Convergence of generalized urn models to non-equilibrium attractors

Mathieu Faure[a,*], Sebastian J. Schreiber[b]

[a] *Aix-Marseille University (Aix-Marseille School of Economics), CNRS and EHESS, Centre de la vieille Charité, 2 rue de la Charité, 13002 Marseille, France*
[b] *Department of Evolution and Ecology, University of California, Davis, CA, 95616, USA*

Abstract

Generalized Polya urn models have been used to model the establishment dynamics of a small founding population consisting of k different genotypes or strategies. As population sizes get large, these population processes are well-approximated by a mean limit ordinary differential equation whose state space is the k simplex. We prove that if this mean limit ODE has an attractor at which the temporal averages of the population growth rate is positive, then there is a positive probability of the population not going extinct (i.e. growing without bound) and its distribution converging to the attractor. Conversely, when the temporal averages of the population growth rate are negative along this attractor, the population distribution does not converge to the attractor. For the stochastic analog of the replicator equations which can exhibit non-equilibrium dynamics, we show that verifying the conditions for convergence and non-convergence reduces to a simple algebraic problem. We also apply these results to selection–mutation dynamics to illustrate convergence to periodic solutions of these population genetics models with positive probability.
© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Biological invasions, where a species is introduced in a novel habitat, are occurring repeatedly throughout the world and often start with small founding population. Whether or not these

---

* Corresponding author.
 E-mail address: mathieu.faure@univ-amu.fr (M. Faure).

founding populations establish or go extinct in their new environment depends on a diversity of factors including local environmental conditions, demographic stochasticity, genetic diversity of the founding population, and nonlinear feedbacks between individuals in the founding population. One commonly used approach to understanding the roles of the first two factors is modeling the dynamics of establishment with branching processes [2]. This approach assumes that individuals survive, grow, and reproduce independently of one another and has provided fundamental insights into fixation of beneficial alleles [8], the build up of biodiversity on islands [11], the viability of endangered populations [15], and the evolution of disease emergence in novel host populations [1,12]. However, even when populations are at low abundance, individuals may interact with one another (e.g. finding or competing for mates in sexually reproducing populations) and thereby violate the assumption of independence of these classical branching processes. When these interactions occur between different types of individuals, they lead to frequency-dependent feedbacks on the population dynamics.

To account for these frequency-dependent interactions within a founding population, Schreiber [13] introduced a class of generalized urn models which were studied more extensively by Benaïm et al. [5]. These models consider an urn containing a finite number of balls (the population) of different colors (the different genotypes or phenotypes). At each stage, balls of possibly different colors can be added or removed from the urn, modeling deaths, births, or changes of state due to interactions between individuals.

Two key questions about these Markov processes are (a) when is there a positive probability that the population never goes extinct (i.e. the population establishes)? and (b) on the event of non-extinction, what can be said about the long-term frequency dynamics? To address these questions, Schreiber [13] introduced a mean limit ordinary differential equation (ODE) on the simplex (corresponding to all possible population frequencies) associated with the urn models. Using these mean field ODEs Benaïm et al. [5] proved: (i) if the population is expected to grow uniformly in the neighborhood of an attractor of this mean limit ODE, then with positive probability the population never goes extinct and the frequencies of the population converge to this attractor (see Theorem 2 next section); (ii) conversely if the population is expected to decrease uniformly in the neighborhood of a given set, then convergence toward this set occurs with probability zero (see Theorem 3 next section). As the expected growth rate of the population typically varies along non-equilibrium attractors, these two results, however, are most useful for equilibrium attractors of the mean limit ODE.

Here, we study the case where the underlying mean limit ODE admits non-equilibrium attractors with non-constant growth rates. As we show in the applications section, this case arises quite naturally in stochastic models for evolutionary games and population genetics. We extend the results of Benaïm et al. [5] to a more general framework (see Theorems 4 and 5 in Section 3). Most notably, we replace the assumption of uniform positive (respectively, negative) population growth near the attractor with the assumption that the temporal average of the population growth rate is positive (respectively, negative) for initial conditions near the attractor (see assumptions (7) and (8)).

The remainder of this paper is organized as follows. In next section, we define the class of generalized urn models, and recall the main results of Benaïm et al. [5]. We also discuss the stochastic approximation methods, and briefly explain how they were used to derive these results. In Section 3, we state and prove our main results: convergence with positive probability toward an attractor with average positive growth and non-convergence to an invariant set with negative average growth. Section 4 is devoted to applications to evolutionary games and population genetics. The proofs of some technical estimates are given in the Appendix.

2. Generalized urn models

In Section 2.1 we give the definitions of a class of generalized urn models introduced by Schreiber [13]. The approach used to study these models is the so-called ODE method, which relates the asymptotic behavior of a stochastic difference equation to an ODE. This method is described in Section 2.2.

2.1. The urn models

We consider a finite population consisting of individuals that are one of k types. Therefore, the state space for this Markov chain is the non-negative cone

$$\mathbb{Z}_+^k = \{z = (z^1, \ldots, z^k) \in \mathbb{Z}^k : z^i \geq 0 \text{ for all } i\}$$

where $\mathbb{Z}^k$ is the space of k-tuples of integers. Given a vector $w = (w^1, \ldots, w^k) \in \mathbb{Z}^k$ define

$$|w| = |w^1| + \cdots + |w^k| \quad \text{and} \quad \alpha(w) = w^1 + \cdots + w^k.$$

We let $\| \cdot \|$ denote the Euclidean norm on $\mathbb{R}^k$.

Let $z(n) = (z^1(n), \ldots, z^k(n))$ be a homogeneous Markov chain with state space $\mathbb{Z}_+^k$. In our context, $z^i(n)$ corresponds to the number of individuals of type i at the nth update. Associated with $z(n)$ is the random process $x(n)$ defined by

$$x(n) = \begin{cases} \dfrac{z(n)}{|z(n)|} & \text{if } z(n) \neq 0 \\ 0 & \text{if } z(n) = 0 \end{cases}$$

which is the distribution of types in the population at the nth update. When there are no individuals in the population at the nth update, we arbitrarily set $x(n)$ to zero which we view as the "null" distribution. Provided, $z(n)$ is non-zero, the population distribution $x(n)$ lies on the unit simplex

$$S_k = \left\{ x = (x^1, \ldots, x^k) \in \mathbb{R}^k : x^i \geq 0, \sum_{i=1}^k x^i = 1 \right\}.$$

We let $\Pi : \mathbb{Z}_+^k \times \mathbb{Z}_+^k \mapsto [0, 1]$ denote the transition kernel of the Markov chain $z(n)$. Specifically,

$$\Pi(z, z') = \mathbb{P}[z(n+1) = z' | z(n) = z].$$

Our standing assumptions on the Markov chains $z(n)$ are as follows:

(A1) At each update, there is a maximal number of individuals that can be added or removed. In other words, there exists a positive integer m such that $|z(n+1) - z(n)| \leq m$ for all n.

(A2) There exist Lipschitz maps

$$\{p_w : S_k \to [0, 1] \ : w \in \mathbb{Z}^k, \ |w| \leq m\}$$

and a real number $a > 0$ such that

$$|p_w(z/|z|) - \Pi(z, z + w)| \leq a/|z|$$

for all non-zero $z \in \mathbb{Z}_+^k$ and $w \in \mathbb{Z}^k$ with $|w| \leq m$.

Assumption (A2) implies that, as the population gets large, the transition probabilities tend to only depend on the frequency vector $z/|z|$.

2.2. Mean limit ODEs

The following lemma which was proved by Benaïm et al. [5] expresses the random process $(x(n))_n$ as a stochastic approximation algorithm.

