



Persistence and extinction for stochastic ecological models with internal and external variables

Michel Benaïm¹ · Sebastian J. Schreiber²

Received: 26 September 2018 / Revised: 27 March 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

The dynamics of species' densities depend both on internal and external variables. Internal variables include frequencies of individuals exhibiting different phenotypes or living in different spatial locations. External variables include abiotic factors or non-focal species. These internal or external variables may fluctuate due to stochastic fluctuations in environmental conditions. The interplay between these variables and species densities can determine whether a particular population persists or goes extinct. To understand this interplay, we prove theorems for stochastic persistence and exclusion for stochastic ecological difference equations accounting for internal and external variables. Specifically, we use a stochastic analog of average Lyapunov functions to develop sufficient and necessary conditions for (i) all population densities spending little time at low densities i.e. stochastic persistence, and (ii) population trajectories asymptotically approaching the extinction set with positive probability. For (i) and (ii), respectively, we provide quantitative estimates on the fraction of time that the system is near the extinction set, and the probability of asymptotic extinction as a function of the initial state of the system. Furthermore, in the case of persistence, we provide lower bounds for the expected time to escape neighborhoods of the extinction set. To illustrate the applicability of our results, we analyze stochastic models of evolutionary games, Lotka–Volterra dynamics, trait evolution, and spatially structured disease dynamics. Our analysis of these models demonstrates environmental stochasticity facilitates coexistence of strategies in the hawk–dove game, but inhibits coexistence in the rock–paper–scissors game and a Lotka–Volterra predator–prey model. Furthermore, environmental fluctuations with positive auto-correlations can promote persistence of evolving populations and persistence of diseases in patchy landscapes. While our results help close the gap between the persistence theories for deterministic and stochastic systems, we highlight several challenges for future research.

Keywords Stochastic difference equations · Stochastic persistence · Coexistence · Extinction

✉ Sebastian J. Schreiber
sschreiber@ucdavis.edu

Extended author information available on the last page of the article

1 Introduction

In population biology, environmental stochasticity refers to the effects of fluctuations in environmental factors (e.g. temperature, precipitation) on demography. These demographic effects include fluctuations in survival, growth, and reproduction and can result in fluctuations in population densities, disease prevalence, and genotypic frequencies. This stochasticity can drive populations extinct (Lewontin and Cohen 1969; Gyllenberg et al. 1994; McLaughlin et al. 2002; Benaïm and Schreiber 2009; Hening and Nguyen 2018b), facilitate coexistence (Hutchinson 1961; Chesson and Warner 1981; Chesson 1982, 1985; Chesson and Ellner 1989; Chesson 1994; Kuang and Chesson 2009; Chesson 2018), reverse competitive dominance (Benaïm and Lobry 2016), maintain or disrupt genetic diversity (Gillespie 1973, 1978; Gillespie and Turelli 1989), and alter the persistence and spread of infectious diseases (Altizer et al. 2006).

One approach to studying these effects is to analyze stochastic difference or differential equations. For these equations, stochastic persistence corresponds to a statistical tendency for the population dynamics to avoid low densities (Chesson 1982; Chesson and Ellner 1989; Schreiber 2012). Thus far, general methods for identifying stochastic persistence exist for unstructured populations (Schreiber et al. 2011; Hening and Nguyen 2018a), and structured (e.g. spatial or stage structure) discrete-time models on compact state spaces (Roth and Schreiber 2014b). Few general results, however, exist for identifying when one or more species are asymptotically tending toward extinction. A notable exception is the work of Hening and Nguyen (2018a) who proved a general condition for extinction for stochastic differential equation models of unstructured populations. This condition, however, does not apply to models with intransitive outcomes, such as the evolutionary game of rock–paper–scissor (Hofbauer and Sigmund 2003). Moreover, none of these aforementioned persistence or extinction results apply to models with general forms of internal and external variables (Patel and Schreiber 2018).

All populations have internal and external variables that impact their population dynamics and, more generally, the ecological dynamics of the communities in which they reside. Internal variables are intrinsic to the population. For example, most populations exhibit a multitude of genotypes which vary in demographically important traits. Differential survival and reproduction of the genotypes alter the distributions of these traits. These frequency changes within the population may alter the rate at which the population grows creating a feedback between the population density and the trait distribution (the internal variable) (Vincent and Brown 2005; Schoener 2011; Schreiber et al. 2018). Similarly, for populations structured by age or space, the distribution of ages or spatial locations are internal variables that often generate feedbacks with the total population density (Chesson 2000; Evans et al. 2013; Hening et al. 2018).

External variables are extrinsic to the population. For example, environmental conditions (e.g. precipitation, temperature, nitrogen availability) often fluctuate over time and impact survivorship, growth, and reproduction of individuals. Indeed, the traditional view of environmental stochasticity corresponds to fluctuations in these external variables driving ecological dynamics, but not vice-versa. However, there is evidence

of active, bidirectional feedbacks between the abundance and composition of ecological communities and local weather patterns (Eltahir 1998; Zeng et al. 1999; Kucharski et al. 2013) and fire (Staver and Levin 2012). These bidirectional feedbacks also arise in ecological communities with ecosystem engineers, such as beavers and oysters, that modulate the environment in a way that influences their demography, as well as, other species in the community (Jones et al. 1994; Cuddington et al. 2009). External variables also may include other species that interact with the focal species of a given model or system.

