test function $\hat{\phi}_T = \phi - \mu_T(\phi)$:

$$\mathbb{E} \left[ \left( \mathcal{M}(W,T)(\hat{\phi}_T) \right)_t^T \right]^p \leq C_p \left( \mathbb{E} \left[ \mathcal{M}(W,T)(\phi) \right]_t^T \right)^p.$$

The proofs of Lemmas C.3 and C.4 are analogous to those of Lemmas 2.3 and 2.5, respectively. They are obtained using Lemma C.2 instead of (19). In particular, the second inequalities in both results are consequences of part b) of Lemma C.2. See also the proofs of Lemma 5.3 and Theorem 5.4 in [23]. We skip the proofs of Lemmas C.3 and C.4 for the sake of brevity.

Let us define the nonlinear propagator associated to $(\mu_t)_{t \geq 0}$ as follows

$$\Phi_{t,T}(v) := \frac{vP_T^\lambda}{\nu P_T^\lambda - (1)} \in \mathcal{M}_t(E).$$

By the semigroup property, $\Phi_{t,T}$ satisfies the propagation equation $\mu_T = \Phi_{t,T}(\mu_0)$. Using Assumption (C2) we can ensure the existence of $\rho \in (0,1)$ such that $\sup_{t \in \mathcal{M}_t(E)} \|\Phi_{t,T}(\nu) - \mu_\infty\|_{TV} \leq C \rho^{T-t}$.

Let us define

$$I_\rho(N) = \sup_{T \geq 0} \sup_{\phi \in \mathcal{B}_1(E)} \mathbb{E} \left[ (m(\eta_T)(\phi) - \mu_T(\phi))^p \right].$$

Our goal is to prove that $I_\rho(N) \leq C / N^{p/2}$. The method we use is similar to the one used by Rousset [23] and Angeli et al. [25]. Broadly speaking, it consists in an induction principle. First let us prove the following result providing the initial case of the induction.

**Lemma C.5** (Initial case). There exists $\epsilon > 0$ independent of $p$, such that

$$I_\rho(N) \leq \frac{C}{N^{p/2}}.$$

**Proof.** Fix $T > 0$ and consider

$$m(\eta_T)(\phi) - \mu_T(\phi) = m(\eta_T)(\phi) - \Phi_{t,T}(m(\eta_0))(\phi) + \Phi_{t,T}(m(\eta_0))(\phi) - \mu_T(\phi). \quad (C.1)$$

The idea is to control $a(t)$ using the stochastic error between $t$ and $T$, and $b(t)$ using the limiting stability. Moreover, $b(0)$ is controlled by the error made by the initial condition.

Let us first control the term $\mathbb{E}[\dot{a}(t)^p]$. Consider the finite variation process

$$A^\eta_{\rho} := \exp \left\{ \int_0^t m(\eta_s)(\Lambda) - \mu_\Lambda(\Lambda)ds \right\}.$$
Then,
\[
\partial_t \left( A_t^c m(\eta_t)(W_{s,t}(\phi)) \right) = A_t^c dM_t(W_{s,t}(\phi)).
\] (C.2)
Indeed, the martingale problem in Proposition 2.2, for the function \(W_{s,t}(\phi)\) yields
\[
d \left( m(\eta_t)(W_{s,t}(\phi)) \right) = dM_t(W_{s,T}(\phi)) + (\mu_t(\Lambda) - m(\eta_t)) m(\eta_t)(W_{s,t}(\phi)) dt.
\]
Hence,
\[
\partial_t \left( A_t^c m(\eta_t)(W_{s,t}(\phi)) \right) = \partial_t \left( A_t^c m(\eta_t)(W_{s,T}(\phi)) \right) + A_t^c d \left( m(\eta_t)(W_{s,T}(\phi)) \right),
\]
\[
= A_t^c (m(\eta_t)(\Lambda) - \mu_t(\Lambda)) m(\eta_t)(W_{s,t}(\phi)) + A_t^c d \left( m(\eta_t)(W_{s,T}(\phi)) \right)
\]
\[
= A_t^c dM_t(W_{s,t}(\phi)),
\]
where the last expression is obtained using (26).