Lemma 1. *Let $z(n)$ be a Markov chain on $\mathbb{Z}_+^k$ satisfying assumptions* (A1) *and* (A2) *with mean limit transition probabilities $p_w : S_k \to [0, 1]$. Let $\mathcal{F}_n$ denote the σ-field generated by $\{z(0), z(1), \ldots, z(n)\}$. There exist sequences of random variables $\{U_n\}$ and $\{b_n\}$ adapted to $\mathcal{F}_n$, and a real number $K > 0$ such that*

(i) *if $z(n) \neq 0$, then*

$$x(n + 1) - x(n) = \frac{1}{|z(n)|} \left(\sum_{w \in \mathbb{Z}^k} p_w(x(n))(w - x(n)\alpha(w)) + U_{n+1} + b_{n+1} \right). \quad (1)$$

(ii) $\mathbb{E}[U_{n+1}|z(n)] = 0.$
(iii) *The random variables $\|U_n\|$ and $\mathbb{E}[\|U_{n+1}\|^2|\mathcal{F}_n]$ are uniformly bounded.*
(iv) $\|b_{n+1}\| \leq \frac{K}{\max\{1, |z(n)|\}}.$

The recurrence relationship (1) can be viewed as a "noisy" Cauchy–Euler approximation scheme with step size $1/|z(n)|$ for solving the ordinary differential equation

$$\frac{dx}{dt} = g(x) := \sum_{w \in \mathbb{Z}^k} p_w(x)(w - x\alpha(w)). \quad (2)$$

The limiting behavior of the $x(n)$ is therefore related to the solutions of (2). Indeed, when the number of individuals in the population grows without bound, the step size decreases to zero and it seems reasonable that there is a strong relationship between the limiting behavior of the mean limit ODE and the population distribution $x(n)$. To make the relationship between the stochastic process $x(n)$ and the mean limit ODE more transparent, it is useful to define a continuous time version of $x(n)$ where time is scaled in an appropriate manner. Since the number of events (updates) that occur in a given time interval is likely to be proportional to the size of the population, we define *the time $\tau(n)$ that has elapsed by update n* as

$$\tau(0) = 0$$

$$\tau(n + 1) = \begin{cases} \tau(n) + \dfrac{1}{|z(n)|} & \text{if } z(n) \neq 0 \\ \tau(n) + 1 & \text{if } z(n) = 0. \end{cases}$$

The continuous time version of $x(n)$ is given by

$$X(t) = x(n) \quad \text{for } \tau(n) \leq t < \tau(n + 1). \quad (3)$$

To relate the limiting behavior of the solutions of (2) to the limiting behavior of $X(t)$, we need a few more definitions to state a key result of Schreiber [13]. Let $x.t$ denote the solution of (2) with initial condition x, at time t. A set C is called *invariant* for (2) provided that $C.t = C$ (where $C.t = \{x.t : x \in C\}$) for all $t \in \mathbb{R}$. A compact invariant set $\mathcal{A} \subset S_k$ is an *attractor* (in the sense of Conley, see [7]) if there is an open neighborhood $U \subset S_k$ of $\mathcal{A}$ such that

$$\cap_{t > 0} \overline{\cup_{s \geq t} U.s} = \mathcal{A}.$$

The basin of attraction $B(\mathcal{A})$ of $\mathcal{A}$ is the set of points $x \in S_k$ satisfying $\inf_{y \in \mathcal{A}} \|x.t - y\| \to 0$ as $t \to \infty$. Finally, a compact invariant set C is *internally chain transitive* provided that (2) restricted to C admits no proper attractor.

Given a function $X : \mathbb{R}_+ \to \mathbb{R}^k$ or a sequence $\{x(n)\}_{n \geq 0}$ in $\mathbb{R}^k$, we define the *limit sets*, $L(X(t))$ and $L(x(n))$, of $X(t)$ and $x(n)$ as follows. $L(X(t))$ is the set of $p \in \mathbb{R}^k$ such that $\lim_{k \to \infty} X(t_k) = p$ for some subsequence $\{t_k\}_{k \geq 0}$ with $\lim_{k \to \infty} t_k = \infty$. $L(x(n))$ is the set of $p \in \mathbb{R}^k$ such that $\lim_{k \to \infty} x(n_k) = p$ for some subsequence $\{n_k\}_{k \geq 0}$ with $\lim_{k \to \infty} n_k = \infty$.

Using the methods of Benaïm [3,4], the following result of Schreiber [13] demonstrates the relationship between the asymptotic behavior of $X(t)$ and $x.t$ on the event that the population is growing sufficiently rapid.

Theorem 1 (*Schreiber [13]*). *Let $z(n)$ be a Markov process satisfying the assumptions of Lemma 1. Then, on the event*

$$\left\{ \sum_n \frac{1}{|z(n)|^{1+a}} < \infty, \text{ for some } a > 0 \right\},$$

1. *the interpolated process $X(t)$ is almost surely an* asymptotic pseudotrajectory *for the flow of the mean limit ODE (2). In other words, $X(t)$ almost surely satisfies*

$$\lim_{t \to \infty} \sup_{0 \leq h \leq T} \|X(t).h - X(t + h)\| = 0 \tag{4}$$

   for any $T > 0$.
2. *the limit set $L(X(t))$ of $X(t)$ is almost surely an internally chain transitive set for the mean limit ODE.*

The first assertion of the theorem roughly states that $X(t)$ tracks the solutions of the mean limit ODE (2), with increasing accuracy far into the future. The second assertion of the theorem states that the only candidates for limit sets of the population distribution $x(n)$ are connected compact internally chain recurrent sets for the mean limit ODE. With regards to attractors, we have the following useful property of internally chain recurrent sets (see, e.g., [4, Cor. 5.4]).

Remark 1. If an internally chain transitive set C meets the basin of attraction of a given attractor $\mathcal{A}$ then it is contained in $\mathcal{A}$.

Consider an attractor $\mathcal{A}$ for the mean limit ODE. Now suppose that there is a neighborhood $U \subset B(\mathcal{A})$ of $\mathcal{A}$ such that whenever the population is large and its distribution $x(n)$ lies in U, the population is expected to grow. Specifically, $f(x(n)) > 0$ whenever $x(n) \in U$ where

$$f(x) = \sum_w p_w(x)\alpha(w)$$

is the limiting expected change in the population size. Under such circumstances, we would expect the population to increase in size and, consequently, the frequencies to follow the solutions of the mean limit ODE more closely. As $x(n)$ lies in the basin of attraction of $\mathcal{A}$, $x(n)$ would tend to remain near $\mathcal{A}$ and the population would be expected to increase further. One would expect that this positive feedback loop would result in the population growing with positive probability and its distribution converging to the attractor $\mathcal{A}$. Indeed, as the next theorem shows, this argument holds when the attractor $\mathcal{A}$ is *attainable*, which basically means that the random process can reach the basin of attraction of $\mathcal{A}$, at any time. More specifically, we define the set of *attainable points*,

$Att_\infty(X)$, as the set of points $x \in S_k$ such that, for all $M \in \mathbb{N}$ and every open neighborhood U of x

$$\mathbb{P}[|z(n)| \geq M \text{ and } x(n) \in U \text{ for some } n] > 0.$$

Theorem 2 (*Benaïm, Schreiber and Tarres [5]*). *Let $z(n)$ be a generalized urn process verifying* (A1) *and* (A2). *Assume that*

$$\lambda = \inf_{x \in \mathcal{A}} f(x) > 0. \tag{5}$$

If $B(\mathcal{A}) \cap Att_\infty(X) \neq \emptyset$, then

$$\mathbb{P}\left[\liminf_n \frac{|z(n)|}{n} \geq \lambda \text{ and } L(x(n)) \subset \mathcal{A}\right] > 0.$$

The next theorem provides a partial converse to Theorem 2. Roughly, it states that if there is a compact set $K \subset S_k$ near which the population is expected to decrease every update, then the population distribution $x(n)$ cannot converge to K.

Theorem 3 (*Benaïm, Schreiber and Tarres [5]*). *Let $z(n)$ be a generalized urn process verifying* (A1) *and* (A2) *and $K \subset S_k$ be any compact set. Assume that*

$$\sup_{x \in K} f(x) < 0. \tag{6}$$

Then there exists $M > 0$ such that

$$\mathbb{P}\left(|z(n)| > M \text{ for } n \text{ large enough and } L(x(n)) \subset K\right) = 0.$$

3. Main results

While the assumption of uniform growth is not restrictive for equilibrium attractors of the mean limit ODE, the long-term behavior of these mean limit ODEs may be governed by non-equilibrium behavior, such as limit cycles, quasi-periodic motions, or chaotic attractors. For these types of attractors, the growth rate $f(x)$ of a population typically varies along points of the attractor and, consequently, the uniform growth assumptions (5) and (6) are too restrictive. Throughout the section, $\mathcal{A}$ denotes an attractor for the mean limit ODE, with basin of attraction $B(\mathcal{A})$. Rather than a uniform growth assumption, we assume that the long-term temporal average of the growth rate f is positive along orbits of the mean limit ODE. Specifically,

$$\liminf_{t \to +\infty} \frac{1}{t} \int_0^t f(x.s)ds > 0 \quad \text{for all } x \in B(\mathcal{A}). \tag{7}$$

Our main result is a generalization of Theorem 2, under this less restrictive positive growth assumption (7). This result states, roughly, that if the basin of attraction is attainable and the temporal averages of the growth rate are positive along the attractor, then the population grows without bound and its distribution converges to the attractor with positive probability.

Theorem 4 (*Positive Average Growth and Convergence*). *Let $z(n)$ be a generalized urn process satisfying* (A1) *and* (A2). *Assume that* (7) *holds and*

$$B(\mathcal{A}) \cap Att_\infty(X) \neq \emptyset,$$

then

$$\mathbb{P}\left[\sum_n \frac{1}{|z(n)|^{1+\delta}} < +\infty \;\forall \delta > 0, \;\; and \;\; L(x(n)) \subset \mathcal{A}\right] > 0.$$

Unlike Theorem 2, Theorem 4 can no longer guarantee linear population growth with positive probability. However, this is not surprising, as the population growth rate can be negative as well as positive despite having a positive temporal average.