As these internal and external variables may influence persistence and extinction, we develop criteria for these outcomes in stochastic difference equations allowing for these auxiliary variables. In Sect. 2, we introduce the class of multi-species stochastic difference equations with auxiliary variables and our main assumptions. In Sect. 3, we discuss two concepts of stochastic persistence (almost-sure stochastic persistence and stochastic persistence in probability) and define realized per-capita growth rates with respect to invariant measures. To prove results about both forms of stochastic persistence, we use the methods of Benaïm (2018) that introduces stochastic analogs of Hofbauer (1981)'s average Lyapunov functions. The proofs in our setting (compact and discrete-time) simplify substantially when compared to the general results of (Benaïm 2018) and, thereby, highlight the core ingredients of these new methods. In Sect. 4, we prove two new classes of extinction results for stochastic difference equations. One of these results shows if there is a stochastically persisting subcommunity that can not be invaded by the other species, then the other species are extinction prone. This result is a discrete-time analog of a result proven by (Hening and Nguyen 2018a) for stochastic differential equation models of unstructured interacting species. Our second extinction result naturally complements our persistence theorem and provides a sufficient condition for convergence with positive probability to the extinction set, where one or more species have density zero. Unlike the first extinction result, the second extinction result allows us to handle rock–paper–scissor type dynamics. In Sect. 5, we apply our results to models of evolutionary games, multi-species interactions of the Lotka–Volterra type, trait evolution in a fluctuating environment, and spatially structured disease dynamics. Each of these examples is designed to highlight different ways that our results can be used. In Sect. 6, we conclude with discussing how our results relate to the existing literature and raising open mathematical challenges. Proofs are given in Sect. 7.

2 Models and assumptions

We consider n interacting species whose densities at time t are given by $X_t = (X_t^1, X_t^2, \dots, X_t^n)$ which lies in $[0, \infty)^n =: \mathbb{R}_+^n$. The species dynamics interact with the auxiliary variable Y_t which lies in $(-\infty, \infty)^k =: \mathbb{R}^k$. As discussed below, Y_t may correspond to abiotic forcing, feedbacks with environment variables, the internal structure of each species, or any combination thereof. Both of these variables may be influenced by stochastic forces captured by a sequence of independent and identically distributed (i.i.d.) random variables ξ_1, ξ_2, \dots taking values in a Polish space \mathcal{E} i.e. a separable completely metrizable topological space. The fitness f^i of individuals in

species i depends on the densities X_t of all species, the auxiliary variables Y_t , and the random variable ξ_{t+1} that captures the stochastic changes that occur over the time interval $(t, t + 1]$. Similarly, the update rule G of the auxiliary variables depends on X_t , Y_t , and ξ_{t+1} .

Under these assumptions, our model is

$$\begin{aligned} X_{t+1}^i &= X_t^i f^i(X_t, Y_t, \xi_{t+1}) \quad i = 1, 2, \dots, n && \text{(species densities)} \\ Y_{t+1} &= G(X_t, Y_t, \xi_{t+1}) && \text{(auxiliary variables).} \end{aligned} \quad (2.1)$$

Our standing assumptions for (2.1) are:

- A1 For each $i = 1, 2, \dots, n$, the fitness function $f^i(z, \xi)$ is continuous in $z = (x, y)$, measurable in (z, ξ) , and strictly positive.
- A2 The auxiliary variable update function G is continuous in $z = (x, y)$ and measurable in (z, ξ) .
- A3 There is a compact subset \mathcal{S} of $\mathbb{R}_+^n \times \mathbb{R}^k$ such that all solutions $Z_t = (X_t, Y_t)$ to (2.1) satisfy $Z_t \in \mathcal{S}$ for t sufficiently large.
- A4 For all $i = 1, 2, \dots, n$, $\sup_{z, \xi} |\log f^i(z, \xi)| < \infty$.

Assumptions **A1–A2** ensure that the Markov chain Z_t is Feller (see Sect. 3.3 for a definition) and that positive densities are mapped to positive densities. Assumption **A3** implies the dynamics remain bounded. Assumption **A4** uniformly bounds the log fitness of all species. These assumptions are often met in models as illustrated in Sect. 5.

Some examples of the auxiliary variable Y_t include:

Markovian environmental forcing: To model Markovian fluctuations in the environment, Y_{t+1} can represent the state of the environment over the time interval $(t, t + 1]$ which determines the fitness of the species. Many forms of Markovian dynamics can be represented as $Y_{t+1} = G(Y_t, \xi_{t+1})$ where ξ_1, ξ_2, \dots are i.i.d. For example, autoregressive processes of the form $Y_{t+1} = AY_t + \xi_{t+1}$ where A is a matrix whose spectral radius is less than one. In this case, if ξ_t take values in a compact set, then Y_t converges to a unique stationary distribution which is compactly supported (Diaconis and Freedman 1999). Alternatively, suppose there a finite number of environmental states $\{1, 2, \dots, k\}$ with transition probabilities p^{ij} i.e. $\mathbb{P}[Y_{t+1} = j | Y_t = i] = p^{ij}$. One can represent this Markov chain Y_t as a composition of random maps by defining $\xi_t = (\xi_t^1, \dots, \xi_t^k)$ to be a random vector such that $\mathbb{P}[\xi_t^i = j] = p^{ij}$ and defining $G(Y, \xi) = \xi^Y$. The representation of Markov chains as random maps can be done more generally (e.g. continuous state spaces in Theorem 1.2 of Kifer (1986)). For any of these choices of the auxiliary variable dynamics, the fitness function are of the form $f^i(X_t, Y_{t+1}) = f^i(X_t, G(Y_t, \xi_{t+1}))$.

Population structure: Many forms of discrete population structure (e.g. age, stages, spatial) can be represented by (2.1) in the Y dynamics. For example, if the interacting species live in k distinct patches, then the auxiliary variable $Y_t = (Y_t^{ij})_{\{1 \leq i \leq n, 1 \leq j \leq k\}}$ can be a matrix where Y_t^{ij} is the fraction of species i living in patch j . The dynamics of these spatial frequencies of species i , even in

the limit $X_t^i = 0$, can be derived directly from a model description of the species densities in each patch.

Trait evolution: The auxiliary variables can represent ecologically-important traits under selection. Changes in these traits can drive changes in population densities and, conversely, the densities of the species can select for different trait values. One classic model of trait evolution is due to Lande (1976) in which the trait Y^i of species i follows the gradient of the log fitness of the species i.e.

$Y_{t+1}^i = Y_t^i + \alpha \frac{\partial \log f^i(X_t, Y_t, \xi_{t+1})}{\partial Y^i}$ for some constant $\alpha > 0$ that determines the speed of evolution. This gradient type dynamic also appears in models of adaptive dynamics (Metz et al. 1996; Michod 2000; Vincent and Brown 2005). More generally, these auxiliary variables may keep track of the frequencies of different genotypes and their associated trait distribution (Gillespie 1978; Schreiber 2019).