Now, integrating from \(t\) to \(T\) in (C.2) and dividing by \(A_t^T\) we get
\[
m(\eta_T)(\phi) - (A_t^T)^{-1} m(\eta_t)(W_{s,t}(\phi)) = (A_t^T)^{-1} \int_t^T A_t^c dM_t(W_{s,t}(\phi)).
\]
Note that
\[
\Phi_{t,T}(m(\eta_t))(\phi) = \frac{(A_t^T)^{-1} m(\eta_t)(W_{s,t}(\phi))}{(A_t^T)^{-1} m(\eta_t)(W_{s,T}(1))}
\]
for all \(t \leq T\). Thus,
\[
a(t) = m(\eta_t)(\phi) - (A_t^T)^{-1} m(\eta_t)(W_{s,t}(\phi)) - \Phi_{t,T}(m(\eta_t))(\phi) \left[ 1 - (A_t^T)^{-1} m(\eta_t)(W_{s,T}(1)) \right]
\]
\[
= (A_t^T)^{-1} \int_t^T A_t^c dM_t(W_{s,t}(\phi)) - \Phi_{t,T}(m(\eta_t))(\phi)(A_t^T)^{-1} \int_t^T A_t^c dM_t(W_{s,T}(1)).
\]
Thus, we obtain the upper bound
\[
|a(t)| \leq (A_t^T)^{-1} \left( \left| \int_t^T A_t^c dM_t(W_{s,t}(\phi)) \right| + \left| \int_t^T A_t^c dM_t(W_{s,T}(1)) \right| \right).
\]
There exists a \(K > 0\) such that
\[
\mathbb{E} \left[ |a(t)|^p \right] \leq K e^{2p|\Lambda||T-t|} \sup_{\phi \in \mathcal{B}_1(E)} \mathbb{E} \left[ \left| \int_t^T A_t^c dM_t(W_{s,t}(\phi)) \right|^p \right]
\]
\[
\leq K e^{2p|\Lambda||T-t|} \sup_{\phi \in \mathcal{B}_1(E)} \mathbb{E} \left[ \left| \int_t^T (A_t^c)^2 d[M(W_{s,t}(\phi))] \right|^p \right],
\]
where the second inequality holds by the BDG inequality. Then, using Lemma C.4 we get
\[
\mathbb{E} \left[ |a(t)|^p \right] \leq K e^{4p|\Lambda||T-t|} \left( T - t + 1 \right)^{p/2} \left( N^{p/2} \right) \leq K \kappa^{p(T-t)} N^{p/2},
\]
where \(\kappa = e^{4|\Lambda|+1/2} > 1\).

Let us now control \(\mathbb{E} [ |b(t)|^p ]\). As a consequence of Assumption (C2) there exists a \(\rho \in (0, 1)\) and a \(C \geq 0\) such that
\[
\mathbb{E} [ |b(t)|^p ] \leq \mathbb{E} [ |\Phi_{0,T}(m(\eta_0))(\phi) - \Phi_{0,T}(\mu_t)(\phi)|^p ] \leq C \rho^{p(T-t)}.
\]
Now, for controlling \(b(0)\), note that
\[
b(0) = \Phi_{0,T}(m(\eta_0))(\phi) - \mu_T(\phi)
\]
\[
= m(\eta_0)(W_{0,T}(\phi)) - \mu_0(W_{0,T}(\phi)) + \Phi_{0,T}(m(\eta_0))(\phi) - m(\eta_0)(W_{0,T}(\phi))
\]
\[
= m(\eta_0)(W_{0,T}(\phi)) - \mu_0(W_{0,T}(\phi)) + \Phi_{0,T}(m(\eta_0))(\phi) \left[ 1 - m(\eta_0)(W_{0,T}(1)) \right].
\]
Thus, using Assumption (I) and the fact that \( \mu_t(W_{0,T}(\mathbf{1})) = 1 \), we get \( \mathbb{E}[|b(0)|^p] \leq C/N^{p/2} \). Let us now establish the global control optimising the choice of the argument \( t \) in (C.1). We have
\[
\mathbb{E}[|a(0) + b(0)|^p] \leq C\frac{pT + 1}{N^{p/2}},
\]
\[
\mathbb{E}[|a(t) + b(t)|^p] \leq C\frac{pT + 1}{N^{p/2}} + Cp^{p(T-\delta)},
\]
for all \( t \in [0, T] \).