We have a partial converse result to Theorem 4. We say that there is an average negative growth rate in an invariant compact set K if

$$\limsup_{t\to+\infty} \frac{1}{t}\int_0^t f(x.s)ds < 0 \quad \text{for all } x \in K. \tag{8}$$

Theorem 5 (*Negative Average Growth and Non-Convergence*). *Let $z(n)$ be a generalized urn process satisfying* (A1) *and* (A2). *Assume that* (8) *holds for a compact invariant set K. Then there exists $M > 0$ such that*

$$\mathbb{P}\left[|z(n)| \geq M \text{ for } n \text{ large enough and } L(x(n)) \subset K\right] = 0.$$

The proofs of Theorems 4 and 5 are given in Sections 3.2 and 3.3, respectively. Several key technical estimates required for these proofs are described in 3.1 and proven in the Appendices.

3.1. Key estimates

To state the key estimates for the proofs of our main results, call L_f the Lipschitz constant of f and $\|f\|_\infty := \sup_x f(x)$. The map $s \mapsto f(x.s)$ is Lipschitz, uniformly in $x \in K$. Consequently, there exists a constant $L' > 0$ such that

$$\|f(x.s) - f(x.s')\| \leq L'|s - s'| \quad \text{for all } x \in K.$$

Recall that $g(x) = \sum_w p_w(x)(w - x\alpha(w))$. Let L_g be the Lipschitz constant for g and $\|g\|_\infty = \sup \|g(x)\|$. Define $L = \max\{L_f, L_g, L', \|g\|_\infty, \|f\|_\infty\}$.

Now assume that $(x(n))_n$ is a stochastic approximation process which satisfies Eq. (1). Define $\|U\| = \sup_n \|U_n\|$. To simplify the presentation of the proof, we assume that $\|b_n\| = 0$. The proof without this assumption is notationally more cumbersome but follows in a nearly identical fashion.

Define $m(t) = \inf\{k : \tau(k) \geq t\}$. Let $T_0 > 0$. We first observe that the population size $|z(r)|$ remains bounded on time intervals of order T_0. Given $r_0 \in \mathbb{N}$, define $r_{k+1} = m(\tau(r_k) + T_0)$ for all $k \geq 1$. Provided $|z(r_k)|$ is large enough, Benaïm et al. [5, Lemma 2] state that

(i) $B|z(r_k)| \geq |z(r)| \geq B^{-1}|z(r_k)|$ for all $r \in [r_k, r_{k+1}]$, and
(ii) $T_0 B^{-1}|z(r_k)| \leq r_{k+1} - r_k \leq T_0 B|z(r_k)|$

where $B = 3e^{mT_0}$. Using Gronwall's inequality, Benaïm [4, Proposition 4.1] proved the following estimate

$$\sup_{r\in[r_k,r_{k+1}]} \|x(r_k).(\tau(r) - \tau(r_k)) - x(r)\| \leq C\left(\Gamma_1(k, T_0) + \Gamma_2(k, T_0)\right) \tag{9}$$

where C is a positive constant, which depends on T_0 and L,

$$\Gamma_1(k, T_0) = \sup_{r \in [r_k, r_{k+1} - 1]} \left| \sum_{i=r_k}^{r} \frac{U(i+1)}{|z(i)|} \right|$$

and[1]

$$\Gamma_2(k, T_0) = \frac{2\|g\|_\infty}{\displaystyle\inf_{r \in [r_k, r_{k+1}]} |z(r)|}.$$

The next lemma refines the statement of the asymptotic pseudotrajectory property (4) for the discrete time process $x(n)$. The proof is given in the Appendix.

Lemma 2. *Let T_0 and δ be positive real numbers. Then we have*

$$\mathbb{P}\left(\sup_{r \in [r_k, r_{k+1}]} \|x(r_k).(\tau(r) - \tau(r_k)) - x(r)\| > \delta \,\big|\, \mathcal{F}_{r_k} \right) \le \frac{C_0(T_0)}{|z(r_k)|\delta^2}, \tag{10}$$

(where $C_0(T_0) := 4\|U\|^2 BC^2 T_0$) on the event $\left\{ |z(r_k)| \ge \frac{4BCL}{\delta} \right\}$.

The next two propositions roughly underestimate the likelihood that the average growth $\frac{1}{T_0} \int_0^{T_0} f(x(r).s)ds$ remains close of its stochastic counterpart on intervals of time of length T_0, provided the population size is initially large enough. Both proofs are given in the Appendix.

Proposition 1. *Let $(x(n))_n$ be a stochastic process satisfying (1). Given a point $y \in S_k$, a time $T_0 > 0$, and $\delta > 0$, there exist a compact neighborhood U of y, $C_1(T_0) > 0$, and $M_1 > 0$ such that*

$$\mathbb{P}\left[\frac{1}{T_0} \left| \sum_{r=r_k}^{r_{k+1}-1} \frac{1}{|z(r)|} f(x(r)) - \int_0^{T_0} f(y.s)ds \right| > \delta \,\Big|\, \mathcal{F}_{r_k} \right] \le \frac{C_1(T_0)}{|z(r_k)|\delta^2} \tag{11}$$

on the event $V_k := \{x(r_k) \in U\} \cap \{|z(r_k)| \ge M_1\}$.

Using the estimate from this proposition, we can get the following result.

Proposition 2. *Let $(x(n))_n$ be a stochastic process satisfying (1). Given a point $y \in S_k$, a time $T_0 > 0$, and $\delta > 0$, there exist a compact neighborhood U of y, $C_2(T_0) > 0$, and $M_2 > 0$ such that*

$$\mathbb{P}\left[\frac{1}{T_0} \left| \sum_{i=r_k}^{r_{k+1}-1} \frac{|z(i+1)| - |z(i)|}{|z(i)|} - \int_0^{T_0} f(x(r_k).s)ds \right| \ge \delta \,\Big|\, \mathcal{F}_{r_k} \right] \le \frac{C_2(T_0)}{|z(r_k)|\delta^2}$$

on the event $W_k := \{x(r_k) \in U\} \cap \{|z(r_k)| \ge M_2\}$.

---

[1] In the more general case where $\|b\| \ne 0$, there is an additional term in Γ_2, namely $\sup_{r \in [r_k, r_{k+1}-1]} \left| \sum_{r_k}^{r} \frac{b(i+1)}{|z(i)|} \right|$.

3.2. *Proof of Theorem 4*

Pick $p \in B(\mathcal{A}) \cap Att_\infty(X)$ and an open neighborhood U of A, which contains p and whose closure K is compact and included in $B(\mathcal{A})$. Since $\mathcal{A}$ is an attractor, there exist a positive time T' and $\delta > 0$ such that, for any $T \geq T'$,

$$X(t) \in U \quad \text{and} \quad \|X(t+T) - X(t).T\| < \delta \Rightarrow X(t+T) \in U.$$

Also, by assumption (7), there exist $a_1 > 0$ and $T'' > 0$ such that, for any $T \geq T''$ and $x \in K$,

$$\frac{1}{T} \int_0^T f(x.s)ds \geq a_1.$$

Define $T_0 = \max\{T', T''\}$.

Let $M > 0$ be fixed (we will need M to be greater than some quantity which depends on T_0, m, a_1 and δ in a manner that will be specified later in the proof). By the attainability condition, there exists an index $r_0 \in \mathbb{N}$ such that

$$\mathbb{P}[x(r_0) \in U, \ |z(r_0)| \geq M] > 0.$$

Consider the following events for all $k \geq 1$

$$E_1(k) = \{|z(r_k)| \geq \zeta^{k-1} B^{-1} M\}; \qquad E_2(k) = \{x(r) \in U, \ \text{for all } r \in [r_k, r_{k+1}]\}$$

where $\zeta = 1 + a_1 T_0/4$. Let $E(0)$ be the event $\{|z(r_0)| \geq M, x(r_0) \in U\}$. For $k \geq 1$, define $E(k) = E(k-1) \cap E_1(k) \cap E_2(k)$. We will show that there exists a constant $F > 0$ such that

$$\mathbb{P}[E(k+1)|E(k)] \geq 1 - F/M\zeta^k \quad \text{for all } k \geq 0. \tag{12}$$

The proof of estimate (12) relies on two ingredients. First, Lemma 3 provides a lower bound to the probability of being inside U on $[r_{k+1}, r_{k+2}]$ if $x(r_k) \in U$. Second, Lemma 4 underestimates the probability that the population grows sufficiently (namely by a multiplicative parameter $\zeta > 1$) between times r_k and r_{k+1}, provided it stayed in U the entire time. Unlike the work of Benaïm et al. [5], the main issue here is that we do not have expected growth for every update of the population in the neighborhood of the attractor, so we need to make use of Proposition 2 and the estimates on the population size given above.