Environmental feedbacks for ecosystem engineers: Ecosystem engineers are organisms, such as beavers and oysters, that modulate the abiotic environment in a manner influencing the demography of itself or other species (Jones et al. 1994; Cuddington et al. 2009). For example, many species of oysters produce large, complex reef structures which are composed of living and dead oysters. These reefs provide a positive feedback on oyster growth and survival, and provide ecosystem services, including water filtration, habitat and predator refuge for a variety of other species (Moore et al. 2016, 2018). To model these system dynamics of oysters, modelers include the reef size Y_t as an auxiliary variable Y_t that can substantially alter the dynamics of the system (Jordan-Cooley et al. 2011; Moore et al. 2018).

Several illustrative examples of these different uses of the auxiliary variable Y_t are given in Sect. 5.

3 Stochastic persistence

In this section, we begin by stating two definitions of stochastic persistence (Schreiber 2012): persistence in probability and almost-sure persistence. The first corresponds to what Chesson (1982) called stochastically bounded coexistence and takes an ensemble point of view. The second takes the perspective of a single, typical realization of the Markov chain (Schreiber et al. 2011). To derive a persistence criterion, we define the realized per-capita growth rates when species are infinitesimally rare. This approach to characterizing coexistence in stochastic models goes back to the work of Turelli (1981). Using these realized per-capita growth rates, we introduce the stochastic analog of the realized community per-capita growth rate when rare due to Hofbauer (1981). This community growth rate allows us to define the stochastic analog of Hofbauer (1981)'s average Lyapunov functions and, thereby, prove sufficient conditions for stochastic persistence of either form.

3.1 Persistence in probability and almost-sure persistence

As we just mentioned, there are two ways to think about the asymptotic behavior of $\{Z_t\}_{t=0}^{\infty}$. First, one can ask what is the distribution of $Z_t = (X_t, Y_t)$ far into the

future. For example, what is the probability that each species' density is greater than ϵ in the long term i.e. $\mathbb{P}[X_t \geq (\epsilon, \dots, \epsilon)]$ for large t ? The answer to this question provides information about what happens across many independent realizations of the ecological dynamics. Alternatively, one might be interested about the statistics associated with a single realization of the process i.e. a single time series. For instance, one could ask what fraction of the time was each species' density greater than ϵ ?

For both ways of quantifying persistence, we define the *extinction set*

$$\mathcal{S}_0 = \{(x, y) \in \mathcal{S} : \min_i x^i = 0\}$$

which corresponds to the set of population states where at least one of the species or genotypes is absent. We also define, for any $\eta > 0$, the η -neighborhood of the extinction set by

$$\mathcal{S}_\eta = \{(x, y) \in \mathcal{S} : \min_i x^i \leq \eta\}$$

that corresponds to states at which at least one of the species has a density less than η .

With respect to the first approach to studying asymptotic behavior, the model (2.1) is *stochastically persistent in probability* (Chesson 1982; Chesson and Ellner 1989; Schreiber 2012) if for all $\epsilon > 0$ there exists an $\eta > 0$ such that

$$\limsup_{t \rightarrow \infty} \mathbb{P}_z[X_t \in \mathcal{S}_\eta] \leq \epsilon \quad \text{whenever } z = (x, y) \in \mathcal{S} \setminus \mathcal{S}_0.$$

In words, there is an arbitrarily small probability of any of the species being at arbitrarily low densities anytime in the future. Moreover, while these upper bounds depend on initial conditions in the short-term, they are independent of initial conditions (provided all species are initially present) in the long-term.

For the second approach to study the asymptotic behavior of $Z_t = (X_t, Y_t)$, we define the *occupation measure* by

$$\Pi_t = \frac{1}{t} \sum_{s=1}^t \delta_{Z_s}$$

where δ_{Z_s} denotes a Dirac measure at Z_s i.e. $\delta_{Z_s}(A) = 1$ if $Z_s \in A$ and 0 otherwise for any (Borel) set $A \subset \mathcal{S}$. $\Pi_t(A)$ equals the proportion of time the community spends in A up to time t . Our model (2.1) is *almost surely stochastically persistent* (Schreiber et al. 2011; Schreiber 2012) if for all $\epsilon > 0$ there exists $\eta > 0$ such that

$$\limsup_{t \rightarrow \infty} \Pi_t(\mathcal{S}_\eta) \leq \epsilon \quad \text{with probability one}$$

whenever $Z_0 = z_0 \in \mathcal{S} \setminus \mathcal{S}_0$. This form of persistence implies that fraction of time spent by the community in \mathcal{S}_η goes to zero as η goes to zero. As with persistence in probability, these upper bounds are asymptotically independent of the initial condition.

Regarding the definition of stochastic persistence, Chesson (1982) wrote:

This criterion requires that the probability of observing a population below any given density, should converge to zero with density, uniformly in time. Consequently it places restrictions on the expected frequency of fluctuations to low population levels. Given that fluctuations in the environment will continually perturb population densities, it is to be expected that any nominated population density, no matter how small, will eventually be seen. Indeed this is the usual case in stochastic population models and is not an unreasonable postulate about the real world. Thus a reasonable persistence criterion cannot hope to do better than place restrictions on the frequencies with which such events occur.

3.2 Invariant measures and realized per-capita growth rates

To determine whether stochastic persistence in either form occurs, we need to understand what happens to a species when it becomes rare in the community. In particular, does the species tend to increase when rare or decrease when rare? To quantify this tendency, we consider the per-capita growth rate of the species averaged over the fluctuations in the community dynamics X_t , the auxiliary variable Y_t , and the i.i.d. random variables ξ_{t+1} . These averages will be taken over invariant probability measures which represent the stationary dynamics of the common species. Recall, a Borel probability measure μ on \mathcal{S} is *invariant* if for all continuous functions $h : \mathcal{S} \rightarrow \mathbb{R}$

$$\int_{\mathcal{S}} h(z)\mu(dz) = \int_{\mathcal{S}} \mathbb{E}_z[h(Z_1)]\mu(dz)$$

where

$$\mathbb{E}_z[h(Z_1)] = \mathbb{E}[h(Z_1)|Z_0 = z].$$

An invariant probability measure μ is *ergodic* if it can not be written as a non-trivial convex combination of invariant probability measures.