The key idea now is to find a \( t_\epsilon \) such that \( \kappa' / N \) and \( \rho' \) are both equal to \( 1/N^\epsilon \). Let us take \( t_\epsilon = \frac{\ln N}{\ln \kappa} \) and \( \epsilon = \frac{\ln \rho}{\ln \kappa} \). Then, we have
\[
\frac{\kappa'}{N} = \exp[-\epsilon \ln N] = \frac{1}{N^\epsilon} \quad \text{and} \quad \rho' = \exp[-\epsilon \ln N] = \frac{1}{N^\epsilon}.
\]

We thus obtain the desired inequality:
\[
\mathbb{E}[|m(\eta_T) - \mu_T|] \leq \frac{C}{N^{p/2}}.
\]
Indeed, this inequality is obtained either from (C.3) when \( T \leq t_\epsilon / 2 \), since the expression in the upper bound is increasing in \( T \), or from (C.4) otherwise taking \( T - t = t_\epsilon / 2 \).

We proceed now to prove the induction step (equation (C.7) below), which, together with the initial case proved in Lemma C.5, concludes the proof of Theorem 1.5.

C.1. Proof of Theorem 1.5

Taking \( t = T \) in (26) reduces to
\[
M_T(W_{0,T} (\psi_T)) = m(\eta_T) (\psi_T) - m(\eta_0) (W_{0,T} (\psi_T)) - \int_0^T (\mu_s(\Lambda) m(\eta_s) (W_{s,T} (\psi_T)) \mathrm{d}s.
\]

Hence,
\[
|m(\eta_T) - \mu_T| \leq C \left( |m(\eta_0) (W_{0,T} (\psi_T))|^p + |M_T(W_{0,T} (\psi_T)))|^p + R_p \right),
\]
where
\[
R_p = \left( \int_0^T (\mu_s(\Lambda) m(\eta_s))(W_{s,T} (\psi_T)) \mathrm{d}s \right)^p.
\]

The initial error can be controlled using Assumption (I). Indeed,
\[
\mathbb{E}[|m(\eta_0)(W_{0,T} (\psi_T))|^p] = \mathbb{E}[|m(\eta_0)(W_{0,T} (\psi_T)) - \mu_T(\phi) + \mu_T(\phi) - \mu_T(\phi) m(\eta_0)(W_{0,T} (\psi_T))|^p]
\]
\[
\leq C \mathbb{E}[|m(\eta_0)(0) - \mu(0)(0)|^p] + C \mathbb{E}[|\mu_T(W_{0,T} (\psi_T))|^p]
\]
\[
\leq \frac{C}{N^{p/2}}.
\]

Furthermore, using Lemma C.4 and BDG inequality we get \( \mathbb{E}[|M_T(W_{0,T} (\psi_T))|^p] \leq C/N^{p/2} \). Note that \( m_s(W_{s,T} (\phi_T)) = 0 \). Using Hölder inequality, we obtain
\[
R_p = \left( \int_0^T (\mu_s(\Lambda) m(\eta_s))(W_{s,T} (\psi_T)) \right)^p
\]
\[
\leq \int_0^T (\mu_s(\Lambda) m(\eta_s))(W_{s,T} (\psi_T)) \left( \frac{\psi_{s,T}}{\|\psi_{s,T}\|} \right) \left( \psi_{s,T} \right) \|\psi_{s,T}\| |\psi_{s,T}| \|\psi_{s,T}\| \mathrm{d}s
\]
\[
\leq \kappa \int_0^T (\mu_s(\Lambda) m(\eta_s))(W_{s,T} (\psi_T)) \left( \frac{\psi_{s,T}}{\|\psi_{s,T}\|} \right) \left( \frac{\psi_{s,T}}{\|\psi_{s,T}\|} \right) \|\psi_{s,T}\| |\psi_{s,T}| \|\psi_{s,T}\| \mathrm{d}s,
\]