Lemma 3. *There exist $M_0' > 0$ and $D > 0$ such that if $M \geq M_0'$, then*

$$P[E_2(k+1)|E(k)] \geq 1 - \frac{D}{M\zeta^{k-1}}$$

for all $k \geq 0$.

Proof. Statement (i) in Section 3.1 implies that $|z(r)| \geq B^{-2}|z(r_k)|$ for $r \in [r_k, r_{k+2}]$. Furthermore, the definition of r_k implies that $T_0 \leq \tau(r_{k+1}) - \tau(r_k) \leq 2T_0$ which implies $4T_0 \geq \tau(r) - \tau(r_k) \geq T_0$ for $r \in [r_{k+1}, r_{k+2}]$. Statements (i) and (ii) in Section 3.1 also imply that $r_{k+2} - r_k \leq 2B^2 T_0|z(r_k)|$ and $r_{k+2} \leq m(\tau(r_k) + 3T_0)$.

On the event $E_1(k)$, $|z(r)| \geq |z(r_k)|B^{-2} \geq \zeta^{k-1}B^{-3}M$ for all $r \in [r_k, r_{k+2}]$. Therefore, we can apply Lemma 2 with $3T_0$ and $|z(r_k)| \geq \zeta^{k-1}B^{-3}M$ which gives, choosing M_0' large enough and M greater than M_0',

$$\mathbb{P}\left[\sup_{r \in [r_{k+1}, r_{k+2}]} \|x(r_k).(\tau(r) - \tau(r_k)) - x(r)\| > \delta \mid \mathcal{F}_{r_k}\right] \leq \frac{B^3 C_0(3T_0)}{\delta^2 M \zeta^{k-1}}.$$

Since $\tau(r_{k+1}) - \tau(r_k) \geq T_0$ and $x(r_k) \in U$, our choice of T_0 and δ implies that $x(r) \in U$ for all $r \in [r_{k+1}, r_{k+2}]$, on the event

$$\left\{ \sup_{r \in [r_{k+1}, r_{k+2}]} \|x(r_k).(\tau(r) - \tau(r_k)) - x(r)\| \leq \delta \right\}.$$

Defining $D = B^3 C_0(3T_0)/\delta^2$, we therefore have

$$\mathbb{P}[E_2(k+1)^c \mid E(k)] \leq \frac{D}{M\zeta^{k-1}}. \quad \blacksquare$$

Lemma 4. *There exist $M_0'' \geq M_0'$ and $D' > 0$ such that if $M \geq M_0''$, then*

$$P[E_1(k+1) \mid E(k)] \geq 1 - \frac{D'}{M\zeta^{k-1}}$$

for all $k \geq 1$.

Proof. On the event $E_1(k+1)^c \cap E(k)$, we have

$$n_k := |z(r_{k+1})| - |z(r_k)| \leq \zeta^k B^{-1} M - \zeta^{k-1} B^{-1} M = B^{-1} M\zeta^{k-1}(a_1 T_0/4).$$

Under the constraint that one cannot add more balls than this fixed quantity between times r_k and r_{k+1}, we prove the following inequality in the [Appendix](#)

$$\sum_{i=r_k}^{r_{k+1}-1} \frac{|z(i+1)| - |z(i)|}{|z(i)|}$$

$$\leq \frac{n_k}{|z(r_k)|} + \frac{m}{2}(r_{k+1} - r_k) \left(\frac{1}{B^{-1}|z(r_k)|} - \frac{1}{B^{-1}|z(r_k)| + m} \right). \tag{13}$$

Consequently, since $r_{k+1} - r_k \leq T_0 B|z(r_k)|$,

$$\sum_{i=r_k}^{r_{k+1}-1} \frac{|z(i+1)| - |z(i)|}{|z(i)|} \leq B^{-1} M\zeta^{k-1}(a_1 T_0/4)\frac{1}{|z(r_k)|} + \frac{T_0 B|z(r_k)|m^2}{2(B^{-2}|z(r_k)|^2)}$$

$$\leq \frac{a_1 T_0}{4} + \frac{T_0 B^3 m^2}{2M}.$$

Hence, choosing M greater than $M_0'' := \max\{M_0', 2T_0 B^3 m^2/a_1\}$,

$$\frac{1}{T_0} \sum_{i=r_k}^{r_{k+1}-1} \frac{|z(i+1)| - |z(i)|}{|z(i)|} < \frac{a_1}{2}.$$

Recall that, by definition of T_0,

$$\frac{1}{T_0} \int_0^{T_0} f(x(r_k).s)ds \geq a_1.$$

Hence, on the event $E_1(k+1)^c \cap E(k)$, we have

$$\frac{1}{T_0} \left| \sum_{i=r_k}^{r_{k+1}-1} \frac{|z(i+1)| - |z(i)|}{|z(i)|} - \int_0^{T_0} f(x(r_k).s)ds \right| \geq \frac{a_1}{2}.$$

Finally, by [Proposition 2](#) and provided M_0'' is large enough, we have

$$\mathbb{P}\left[E_1(k+1)^c \mid E(k)\right] \leq \frac{4C_2(T_0)}{a_1^2 |z(r_k)|}.$$

The proof is complete, taking $D' = \frac{4C_2(T_0)B}{a_1^2}$. ∎

Now choose M larger than M_0''. By [Lemmas 3](#) and [4](#), and denoting $F = D + D' > 0$, we have

$$\mathbb{P}[E(k+1)|E(k)] \geq 1 - \frac{F}{M\zeta^{k-1}}$$

for all $k \geq 0$. Since the sequence of event $\{E(k)\}_k$ is decreasing, it follows that

$$\mathbb{P}\left[\lim_{k\to\infty} E(k)\right] \geq \mathbb{P}[E(0)]\left(1 - \sum_{k=1}^{\infty} \frac{F}{M\zeta^{k-1}}\right)$$

$$\geq \mathbb{P}[E(0)]\left(1 - \frac{\zeta F}{M(\zeta-1)}\right).$$

On the event $\lim_{k\to\infty} E(k)$, we have for any $\delta > 0$

$$\sum_{i=1}^{+\infty} \frac{1}{|z(i)|^{1+\delta}} = \sum_{i=1}^{r_0} \frac{1}{|z(i)|^{1+\delta}} + \sum_{k=1}^{\infty} \sum_{i=r_k+1}^{r_{k+1}} \frac{1}{|z(i)|^{1+\delta}}$$

$$\leq \sum_{i=1}^{r_0} \frac{1}{|z(i)|^{1+\delta}} + \sum_{k=1}^{\infty} \left(\max_{r_k+1\leq i\leq r_{k+1}} \frac{1}{|z(i)|^{\delta}} \times \sum_{i=r_k+1}^{r_{k+1}} \frac{1}{|z(i)|}\right)$$

$$\leq \sum_{i=1}^{r_0} \frac{1}{|z(i)|^{1+\delta}} + T_0 \sum_{k=1}^{\infty} \max_{r_k+1\leq i\leq r_{k+1}} \frac{1}{|z(i)|^{\delta}}$$

$$\leq \sum_{i=1}^{r_0} \frac{1}{|z(i)|^{1+\delta}} + T_0 \sum_{k=1}^{\infty} \frac{1}{(B^{-1}|z(r_k)|)^{\delta}}$$

$$\leq \sum_{i=1}^{r_0} \frac{1}{|z(i)|^{1+\delta}} + T_0 \sum_{k=1}^{\infty} \frac{1}{(\zeta^{k-1}B^{-2}M)^{\delta}}$$

$$= \sum_{i=1}^{r_0} \frac{1}{|z(i)|^{1+\delta}} + T_0 B^{2\delta} M^{-\delta} \frac{\zeta^{\delta}}{\zeta^{\delta}-1} < \infty.$$

Hence the definition of $E_2(k)$ implies that

$$\mathbb{P}[\mathcal{C}] \geq \mathbb{P}[E(0)]\left(1 - \frac{\zeta F}{M(\zeta-1)}\right) > 0,$$

where

$$\mathcal{C} = \left\{\sum_{i=1}^{+\infty} \frac{1}{|z(i)|^{1+\delta}} < \infty, \ \forall \delta > 0 \text{ and } x(n) \in U \ \forall n\right\}.$$

On the event $\mathcal{C}$, [Theorem 1](#) implies that $L(\{x(n)\})$ is a compact internally chain recurrent set for the mean limit ODE. Since $L(\{x(n)\}) \subset K \subset B(\mathcal{A})$ on the event $\mathcal{C}$, [Remark 1](#) and [Theorem 1](#) imply that $L(\{x(n)\}) \subset \mathcal{A}$. □

3.3. Proof of Theorem 5

By assumption (8), there exist an open neighborhood U of K, $T_0 > 0$ and $a_1 > 0$ such that

$$\frac{1}{T_0} \int_0^{T_0} f(x.s)ds < -a_1 \quad \text{for all } x \in U.$$

Let M_2 given by Proposition 2 and choose $M > \max\{M_2, |z(0)|\}$. Define

$$T_\ell = \inf\{n \geq \ell : x(r_n) \notin U \text{ or } |z(r_n)| < M\}.$$

Then

$$\{L((x(n))_n) \subset K, \ |z(n)| \geq M \text{ for } n \text{ large enough}\} \subset \bigcup_{\ell \in \mathbb{N}^*} \{T_\ell = +\infty\}. \tag{14}$$

For large enough M, we prove that $\mathbb{P}[T_\ell = +\infty] = 0$ for all ℓ.