Given any species i , the set $\mathcal{S}^i = \{z = (x, y) \in \mathcal{S} : x^i > 0\}$ is the set of states for which species i has positive density. For an ergodic measure μ , $\mu(\mathcal{S}^i)$ is either 1 or 0 for any species i i.e. either supports the presence of species i or not. Hence, for an ergodic measure μ , we define $S(\mu) = \{1 \leq i \leq n : \mu(\mathcal{S}^i) = 1\}$ as the *species support* of μ .

To understand coexistence, we can ask: what happens to a species when it becomes rare in the community? To this end, imagine that some species are infinitesimally rare and the dynamics of the remaining, common species are characterized by an ergodic measure μ . As μ supports only a subset of species, $S(\mu)$ is a proper subset of $\{1, \dots, n\}$. For one of the infinitesimally rare species $i \notin S(\mu)$, its rate of growth is determined by the linearized model

$$\tilde{X}_{t+1}^i = f^i(Z_t, \xi_{t+1})\tilde{X}_t^i \quad \text{with } Z_t = (X_t, Y_t)$$

for μ -almost every $Z_0 = z_0 \in \mathcal{S}_0$. The solution to this stochastic linear difference equation is

$$\tilde{X}_t^i = \prod_{\tau=t-1}^0 f^i(Z_\tau, \xi_{\tau+1}) \tilde{X}_0^i.$$

Taking the log of this solution, dividing by t , and taking the limit (provided it exists), species i tends to increase in density if

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \log f^i(Z_\tau, \xi_{\tau+1}) > 0$$

and tends to decrease if

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \log f^i(Z_\tau, \xi_{\tau+1}) < 0.$$

The following proposition shows that for μ -almost every $Z_0 = z_0 \in S$,

$$r^i(\mu) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \log f^i(Z_\tau, \xi_{\tau+1})$$

where $r^i(\mu)$ is the *realized per-capita growth rate* of species i :

$$r^i(\mu) = \int_S \mathbb{E}[\log f^i(z, \xi_t)] \mu(dz). \quad (3.1)$$

For any infinitesimally rare species $i \notin S(\mu)$, $r^i(\mu)$ determines how quickly the species tends to increase when introduced at low densities. Alternatively, the realized per-capita growth rate $r^i(\mu)$ equals zero for any species i supported by μ . Intuitively, if the species' density is neither asymptotically growing or declining, then the species' realized per-capita growth must be zero. A proof of the proposition is given in Sect. 7.1.

Proposition 1 *Let μ be an invariant probability measure and $i \in \{1, \dots, n\}$. Then there exists a bounded Borel map $\hat{r}^i : S \rightarrow \mathbb{R}$ such that:*

(i) *With probability one and for μ -almost every $z \in S$*

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \log f^i(Z_\tau, \xi_{\tau+1}) = \hat{r}^i(z) \text{ when } Z_0 = z; \quad (3.2)$$

and

$$\lim_{t \rightarrow \infty} \frac{1}{t} \sum_{\tau=0}^{t-1} \mathbb{E}_z \left[\log f^i(Z_\tau, \xi_{\tau+1}) \right] = \hat{r}^i(z) \text{ when } Z_0 = z; \quad (3.3)$$

- (ii) $\int_{\mathcal{S}} \widehat{r}^i(z) \mu(dz) = r^i(\mu)$;
- (iii) if μ is ergodic, then $\widehat{r}^i(z) = r^i(\mu)$ μ -almost surely and $r^i(\mu) = 0$ for all $i \in S(\mu)$.

3.3 Realized per-capita community growth rates and stochastic persistence

Following the approach introduced by Josef Hofbauer (Hofbauer 1981; Hofbauer and Sigmund 1998), our criterion for stochastic persistence is

$$\begin{aligned} &\text{there exist positive } p^1, \dots, p^n \text{ s.t. } \sum_i p^i r^i(\mu) > 0 \\ &\text{for all ergodic } \mu \text{ with } \mu(\mathcal{S}_0) = 1. \end{aligned} \tag{3.4}$$

Criterion (3.4) requires that there is some weighting of the species such that the weighted average of the realized per-capita growth is positive for all ergodic measures supporting a subset of the community. In particular, for each ergodic measure μ , (3.4) requires there is at least one species such that one of the missing species can increase when rare i.e. $r^i(\mu) > 0$ for some $i \notin S(\mu)$. In the case of two species models, positive realized per-capita growth rates for at least one missing species is sufficient to ensure that there are positive weights p^1, p^2 satisfying (3.4). However, for more species, the criterion is more subtle as we illustrate in the evolutionary game of rock–paper–scissors (see Sect. 5.1).

To see why criterion (3.4) is useful, we follow the approach developed in Benaïm (2018); Benaïm and Lobry (2016). To describe this approach, we observe that assumptions **A1**–**A2** imply that (2.1) is (weak) Feller (Meyn and Tweedie 2009): the transition operator P takes continuous functions $h : \mathcal{S} \rightarrow \mathbb{R}$ to continuous functions Ph defined by

$$(Ph)(z) = \mathbb{E}_z[h(Z_1)].$$

For any $t \geq 0$, let P_t denote the t -time step transition operator of our Markov model (2.1) defined by

$$(P_t h)(z) = \mathbb{E}_z[h(Z_t)] \quad \text{for all } z \in \mathcal{S}.$$

Now, if we define

$$V(z) = - \sum_{i=1}^n p^i \log(x^i) \quad \text{for } z = (x, y) \in \mathcal{S},$$

then

$$P_t V(z) - V(z) = - \sum_{i=1}^n p^i \left(\sum_{\tau=0}^{t-1} \mathbb{E}_z[\log f^i(Z_\tau, \xi_{\tau+1})] \right).$$

Proposition 1 and criterion (3.4) imply that there is a $T \geq 1$ and $\alpha > 0$ such that (see proof in Sect. 7.3)

$$P_T V(z) - V(z) \leq -\alpha \quad \text{for all } z \in \mathcal{S}_0. \tag{3.5}$$

For this choice of $T \geq 1$, we prove the following key result in Sect. 7.2.