25
where \( \psi_{s,T} = W_{s,T}(\tilde{\phi}_T) \). Taking expectation and using the Cauchy–Schwarz inequality yield
\[
\mathbb{E} \left[ \left\| \int_0^T (\mu_s(\Lambda) - m(\eta_s(\Lambda)))m(\eta_s)(W_{s,T}(\tilde{\phi}_T))ds \right\|^p \right] \leq \kappa \int_0^T \| I_{2p}(N) \| W_{s,T}(\tilde{\phi}_T) \| ds 
\leq K I_{2p}(N). \tag{C.6}
\]
Thus, for every \( p \geq 1 \) we get the inequality
\[
I_p(N) \leq C \left( \frac{1}{N^{p/2}} + I_{2p}(N) \right), \tag{C.7}
\]
which using Lemma C.5 reduces to \( I_p(N) \leq C/N^{\min\{2p,1\}/2} \). By induction we obtain the bound \( I_p(N) \leq C/N^{p/2} \).

For a fixed \( N \), if the process \( (\eta_t)_{t \geq 0} \) generated by \( Q \) allows a stationary distribution \( \nu_N \), then under the hypothesis in Theorem 1.5 we get
\[
\sup_{\phi \in \mathcal{B}_b(E)} E_{\nu_N}[|m(\eta_\infty)(\phi) - \mu_\infty(\phi)|^p]^{1/p} \leq \frac{C_p}{\sqrt{N}}, \tag{C.8}
\]
for all \( p \geq 1 \).

We recall that \( \xi_i^{(t)} \) stands for the type of the \( i \)-th individual, for \( 1 \leq i \leq N \), at time \( t \geq 0 \). Let us denote by \( \text{Law}(\xi_i^{(t)}) \) the law of \( \xi_i^{(t)} \).

**Theorem C.6** (Bias estimate and ergodicity of one particle). Under Assumptions (I), (C1) and (C2), there exists a constant \( C > 0 \) such that
\[
\sup_{t \geq 0} \| \bar{m}(\eta_t) - \mu_t \|_{TV} \leq \frac{C}{N}.
\]
Moreover, if the initial distribution of the \( N \) particles is exchangeable, then
\[
\sup_{t \geq 0} \| \text{Law}(\xi_i^{(t)}) - \mu_t \|_{TV} \leq \frac{C}{N}.
\]

**Proof of Theorem C.6.** Taking expectation in (C.5) we get
\[
\mathbb{E} \left[ m(\eta_T)(\phi) - \mu_T(\phi) \right] = \int_0^T \mathbb{E} \left[ (\mu_s(\Lambda) - m(\eta_s(\Lambda)))m(\eta_s)(W_{s,T}(\tilde{\phi}_T)) \right] ds.
\]
Using Cauchy–Schwarz inequality we obtain
\[
|\mathbb{E} \left[ m(\eta_T)(\phi) - \mu_T(\phi) \right]| \leq C_1 \int_0^T I_{2}(N) \| W_{s,T}(\tilde{\phi}_T) \| ds \leq \frac{C}{N}.
\]

Now, assume that initially the \( N \) particles are sampled according to an exchangeable distribution. Note that
\[
\mathbb{E} \left[ \eta(x) \right] = \frac{1}{N} \sum_{i=1}^N \mathbb{P}[\xi_i = x] = \mathbb{P}[\xi_i = x], \quad \forall j \in \{1, \ldots, N\},
\]
where \( \xi_i \) denotes the position of the \( i \)-th particle of the process \( (\eta_t)_{t \geq 0} \) at time \( t \geq 0 \). Note that the second equality holds because of the exchangeability of the particles. Thus, as a consequence of the first part of the theorem and the previous equality we get \( \| \text{Law}(\xi_i^{(t)}) - \mu_t \|_{TV} \leq C/N \).