By Proposition 2, there exists $\alpha > 0$ which depends on T_0 and a_1 such that

$$\mathbb{P}\left[\frac{1}{T_0} \sum_{i=r_k}^{r_{k+1}-1} \frac{|z(i+1)| - |z(i)|}{|z(i)|} \geq -a_1/2 \right] \leq \frac{\alpha}{|z(r_k)|}$$

for any $k \geq \ell$, on the event $T_\ell > k$. Let $N(i) = |z(i+1)| - |z(i)|$. On the event $\sum_{i=r_k}^{r_{k+1}-1} \frac{N(i)}{|z(i)|} < 0$, we claim that

$$\sum_{i=r_k}^{r_{k+1}-1} \frac{N(i)}{|z(i)|} \geq \frac{|z(r_{k+1})| - |z(r_k)|}{B^{-1}|z(r_k)| + 1}. \tag{15}$$

To prove this inequality, let us first assume that by contradiction $|z(r_{k+1})| - |z(r_k)| \geq 0$. Then in the interval of time from r_k to r_{k+1}, more balls were added than removed. Therefore, for every ball removed when the number of balls is $|z(i)|$ for some $r_k \leq i \leq r_{k+1}$, there is ball added when the number of balls is at most $|z(i)| - 1$. Consequently $\sum_{i=r_k}^{r_{k+1}-1} N(i)/|z(i)| > 0$, a contradiction. Hence $|z(r_{k+1})| - |z(r_k)| < 0$: more balls are removed than added. For every ball that is added when the number of balls is $|z(i)|$ for some $r_k \leq i \leq r_{k+1}$, a ball is removed when the state is at least $|z(i)| + 1$. Now the remaining $|z(r_k)| - |z(r_{k+1})|$ balls were removed when the number of balls was at least $B^{-1}|z(r_k)| + 1$. This proves (15). As a consequence

$$|z(r_{k+1})| - |z(r_k)| \leq B^{-1}|z(r_k)| \sum_{i=r_k}^{r_{k+1}-1} \frac{N(i)}{|z(i)|}$$

and

$$|z(r_{k+1})| - |z(r_k)| \leq B^{-1}|z(r_k)| \frac{-a_1 T_0}{2}$$

with probability greater than $1 - \alpha/|z(r_k)|$ on the event $\{T_\ell > k\}$. Moreover, $|z(r_{k+1})| - |z(r_k)|$ can never be larger than $r_{k+1} - r_k < T_0 B|z(r_k)|$. Hence,

$$\mathbb{E}\left[|z(r_{k+1})| - |z(r_k)| \mid T_\ell > k \right] \leq \frac{-a_1 T_0 |z(r_k)|}{2B} \left(1 - \frac{\alpha}{|z(r_k)|} \right) + T_0 B|z(r_k)| \frac{\alpha}{|z(r_k)|}$$

$$\leq |z(r_k)| \frac{-a_1 T_0}{2B} + \frac{\alpha a_1 T_0}{2B} + T_0 B\alpha$$

$$\leq -\epsilon |z(r_k)| \leq -\epsilon M$$

for some $\epsilon > 0$. The remainder of the proof is similar to the proof of Benaïm et al. [5, Proposition 2]: for $n > \ell$ we have

$$
\begin{aligned}
0 \leq \mathbb{E}(|z(r_{n \wedge T_\ell})|) &= \sum_{k=l+1}^{n} \mathbb{E}[|z(r_{k \wedge T_\ell})| - |z(r_{(k-1) \wedge T_\ell})|] + \mathbb{E}(|z(r_\ell)|) \\
&= \sum_{k=l+1}^{n} \mathbb{E}\left[|z(r_k)| - |z(r_{k-1})| \mid T_\ell \geq k\right] \mathbb{P}[T_\ell \geq k) + \mathbb{E}(|z(r_\ell)|] \\
&\leq -\epsilon M \sum_{k=l+1}^{n} \mathbb{P}[T_\ell \geq k] + \mathbb{E}[|z(r_\ell)|].
\end{aligned}
$$

Taking the limit as $n \to \infty$, we get that $\sum_{k=\ell+1}^{\infty} \mathbb{P}[T_\ell \geq k] \leq \mathbb{E}[|z(r_\ell)|]/\epsilon M$. The Borel–Cantelli lemma implies that $\mathbb{P}[T_\ell = \infty] = 0$. Relation (14) implies that

$$\mathbb{P}[\{L(x(n)) \subset K\} \cap \{|z(n)| \geq M \text{ for } n \text{ large enough}\}] = 0. \quad \blacksquare$$

4. Applications

To illustrate the applicability of our results, we examine stochastic analogs of two well-known evolutionary dynamics: replicator dynamics from evolutionary game theory [14,10] and selection–mutation dynamics from population genetics [9,6,10].

4.1. Replicator processes

Consider a population consisting of individuals playing k different strategies. In the absence of interactions, each individual produces offspring at rate b and dies at rate d. Each individual initiates an encounter with another individual at rate v. Individual encounters are random. If an individual with strategy i initiated an encounter with an individual with strategy j, then the individual with strategy i either gives birth with probability b_{ij}, dies with probability d_{ij} or is unaffected by the encounter with probability $u_{ij} = 1 - b_{ij} - d_{ij}$. Similarly, the individual with strategy j gives births, dies, or is unaffected with probabilities b_{ji}, d_{ji}, and u_{ji}. While this description yields a continuous-time Markov chain $\tilde{z}(t)$, we focus on the embedded discrete-time Markov chain $z(n)$ corresponding to the state of $\tilde{z}(t)$ at the nth update via a birth, death, or encounter between two individuals. We note that the probability of unbounded population growth (i.e. non-extinction) and convergence of the population distribution to a particular set in S_k are equal for the embedded and continuous-time processes. Hence, for the questions we are interested in, there is no loss of information by restricting our attention to the discrete-time model.

Let $e_1 = (1, 0, 0, \ldots, 0)$, $e_2 = (0, 1, 0, 0, \ldots, 0)$, $\ldots$, $e_k = (0, \ldots, 0, 0, 1)$ be the standard basis of $\mathbb{R}^k$. At each update, $z(n)$ can change either by one individual giving birth $(+e_i)$ possibly following an encounter, one individual dying $(-e_i)$ possibly following an encounter, or pairs of births or deaths following an encounter between individuals $(e_i - e_j, -e_i - e_j$ or $e_i + e_j$ for $i, j \in \{1, 2, \ldots, k\})$. Setting $\gamma = \frac{1}{b+d+v}$, the limiting transition functions p_w, as $|z| \to \infty$, are defined by

$$p_{e_i}(x) = \gamma x_i \left(b + 2v \sum_{j=1}^{k} x_j b_{ij} u_{ji} \right)$$

$$p_{e_i+e_j}(x) = 2\gamma v x_i x_j b_{ij} b_{ji} \quad \text{for } i \neq j$$

$$p_{2e_i}(x) = \gamma v x_i^2 b_{ii}^2$$
$$p_{e_i-e_j}(x) = 2\gamma v x_i x_j b_{ij} d_{ji} \quad \text{for } i \neq j$$

and analogously for p_{-e_i}, $p_{-e_i-e_j}$ and p_{-2e_i}. The transition functions, $\Pi(z, z+w)$, for the non-limiting process $z(n)$ are given by $p_w(x = \frac{z}{|z|})$ with the exception that

$$\Pi(z, z+2e_i) = \gamma v x_i \frac{(x_i|z|-1)}{|z|} b_{ii}^2$$

(and analogously for $-2e_i$) to ensure that these interactions are between *different* individuals playing the same strategy i.