Proposition 2 *Assume (3.4) holds. For $\theta > 0$ sufficiently small, there exist $\rho \in (0, 1)$ and $\beta > 0$ such that*

$$(P_T V_\theta)(z) \leq \rho V_\theta(z) + \beta \quad \text{for all } z \in \mathcal{S} \setminus \mathcal{S}_0 \text{ where } V_\theta(z) = \exp(\theta V(z)).$$

Using Proposition 2, criterion (3.4) implies stochastic persistence in probability. To understand why, note that the t -step transition operator P_t maps Borel probability measures μ to Borel probability measures μP_t by duality i.e. μP_t is the probability measure defined by

$$\int_{\mathcal{S}} h(z) (\mu P_t)(dz) := \int_{\mathcal{S}} (P_t h)(z) \mu(dz) \quad \text{for all continuous } h : \mathcal{S} \rightarrow \mathbb{R}.$$

Namely, if $Z_0 = z$ is drawn randomly according to $\mu(dz)$, then μP_t is the law of Z_t . In particular, if $\mu = \delta_{z_0}$ for some $z_0 \in \mathcal{S}$, then $\delta_{z_0} P_t$ is the law of Z_t given $Z_0 = z$.

Now, for simplicity, assume $T = 1$ (see Sect. 7.3 for the general case $T \geq 1$). Given any $z_0 \in \mathcal{S} \setminus \mathcal{S}_0$, any integer t and $\eta > 0$, Proposition 2 implies

$$\mathbb{P}_{z_0}[Z_t \in \mathcal{S}_\eta] \min_{z \in \mathcal{S}_\eta \setminus \mathcal{S}_0} V_\theta(z) \leq \int_{\mathcal{S}} V_\theta(z) d(\delta_{z_0} P_t) \leq (\rho)^t V_\theta(z_0) + \frac{\beta}{1 - \rho}. \tag{3.6}$$

For $\eta \leq 1$ and $z = (x, y) \in \mathcal{S}_\eta \setminus \mathcal{S}_0$,

$$V_\theta(z) = \prod_{i=1}^n (x^i)^{-\theta p^i} \geq a_0(\eta)^{-b} \quad \text{where } a_0 = \min_{z=(x,y) \in \mathcal{S} \setminus \mathcal{S}_0} \|x\|^{b(1-n)}, \quad b = \theta \max_i p^i. \tag{3.7}$$

Thus, for $\eta \leq 1$, inequalities (3.6)–(3.7) and $\rho < 1$ imply

$$\limsup_{t \rightarrow \infty} \mathbb{P}_{z_0}[Z_t \in \mathcal{S}_\eta] \leq a(\eta)^b \quad \text{where } a = \frac{\beta}{a_0(1 - \rho)}$$

which implies persistence in probability.

To show that criterion (3.4) implies almost sure stochastic persistence requires two ingredients. First, using an extension of an argument presented in Schreiber et al. (2011), we show in Sect. 7.3 that if $Z_0 \in \mathcal{S} \setminus \mathcal{S}_0$, then all weak* limit points of Π_t are, with probability one, invariant probability measures supported by $\mathcal{S} \setminus \mathcal{S}_0$. Second, we show that the weight placed by positive invariant probability measures near the boundary can be controlled uniformly. Specifically, if μ is an invariant probability

measure with $\mu(\mathcal{S}_0) = 0$ and $\int V_\theta(z)\mu(dz)$ is finite (the proof of the general case without this integrability assumption appears in Sect. 7.3), then Proposition 2 implies for any $k \geq 1$

$$\int V_\theta(z)\mu(dz) = \int (P_{kT} V_\theta)(z)\mu(dz) \leq (\rho)^k \int V_\theta(z)\mu(dz) + \frac{\beta}{1 - \rho}.$$

As $\rho < 1$, taking the limit as $k \rightarrow \infty$ yields

$$\int V_\theta(z)\mu(dz) \leq \frac{\beta}{1 - \rho}.$$

Thus, for any $\eta > 0$,

$$\mu(\mathcal{S}_\eta) \min_{z \in \mathcal{S}_\eta} V_\theta(z) \leq \int V_\theta(z)\mu(dz) \leq \frac{\beta}{1 - \rho}.$$

Inequality (3.7) implies that

$$\mu(\mathcal{S}_\eta) \leq a(\eta)^b \quad \text{for all } \eta \leq 1 \text{ where } a = \frac{\beta}{a_0(1 - \rho)}.$$

Theorem 1 *If criterion (3.4) holds, then (2.1) is almost-surely stochastically persistent and stochastically persistent in probability. In particular, there exists $a, b > 0$ such that for all $\eta \leq 1$ and $z \in \mathcal{S} \setminus \mathcal{S}_0$*

$$\text{(persistence in probability)} \quad \limsup_{t \rightarrow \infty} \mathbb{P}_z [Z_t \in \mathcal{S}_\eta] \leq a(\eta)^b$$

and

$$\text{(almost-sure persistence)} \quad \limsup_{t \rightarrow \infty} \Pi_t [\mathcal{S}_\varepsilon] \leq a(\eta)^b \text{ for } Z_0 = z \text{ almost surely.}$$

3.4 Escaping extinction risk

When stochastic persistence occurs, it is natural to ask: “if one or more species are at very low densities, how long before they reach higher densities?” For $\eta > 0$ sufficiently small, the following proposition provides upper bounds on the time for all species densities to exceed η .

Proposition 3 *Assume that Proposition 2 holds. There exists $\eta > 0$ such that if $Z_0 = (x, y) \in \mathcal{S}_\eta \setminus \mathcal{S}_0$ and*

$$\tau = \inf_{k \geq 1} \{k : Z_{kT} \in \mathcal{S} \setminus \mathcal{S}_\eta\},$$

then

$$\mathbb{P}[\tau > k] \leq \prod_{i=1}^n (x^i)^{-\theta p^i} (\rho)^k. \quad (3.8)$$

Biological implication. Inequality (3.8) implies that the probability that one or more species remain at low densities decreases exponentially over time. Furthermore, the larger the minimal community realized per-capita growth rate (α in (3.5)), the greater this rate of exponential decline.