Let $B = (b_{ij})_{i,j}$ and $D = (d_{ij})_{i,j}$ be the birth and death matrices. The mean limit ODE for the process is given by

$$\frac{dx}{dt} = \sum_w p_w(x)(w - x\alpha(w)) = 2v\gamma \, x \circ \left((B-D)x - x^T(B-D)x \right) \tag{16}$$

where $\circ$ denotes the Hadamard product. Setting $A = 2v\gamma(B-D)$ yields the standard form of the replicator equations [10]:

$$\frac{dx}{dt} = x \circ (Ax - x^T Ax). \tag{17}$$

Two key results for these equations described by Hofbauer and Sigmund [10] are the following.

Theorem 6. *Let A be a $k \times k$ matrix. If there exist $p_1, \ldots, p_k > 0$ such that*

$$\sum_{i=1}^k p_i \left((Ax)_i - x^T Ax \right) > 0$$

for all equilibria of (17) *lying on the boundary ∂S_k of the simplex, then there exists an attractor $\mathcal{A}$ for* (17) *whose basin of attraction is $S_k \backslash \partial S_k$.*

Theorem 7. *Let A be a $k \times k$ matrix. If there exists a compact invariant set $K \subset S_k \backslash \partial S_k$ for* (17), *then there exists a unique positive equilibrium $\hat{x} \in S_k \backslash \partial S_k$ such that*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t (x.s) \, ds = \hat{x}$$

and

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t (x.s)^T A(x.s) \, ds = \hat{x}^T A\hat{x}$$

for any solution of (17) *with initial condition in K.*

These two theorems in conjunction with our main results Theorems 4 and 5, imply the following result. We recall that F is a *face* of S_k if there exists a set $I \subset \{1, \ldots, k\}$ such that $F = \{x \in S_k : x_i = 0 \text{ for } i \in I\}$.

Theorem 8. *Let $z(n)$ be a replicator process with $b > 0$ and $d > 0$ and $\prod_i z_i(0) > 0$. Define $A = 2v\gamma(B-D)$ as above. Let F be a face of the simplex S_k. If there exist $p_1, \ldots, p_k > 0$ such*

that

$$\sum_{i \notin I} p_i \left((Ax)_i - x^T A x \right) > 0 \tag{18}$$

for all equilibria $x \in \partial F$ and

$$\frac{b-d}{b+d+\nu} + \hat{x}^T A \hat{x} > 0 \tag{19}$$

for the interior equilibrium $\hat{x} \in F \backslash \partial F$, then there exists a compact set $\mathcal{A}$ in $F \backslash \partial F$ such that

$$\mathbb{P}\left[\sum_n \frac{1}{|z(n)|^{1+\delta}} < \infty \text{ for all } \delta > 0 \text{ and } L(\{x(n)\}_n) \subset \mathcal{A} \right] > 0.$$

Alternatively, if K is a compact, invariant subset of $F \backslash \partial F$ and

$$\frac{b-d}{b+d+\nu} + \hat{x}^T A \hat{x} < 0 \tag{20}$$

for the interior equilibrium $\hat{x} \in F \backslash \partial F$, then there exists $M > 0$ such that

$$\mathbb{P}\left[|z(n)| \geq M \text{ for n sufficiently large and } L(\{x(n)\}_n) \subset \mathcal{K} \right] = 0.$$

Proof. To prove the first assertion, assume that F is a face of the simplex S_k such that (18) holds. Since $\prod_i z_i(0) > 0$ and $d > 0$, there is a positive probability that $z(n)$ lies in the interior of the face F at some update $n \geq 1$. Hence, without loss of generality and by shifting time, we assume that $z(0)$ lies in the interior of F. Theorem 6 implies that there is an attractor $\mathcal{A}$ in the interior of F for the replicator dynamics restricted to F and $B(\mathcal{A})$ equals the interior of F. The assumptions that $b > 0$ and $d > 0$ imply that $B(\mathcal{A})$ is attainable. Assumption (19) and Theorem 7 imply there is average positive growth rate in $B(\mathcal{A})$ i.e. (7) holds. Applying Theorem 4 completes the proof of the first assertion.

To prove the second assertion, assumption (20) and Theorem 7 imply there is average negative growth rate at K i.e. (8) holds. Applying Theorem 5 completes the proof of the second assertion. $\square$

To illustrate the use of this result, we consider the hypercycle replicator equations featured in [10, Chapter 12]. In this game of k strategies, interactions between an individual playing strategy i with individuals playing strategy $i - 1$ (n in the case $i = 1$) catalyze births. These interactions occur at rate ν, while births and deaths independent of interactions occur at rates b and d, respectively. The birth and death matrices associated with interactions are given by

$$B = \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & \cdot & 1 \\ 1 & 0 & 0 & \cdot & \cdot & \cdot & 0 \\ 0 & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 & 0 \end{pmatrix} \quad \text{and} \quad D = 0. \tag{21}$$

For the replicator equation (17) with $A = 2\nu\gamma(B - D)$, Hofbauer and Sigmund [10, Theorem 12.1.2] prove that there is a globally stable equilibrium $\hat{x} = (1/k, 1/k, \ldots, 1/k)$ in $S_k \backslash \partial S_k$ for $k \leq 4$. In contrast, for $k \geq 5$, this equilibrium is unstable see, e.g., [10, Section 12.1] and there is a global, non-equilibrium attractor $\mathcal{A} \subset S_k \backslash \partial S_k$ which attracts all initial conditions in $S_k \backslash \partial S_k$ except those on the stable manifold of $\hat{x}$ [10, Theorem 12.3.1].
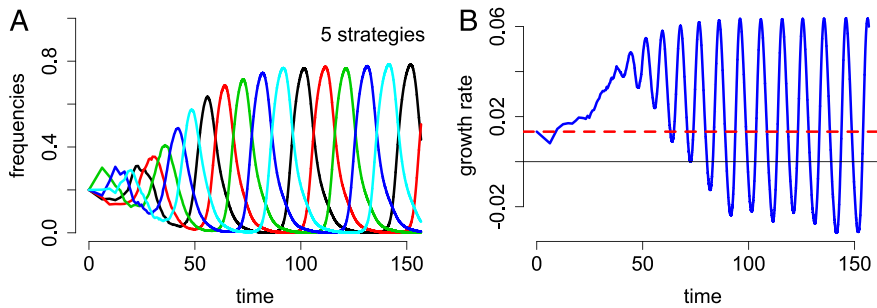
Fig. 1. Sample trajectories of the replicator process $x(n)$ with $b = 1$, $d = 2.5$, $v = 4$, $n = 5$, B and D given by Eqs. (21). Since $(b - d + 2v/5)/(b + d + v) = 1/75 > 0$, there is growth and convergence with positive probability to an equilibrium attractor a non-equilibrium attractor for $n = 5$ (A). In (B), oscillations of the growth function along the stochastic trajectory are shown. The dashed line corresponds to the equilibrium value $1/75$ of the growth function.

For our stochastic analog of this replicator dynamic, we get growth and convergence with positive probability to one of the vertices whenever $b > d$. When $b - d + 2v/n > 0$, we get growth and convergence with positive probability to $\hat{x}$ whenever $n \leq 4$ and growth and convergence with positive probability to the non-equilibrium attractor $\mathcal{A}$ whenever $n \geq 5$. Moreover, if $n \geq 5$, $b - d < 0$ and $b - d + 2v/n > 0$ is sufficiently close to zero, then we conjecture that there are points in $\mathcal{A}$ where the growth function $f(x) = \frac{b-d}{b+d+v} + \frac{2v}{b+d+v} (x_1 x_n + x_2 x_1 + \cdots + x_{n-1} x_{n-2} + x_n x_{n-1})$ is negative. In this case, our Theorem 4 implies convergence with positive probability while Theorem 2 from the earlier work of Benaïm et al. [5] would not. Fig. 1 illustrates converge to the non-equilibrium attractor for $n = 5$ and the oscillations exhibited in the growth function $f(x)$.

4.2. Selection–mutation processes

Two key evolutionary processes are natural selection in which there is differential survival or reproduction amongst genetically distinct individuals and mutation in which parents produce offspring with novel genotypes from their own. A simple, continuous time deterministic model of these processes acting simultaneously on populations tracks the gametes of the populations and assumes that selection acting on diploid individuals and mutation acting on gametes occur independently of one another. While this simplification is somewhat unrealistic, it turns out to be sufficiently realistic to provide useful insights and, for our purposes, illustrates how urn models can be used in the context of population genetics.

We consider a population of haploid individuals (gametes) with a single locus with k possible alleles, $A_1, \ldots, A_k$. All gametes die at a constant rate $d > 0$ and fuse with another gamete at a constant rate $v > 0$. When two gametes, say of types A_i and A_j, fuse to produce an individual with genotype $A_i A_j$ they produce, on average, $f_{ij}/2$ gametes of type A_i and $f_{ij}/2$ gametes of type A_j. For each pair i, j, let $B_{ij}(1), B_{ij}(2), \ldots$ be a sequence of independent, identically distributed random variables taking values in $\{0, 2, \ldots, 2m\}$ and satisfying $\mathbb{E}[B_{ij}(n)] = f_{ij}/2$. At the nth fusion of a gamete of type i and j, $B_{ij}(n)$ gametes of type i and j are added to the population. Gametes of type i mutate to type j at rate $\mu_{ij} > 0$. Namely, on this event, a gamete of type i is replaced with a gamete of type j. Let $\mu = \sum_{i \neq j} \mu_{ij}$ be the total mutation rate.

As in the case of the replicator processes, we are interested in the discrete-time embedded stochastic process where $z(n)$ is the state of the population immediately following the nth demographic event. This process has three types of demographic events: a gamete is removed

due to death ($-e_i$), a gamete changes types due to mutation ($e_i - e_j$ for $i \neq j$), or new gametes are added to the population due to births following the fusing of two gametes ($l(e_i + e_j)$ for some $l \leq m$). Setting $\gamma = \frac{1}{d+\mu+\nu}$, the limiting transition functions p_w, as $|z| \to \infty$, corresponding to these transitions are defined by

$$p_{e_j - e_i}(x) = \gamma \mu_{ij} x_i \quad \text{for } i \neq j$$
$$p_{-e_i}(x) = \gamma d x_i$$
$$p_{\ell(e_i + e_j)}(x) = \gamma \nu x_i x_j \mathbb{P}[B_{ij}(n) = 2\ell] \quad \text{for } \ell \in \{0, 1, \ldots, m\}.$$

The transition functions, $\Pi(z, w)$, for the actual process $z(n)$ are given by $p_w(x = \frac{z}{|z|}, w)$ with the exception that the terms $x_i^2 = \frac{z_i^2}{|z_i|^2}$ are replaced by $x_i(x_i|z| - 1)/|z|$ to ensure that interactions between different individuals playing the same strategy i also take place.

Let $F = (f_{ij})_{i,j}$ and $M = (\mu_{ij})_{i,j}$. Then the mean limit ODE for $z(n)$ is given by the mutation–selection equation [10, Section 20.1]

$$\frac{dx}{dt} = \gamma \nu x \circ (Fx - x^T Fx) + \gamma \mu (M^T x - x). \tag{22}$$

Hofbauer [9] proved that the selection–mutation equation for (22) can exhibit gradient-like dynamics or non-equilibrium dynamics depending on the mutation rates. For example, the following result shows that if rate of mutating to gamete type j is the same for all gamete types, then the dynamics are gradient-like.

Theorem 9 (*Hofbauer [9]*). *If $\mu_{ij} = \mu_j$ for all $i \neq j$, then all solutions $x(t)$ of (22) converge to the set of equilibria of (22).*

Using the earlier work, Theorems 2 and 3 of Benaïm et al. [5], we get the following corollary:

Corollary 1. *Assume that $\mu_{ij} = \mu_j$ for all $i \neq j$ and F, M are such that (22) has a finite number of stable, hyperbolic equilibria $\hat{x}_1, \ldots, \hat{x}_m$ of which $\hat{x}_1, \ldots, \hat{x}_s$ with $s \leq m$ are linearly stable. The fertility selection process $x(n)$ grows and converges to $\hat{x}_i$ for some $1 \leq i \leq s$ with positive probability if*

$$\nu \hat{x}_i^T F \hat{x}_i > d.$$

Furthermore, on the event of linear growth, $x(n)$ converges to $\hat{x}_i$ for some $1 \leq i \leq s$.

Without the strong assumption on mutation rates, the selection–mutation dynamics can give rise to non-equilibrium dynamics. When this occurs, the following result is useful.

Corollary 2. *Assume (22) has a stable periodic attractor $\mathcal{A} \subset S_k \setminus \partial S_k$ of period T. Let $x.t$ denote a solution of (22) with $x \in \mathcal{A}$. If*

$$\frac{\nu}{T} \int_0^T (x.t)^T F(x.t) > d \tag{23}$$

and $x \in S_k \setminus \partial S_k$, then

$$\mathbb{P}\left[\sum_n \frac{1}{|z(n)|^{1+\delta}} < +\infty \, \forall \delta > 0, \text{ and } L(x(n)) \subset \mathcal{A} \right] > 0.$$
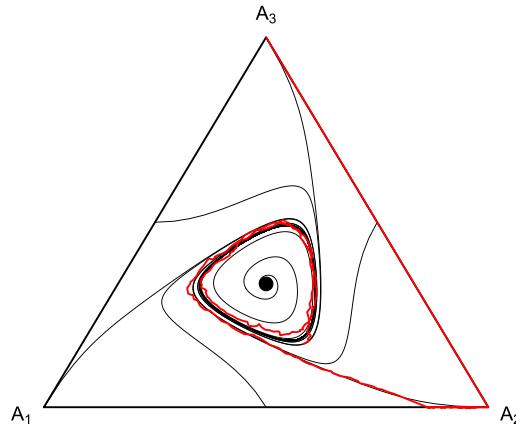
Fig. 2. The dynamics of the selection–mutation process resulting in a limit cycle as described in the text. Trajectories of the mean limit ODE shown in black. A stochastic trajectory is shown in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Alternatively, if inequality (23) *is reversed, then*

$$\mathbb{P}\left[|z(n)| \geq M \text{ for } n \text{ large enough and } L(x(n)) \subset \mathcal{A}\right] = 0.$$

Hofbauer [9] illustrated how selection–mutation equations for $k = 3$ alleles can lead to oscillatory dynamics. Specifically, assume all heterozygous $A_i A_j$ with $i \neq j$ have the same fitness, i.e. $f_{ij} = f$ for all $i \neq j$, all homozygotes $A_i A_i$ have the same fitness, i.e. $f_{ii} = f + s$ for all i, and the mutation rates are cyclic symmetric, i.e. $\mu_{ij} = \mu_{i-j}$ where $\sum_{i=0}^{2} \mu_i = 1$ and the index i is considered as a residue modulo 3. If $\mu_1 \neq \mu_2$ and s is slightly larger than $\frac{9}{2}(\mu_1 + \mu_2)$, then there is a stable periodic orbit (of say period T). Assume x lies on this periodic orbit. If

$$\frac{\nu}{T} \int_0^T s \|x.t\|^2 dt + \nu f > d$$

then Corollary 2 implies that there is a positive probability the population grows and its distribution converges to this periodic orbit. Convergence with positive probability to a periodic orbit is illustrated in Fig. 2. Conversely, if

$$\frac{\nu}{T} \int_0^T s \|x.t\|^2 dt + \nu f < d$$

then the population distribution cannot converge to this periodic orbit.

Acknowledgments

Appendix

In this Appendix, we prove some of the key technical lemmas and estimates used in the proofs of our main results.

A.1. Proof of Lemma 2

On the event $\left\{|z(r_k)| \geq \frac{4BCL}{\delta}\right\}$, we have $\Gamma_2(k, T_0) \leq \frac{2}{B^{-1}|z(r_k)|} \leq \delta/2C$. By (9), the left expression in (10) is smaller than

$$\mathbb{P}\left[\Gamma_1(k, T_0) \geq \frac{\delta}{2C} \middle| \mathcal{F}_{r_k}\right].$$

On the other hand, we have

$$\mathbb{E}\left[\Gamma_1(k, T_0)^2 | \mathcal{F}_{r_k}\right] \leq \mathbb{E}\left[\sum_{r=r_k}^{r_{k+1}-1} \frac{\|U\|^2}{|z(r)|^2}\right]$$

$$\leq \frac{\|U\|^2}{B^{-1}|z(r_k)|}\mathbb{E}\left[\sum_{r=r_k}^{r_{k+1}-1} \frac{1}{|z(r)|}\right]$$

$$\leq \frac{BT_0\|U\|^2}{|z(r_k)|},$$

and (10) holds. $\quad\square$

A.2. Proof of Proposition 1

By the Markov property, it suffices to prove the estimate for $k = 0$ with $x(0) = x \in U$. Define $M_1 = \max\{3B(L + 1/T_0)/\delta, 12L^2BC/\delta\}$. We have

$$\frac{1}{T_0}\left|\int_0^{T_0} f(x.s)ds - \sum_{r=0}^{m(T_0)-1} \frac{1}{|z(r)|}f(x.\tau(r))\right|$$

$$\leq \frac{1}{T_0}\left|\sum_{r=0}^{\tau(r+1)}\left(\int_{\tau(r)}^{\tau(r+1)} f(x.s)ds - \frac{1}{|z(r)|}f(x.\tau(r))\right)\right| + \frac{1}{T_0}\left|\int_{T_0}^{\tau(m(T_0))} f(x.s)ds\right|$$

$$\leq \frac{1}{T_0}\left|\sum_{r=0}^{m(T_0)-1}\left(\int_{\tau(r)}^{\tau(r+1)} f(x.s)ds - \frac{1}{|z(r)|}f(x.\tau(r))\right)\right| + \frac{1}{T_0}(\tau(m(T_0)) - T_0)$$

$$\leq \frac{1}{T_0}\sum_{r=0}^{m(T_0)-1}\int_{\tau(r)}^{\tau(r+1)} \|f(x.s)ds - f(x.\tau(r))\| \, ds + \frac{1}{T_0}\frac{1}{|z(m(T_0))|}$$

$$\leq \frac{1}{T_0}\sum_{r=0}^{m(T_0)-1}\int_{\tau(r)}^{\tau(r+1)} L|\tau(r+1) - \tau(r)|ds + \frac{1}{T_0}\frac{1}{|z(m(T_0))|}$$

$$\leq \frac{1}{T_0}\sum_{r=0}^{m(T_0)-1}\frac{L}{|z(r+1)|^2} + \frac{1}{T_0}\frac{1}{|z(m(T_0))|}$$

$$\leq \frac{1}{|z(0)|} (B(L + 1/T_0))$$

$$\leq \frac{\delta}{3}$$

on the event V_0.