Proof From the proof of Proposition 2, we can choose $\eta > 0$ sufficiently small so that $P_T V_\theta(z) \leq \rho V_\theta(z)$ for all $z \in \mathcal{S}_\eta \setminus \mathcal{S}_0$. Benaim (2018, Proposition 8.2) implies that $\mathbb{E}_z[\rho^{-\tau}] \leq V_\theta(z) = \prod_i (x^i)^{-\theta p^i}$ for $z = (x, y) \in \mathcal{S}_\eta \setminus \mathcal{S}_0$. By the Markov inequality,

$$\mathbb{P}_z[\tau \leq k] = \mathbb{P}_z[\rho^{-\tau} \geq \rho^{-k}] \leq \mathbb{E}_z[\rho^{-\tau}] \rho^k \leq \prod_i (x^i)^{-\theta p^i} \rho^k.$$

□

4 Stochastic extinction

To truly understand the conditions for stochastic persistence, we also need to understand when the solutions of our model (2.1) converge to the extinction set \mathcal{S}_0 with positive probability. To this end, we prove two theorems about extinction. The first theorem is a partial converse to Theorem 1, and shows that when the inequality (3.4) is reversed for all ergodic measures supported by \mathcal{S}_0 , population trajectories starting near \mathcal{S}_0 are likely to asymptotically approach \mathcal{S}_0 . To measure the rate of approach to the extinction set \mathcal{S}_0 , define

$$\text{dist}(z, \mathcal{S}_0) = \min_{z' \in \mathcal{S}_0} \|z - z'\|$$

where $\|\cdot\|$ denotes the standard Euclidean norm. Our second main extinction theorem considers the case when there is a subset of persisting species that can not be invaded by the other species i.e. $r^i(\mu) < 0$ for all species not in the community. In this case, we show that when the missing species are sufficiently rare, they are highly likely to go extinct. For both extinction results, we also introduce an accessibility criterion that, when met, implies that all initial conditions lead to asymptotic extinction at an exponential rate with probability one.

Theorem 2 *If there exist positive weights p^1, p^2, \dots, p^n such that*

$$\sum_{i=1}^n p^i r^i(\mu) < 0 \quad (4.1)$$

for all ergodic μ supported by \mathcal{S}_0 , then there exists $a, b > 0$ such that

$$\mathbb{P}_z \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \log \text{dist}(Z_t, \mathcal{S}_0) < 0 \right] \geq 1 - a \text{dist}(z, \mathcal{S}_0)^b$$

where $\text{dist}(z, \mathcal{S}_0) = \min_{z' \in \mathcal{S}_0} \|z - z'\|$.

The proof of Theorem 2 is given in Sect. 7.4 whose strategy is based on the proof of Theorem 3.3 of Benaïm and Lobry (2016) for random switching between two dimensional Lotka–Volterra competition equations. A key ingredient of the proof is to introduce the function $V(z) = \sum_{i=1}^n p^i \log x^i$ —the negative of the V function used in the proof of Theorem 1. The analog of Proposition 2 holds for this choice of V . However, in this case, the function $V_\theta(z) = \prod_{i=1}^n (x^i)^{p^i}$ tends to decrease after T iterates near the extinction set \mathcal{S}_0 .

To get asymptotic extinction with probability one for all initial conditions, we say \mathcal{S}_0 is accessible if for all $\eta > 0$, there exists $\gamma > 0$ such that

$$\mathbb{P}_z[X_t \in \mathcal{S}_\eta \text{ for some } t \geq 1] \geq \gamma \text{ for all } z \in \mathcal{S}.$$

The following corollary implies that accessibility plus the conditions of Theorem 2 imply almost sure extinction. We prove this corollary in Sect. 7.4.

Corollary 1 Assume the assumptions of Theorem 2 hold and \mathcal{S}_0 is accessible. Then

$$\mathbb{P}_z \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \log \text{dist}(Z_t, \mathcal{S}_0) < 0 \right] = 1 \text{ for all } z \in \mathcal{S}.$$

Our second main extinction result considers the case when there is a set of species that can not be invaded by another other species. Recall that for an ergodic probability measure μ , $S(\mu) \subset \{1, 2, \dots, n\}$ is the set of species supported by μ .

Theorem 3 Let $\{1, \dots, n\} = I \cup J$ where $I \cap J = \emptyset$ and $J \neq \emptyset$. Assume

- (i) (2.1) restricted $\mathcal{S}^I := \{(x, y) | x^j = 0 \text{ whenever } j \in J\}$ satisfies that there exists $p^i > 0$ for $i \in I$ and $\sum_{i \in I} p^i r^i(\mu) > 0$ for ergodic μ with $\mu(\mathcal{S}_0^I) = 1$ where $\mathcal{S}_0^I := \{z = (x, y) \in \mathcal{S}^I : \prod_{i \in I} x^i = 0\}$, and
- (ii) $r^j(\mu) < 0$ for any $j \in J$ and ergodic μ with $S(\mu) = I$.

Then there exist $a, b, c > 0$ such that

$$\mathbb{P}_z \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \log(X_t^j) < 0 \text{ for all } j \in J \right] \geq 1 - c \frac{(\max_{j \in J} x^j)^a}{(\min_{i \in I} x^i)^b} \tag{4.2}$$

whenever $z = (x, y)$ satisfies $\prod_{i \in I} x^i > 0$.

The lower bound in Eq. (4.2) implies that for a given set of positive densities for the species in subcommunity I (i.e. $x^i > 0$ for $i \in I$), then the probability of

asymptotic extinction of an unsupported species increases to one as their densities get sufficiently small. The proof of Theorem 3 follows the proof strategy used for unstructured stochastic differential equations by Hening and Nguyen (2018a). For this proof, the key ingredient is the function $V(z) = -\sum_{i \in I} p^i \log x^i + \delta \max_{j \in J} \log x^j$ for a sufficiently small $\delta > 0$. The corresponding $V_\theta(z) = \exp(\theta V(z))$ for $\theta > 0$ sufficiently small tends to be increasing at low densities of species from I and decreasing at low densities of species from J .