Since

$$\frac{1}{T_0} \left| \sum_{r=0}^{m(T_0)-1} \frac{1}{|z(r)|} (f(x.\tau(r)) - f(x(r))) \right| \leq L \sup_{0 \leq r \leq m(T_0)-1} \|x.\tau(r) - x(r)\|.$$

Lemma 2 implies

$$\mathbb{P} \left[\frac{1}{T_0} \left| \sum_{r=0}^{m(T_0)-1} \frac{1}{|z(r)|} (f(x.\tau(r)) - f(x(r))) \right| \geq \delta/3 \right] \leq \frac{9L^2 C_0(T_0)}{|z(0)|\delta^2}$$

on the event V_0.

Finally,

$$\frac{1}{T_0} \left| \int_0^{T_0} f(x.s) - f(y.s)ds \right| \leq L \sup_{0 \leq s \leq T_0} \|x.s - y.s\| \leq \frac{\delta}{3}$$

for $x \in U$ by choosing U to be a sufficiently small neighborhood of y.

These three estimates plus the triangle inequality complete the proof of the proposition, with $C_1(T_0) = 9L^2 C_0(T_0)$. □

A.3. Proof of Proposition 2

By the Markov property, it suffices to prove the estimate for $k = 0$ with $x(0) = x \in U$. Define $N(i) = |z(i + 1)| - |z(i)|$, $D(i) = N(i) - E[N(i)|z(i)]$ and

$$G = \frac{1}{T_0} \sum_{i=0}^{m(T_0)-1} \frac{D(i)}{|z(i)|}.$$

Observe that $|D(i)| \leq 2m$. Therefore

$$E \left[G^2 | z(0) \right] \leq \frac{1}{T_0^2} E \left[\sum_{i=0}^{m(T_0)-1} \frac{D(i)^2}{|z(i)|^2} \mid z(0) \right]$$

$$\leq \frac{4m^2 B^2}{T_0^2 |z(0)|^2} m(T_0) \leq \frac{4m^2 B^3}{T_0 |z(0)|}$$

where we have used the fact that $T_0 B^{-1}|z(0)| \leq m(T_0) \leq T_0 B|z(0)|$ (see point (ii) before Proposition 2). Chebyshev's inequality implies

$$P \left[|G| \geq \frac{\delta}{3} \mid z(0) \right] \leq \frac{36m^2 B^3}{T_0 |z(0)|\delta^2}. \tag{24}$$

Let $C_1(T_0)$ and M_1 be as defined by Proposition 1. Notice that, by definition of f and assumption (A2), we have

$$|\mathbb{E}(N(i) \mid z(i)) - f(x(i))| \leq \frac{a}{|z(i)|}.$$

Hence,

$$\frac{1}{T_0} \left| \sum_{i=0}^{m(T_0)-1} \frac{\mathbb{E}(N(i) \mid z(i)) - f(x(i))}{|z(i)|} \right| \leq \frac{aBT_0}{|z(0)|} \leq \frac{\delta}{3}$$

provided M_2 is large enough.

Consequently, Proposition 1 with a δ value of $\delta/3$ and inequality (24) imply

$$\mathbb{P}\left[\frac{1}{T_0} \left| \sum_{i=0}^{m(T_0)-1} \frac{N(i)}{|z(i)|} - \int_0^{T_0} f(x(0).s)ds \right| \geq \delta \mid z(0) \right]$$

$$\leq \mathbb{P}\left[\frac{1}{T_0} \left| \sum_{i=0}^{m(T_0)-1} \frac{D(i)}{|z(i)|} \right| \geq \frac{\delta}{3} \mid z(0) \right]$$

$$+ \mathbb{P}\left[\frac{1}{T_0} \left| \sum_{i=0}^{m(T_0)-1} \frac{f(x(i))}{|z(i)|} - \int_0^{T_0} f(x(0).s)ds \right| \geq \frac{\delta}{3} \mid z(0) \right]$$

$$\leq \frac{36m^2 B^3}{T_0|z(0)|\delta^2} + \frac{9C_1(T_0)}{|z(0)|\delta^2} \leq \frac{C_2(T_0)}{|z(0)|\delta^2}$$

on the event W_0 with $M_2 \geq M_1$ large enough and $C_2(T_0) = 9C_1(T_0) + \frac{36m^2 B^3}{T_0}$. □

A.4. *Proof of claim (13)*

Let us prove (13):

$$\sum_{i=r_k+1}^{r_{k+1}-1} \frac{N(i)}{|z(i)|} \leq \frac{n_k}{B^{-1}|z(r_k)|} + \frac{m}{2}(r_{k+1} - r_k)\left(\frac{1}{B^{-1}|z(r_k)|} - \frac{1}{B^{-1}|z(r_k)| + m} \right) \quad (25)$$

where $n_k = |z(r_{k+1})| - |z(r_k)|$. Assume without loss of generality that $n_k > 0$. For every ball which is removed between times $\tau(r_k)$ and $\tau(r_{k+1})$ when state is $z(r)$, a ball is added when the state is at least $|z(r)| - m$. The quantity

$$\frac{1}{|z(r)| - m} - \frac{1}{|z(r)|}$$

is maximal when $|z(r)| - m$ is minimal, and $z(r) - m$ can never be smaller than $B^{-1}|z(r_k)|$ for $r \in [r_k, r_{k+1} - 1]$. Moreover there is at most $m(r_{k+1} - r_k)/2$ balls removed in the process. Finally the remaining n_k balls were added when the state was at least $B^{-1}|z(r_k)|$. A very rough upper bound is given by (25). □

References

[1] R. Antia, R. Regoes, J. Koella, C. Bergstrom, The role of evolution in the emergence of infectious diseases, Nature 426 (6967) (2003) 658–661.

[2] K.B. Athreya, P.E. Ney, Branching Processes, Dover Publications Inc., Mineola, NY, 2004, reprint of the 1972 original [Springer, New York].

[3] M. Benaïm, A dynamical system approach to stochastic approximations, SIAM J. Control Optim. 34 (1996) 437.

[4] M. Benaïm, Dynamics of stochastic approximation algorithms, Sémin. Probab. Strasbg. 33 (1999) 1–68.

[5] M. Benaïm, S. Schreiber, P. Tarres, Generalized urn models of evolutionary processes, Ann. Appl. Probab. 14 (3) (2004) 1455–1478.

[6] M. Bulmer, The selection–mutation-drift theory of synonymous codon usage, Genetics 129 (1991) 897–907.

[7] C.C. Conley, Isolated Invariant Sets and the Morse Index. Vol. 38, American Mathematical Soc., 1978.

[8] J. Haldane, A mathematical theory of natural and artificial selection, in: Mathematical Proceedings of the Cambridge Philosophical Society. Vol. 23, Cambridge University Press, 1927, pp. 607–615.

[9] J. Hofbauer, The selection mutation equation, J. Math. Biol. 23 (1) (1985) 41–53.

[10] J. Hofbauer, K. Sigmund, Evolutionary Games and Population Dynamics, Cambridge University Press, 1998.

[11] R. McArthur, E.O. Wilson, The Theory of Island Biogeography, Princeton University Press, 1967.

[12] M. Park, C. Loverdo, S. Schreiber, J. Lloyd-Smith, Multiple scales of selection influence the evolutionary emergence of novel pathogens, Philos. Trans. R. Soc. B 368 (2013) 20120333.

[13] S. Schreiber, Urn models, replicator processes, and random genetic drift, SIAM J. Appl. Math. 61 (2001) 2148.

[14] P. Schuster, K. Sigmund, Replicator dynamics, J. Theoret. Biol. 100 (1983) 533–538.

[15] M. Soule, Viable Populations for Conservation, Cambridge University Press, 1987.