To get extinction with probability one, we need an accessibility assumption. To allow for multiple, uninhabitable subcommunities, we consider a finite collection of proper subsets of species, $I_1, \dots, I_m \subset \{1, \dots, n\}$. For each $1 \leq i \leq m$ and $\eta > 0$, define $\mathcal{S}_\eta^{I_i} = \{(x, y) \in \mathcal{S} : x^j \leq \eta \text{ for all } j \in I_i\}$. We say $\cup_i \mathcal{S}^{I_i}$ is *accessible* if for all $\eta > 0$, there exists $\gamma > 0$ such that

$$\mathbb{P}_z[Z_t \in \cup_i \mathcal{S}_\eta^{I_i} \text{ for some } t \geq 1] \geq \gamma$$

whenever $z = (x, y)$ satisfies $\prod_i x^i > 0$.

Corollary 2 *Assume the assumptions of Theorem 3 hold for each I_i , $J_i = \{1, \dots, n\} \setminus I_i$ and $\cup_i \mathcal{S}^{I_i}$ is accessible. Then*

$$\mathbb{P}_z \left[\limsup_{t \rightarrow \infty} \frac{1}{t} \log \text{dist}(Z_t, \mathcal{S}_0) < 0 \right] = 1 \quad \text{for all } z \in \mathcal{S}.$$

The proof of this corollary uses the same strategy as the proof of Corollary 1 in Sect. 7.4. The details are left to the reader.

5 Applications

To illustrate the utility of our theorems, we apply them to four types of models. Our first set of models considers a stochastic version of two and three strategy evolutionary games. It highlights how Theorems 1–3 can be used to fully characterize the possible dynamics of coexistence and extinction in these evolutionary games. The second set of models are stochastic counterparts to the Lotka–Volterra differences equations (Hofbauer et al. 1987). These models exhibit an averaging property that allows one to readily compute the realized per-capita growth rates $r^i(\mu)$ with respect to any ergodic measure. We illustrate the use of this averaging property with a predator–prey model. Our last two models are single species models with different types of auxiliary variables. The first of these models considers a species evolving in a fluctuating environment. This example illustrates how, simultaneously, an evolving trait can be incorporated as an internal variable and how auto-correlated fluctuations can be incorporated as an external variable. The second of these two models explores the dynamics of disease spread in a spatially structured population. This example illustrates how discrete population structure can be represented as internal variables.

5.1 Stochastic evolutionary games

Evolutionary game theory is an important framework for studying frequency-dependent population dynamics (Maynard Smith 1982; Hofbauer and Sigmund 1998, 2003). Remarkably, the three basic games (the Prisoner’s Dilemma, the Hawk–Dove game, and the rock–paper–scissor game) have provided fundamental insights about the evolution of cooperation (Axelrod 1984; Axelrod and Hamilton 1981), animal contests (Maynard Smith 1982; Maynard Smith and Price 1973), and Red Queen dynamics (Sinervo and Lively 1996; Kerr et al. 2002; Kirkup and Riley 2004; Nahum et al. 2011). An important, yet often under-appreciated, consideration in these games is the effect of environmental stochasticity on the maintenance or loss of polymorphisms i.e. coexistence of multiple strategies. Here, we show how our results characterize (generically) persistence in probability and extinction for stochastic versions of these classic games in correlated, as well as, uncorrelated environments.

For the stochastic evolutionary games, there are n competing asexual genotypes or strategies. The frequency of strategy or genotype i is X^i . Individuals interact randomly resulting in fitness payoffs that modify the basal fitness of each strategy. Let $b^i(Y)$ be the basal fitness (possibly zero) of strategy i and $A^{ij}(Y)$ be the fitness payoff (possibly negative) to an individual of strategy i following an interaction with an individual of strategy j . Here Y is an environmental variable that determines the fitness payoffs in any time step. If Y is given by a first-order, multivariate auto-regressive process, then the evolutionary game dynamics are

$$\begin{aligned}
 X_{t+1}^i &= X_t^i \frac{\sum_j A^{ij}(Y_t) X_t^j + b^i(Y_t)}{\sum_{j\ell} A^{j\ell}(Y_t) X_t^j X_t^\ell + \sum_j b^j(Y_t) X_t^j} \\
 Y_{t+1} &= CY_t + \xi_{t+1}
 \end{aligned}
 \tag{5.1}$$

where C is a matrix with a spectral radius strictly less than one, and ξ_1, ξ_2, \dots is a sequence of i.i.d random variables taking value in a compact subset of \mathbb{R}^n . Under these assumptions on the auxiliary variable dynamics, they enter and remain in a compact set $K \subset \mathbb{R}^k$ with probability one. Hence, the state space is given by $\mathcal{S} = \Delta \times K$ where $\Delta = \{(x^1, \dots, x^n) \in [0, 1] : \sum_i x^i = 1\}$ is the probability simplex.

We illustrate our results with two important evolutionary games: the hawk–dove game and the rock–paper–scissors game.

5.1.1 Hawk–dove game

For $n = 2$ strategy games, $\mathcal{S}_0 = \{(1, 0), (0, 1)\} \times K$ which only supports two ergodic measures $\mu^1 = \delta_{(1,0)} \otimes \nu$ and $\mu^2 = \delta_{(0,1)} \otimes \nu$ where \otimes denotes the product of two measures, and ν is the law of the stationary distribution \widehat{Y} of the auto-regressive process Y_t . It follows that

$$r^j(\mu^i) = \mathbb{E} \left[\log \frac{A^{ji}(\widehat{Y}) + b^j(\widehat{Y})}{A^{ii}(\widehat{Y}) + b^i(\widehat{Y})} \right].$$

If both strategies can invade each other ($r^j(\mu^i) > 0$ for $i \neq j$), Theorem 1 implies stochastic persistence. Alternatively, if $r^j(\mu^i) < 0$ for some $i \neq j$, then Theorem 3 implies strategy j is excluded with high probability when its initial frequency is sufficiently low.

For example, consider the hawk–dove game in which hawks (strategy 1) and doves (strategy 2) engage in pairwise contest over a resource worth $V(Y)$ offspring. Interacting doves, split the resource, while hawks get all of the resource when encountering a dove. Encounters between hawks result in one getting all of the resource and the other paying a cost C . Assume that all individuals have a base pay-off of $b(Y) \equiv 1$ and that $V(Y) - C > -2$ to ensure that payoffs are always positive. Then

$$A(Y) = 1 + \begin{pmatrix} \frac{V(Y)-C}{2} & V(Y) \\ 0 & V(Y)/2 \end{pmatrix}.$$

Hawks always invade a population of doves as

$$r^1(\mu^2) = \mathbb{E} \left[\log \left(\frac{1 + V(\widehat{Y})}{1 + V(\widehat{Y})/2} \right) \right] > 0.$$

Doves can invade a population of hawks only if

$$r^2(\mu^1) = \mathbb{E} \left[\log \frac{1}{(V(\widehat{Y}) - C)/2 + 1} \right] > 0.$$

Provided $\text{Var}[V(\widehat{Y})] > 0$, Jensen’s inequality implies

$$\mathbb{E} \left[\log \frac{1}{(V(\widehat{Y}) - C)/2 + 1} \right] > \log \frac{1}{(\mathbb{E}[V(\widehat{Y})] - C)/2 + 1}.$$

Biological implication. *Environmental fluctuations in the payoff $V(Y)$ can facilitate the coexistence of hawk and dove strategies. Importantly, unlike the deterministic hawk–dove game, hawks and doves can coexist even if the mean reward for winning a contest is greater than the cost paid by a hawk losing a battle to another hawk .*

5.1.2 Rock–paper–scissor dynamics

To illustrate the use of Theorem 2, we consider the rock–paper–scissor game. For mathematical convenience, we number the strategies 0 (rock), 1 (paper), and 2 (scissors). When strategy i interacts with strategy $i + 1 \pmod{3}$, it receives a negative payoff $-L_i^i < 0$ while the other strategy receives a positive payoff W_i^{i+1} . Let $Y_t = (W_t^0, W_t^1, W_t^2, L_t^0, L_t^1, L_t^2)$. Then the payoff matrix equals

$$A(Y_t) = \begin{pmatrix} 0 & -L_t^1 & W_t^1 \\ W_t^2 & 0 & -L_t^2 \\ -L_t^3 & W_t^3 & 0 \end{pmatrix}.$$

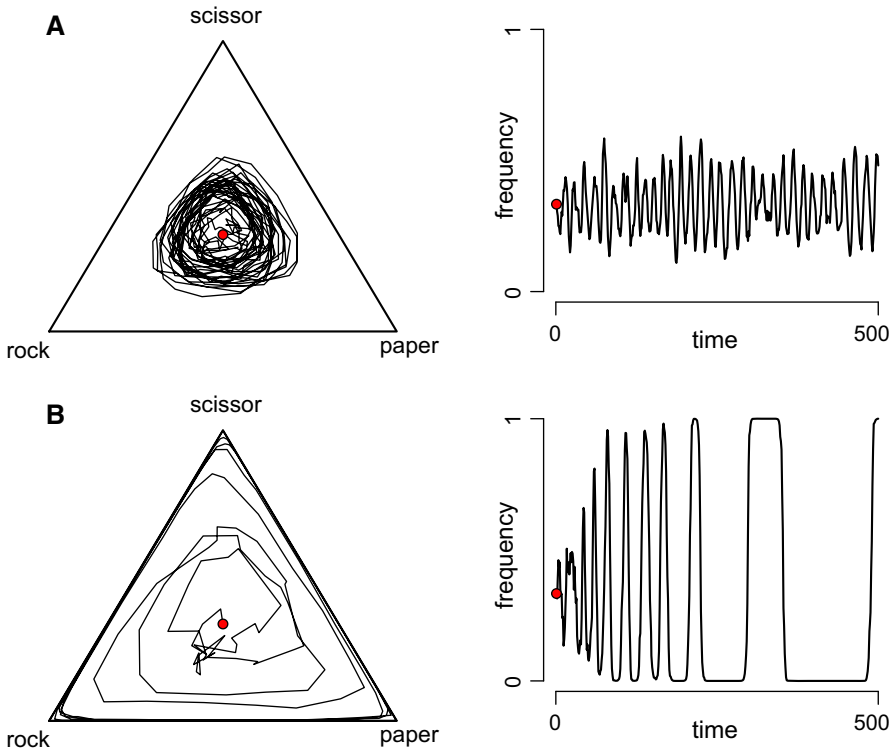


Fig. 1 Noise induced exclusion in the rock–paper–scissors dynamic. In both panels, payoffs fluctuate around the mean values $\mathbb{E}[\widehat{W}_t^i] = 2.3$ and $\mathbb{E}[L_t^i] = 0.5$. In **a**, payoffs are equally likely to be $\pm 40\%$ of the mean payoff. In **b**, payoffs are equally likely to be $\pm 90\%$ of the mean payoff

Assume a base payoff $b(Y) \equiv 1$ of one and $L_t^i < 1$. Furthermore, for simplicity, assume L_t^0, L_t^1, L_t^2 are identically distributed and W_t^0, W_t^1, W_t^2 are identically distributed.

As $1 - L_t^i < 1 + W_t^{i+1}$ implies that for any pair of strategies i and $i + 1 \pmod{3}$, strategy i gets excluded by strategy $i + 1 \pmod{3}$ i.e. $\lim_{t \rightarrow \infty} X_t^{i+1} = 1$ with probability one whenever $X_0^i + X_0^{i+1} = 1$ and $X_0^{i+1} > 0$. Hence, the only invariant measures on the boundary S_0 are given by the pure strategy invariant measures $\delta_i \otimes \nu$ where δ_i is a Dirac measure only supporting strategy i and ν is the law of the stationary distribution \widehat{Y} of the autoregressive process. The realized per-capita growth rates at these invariant measures are given by

$$r^{i-1 \pmod{3}}(\delta_i \otimes \nu) = \mathbb{E}[\log(1 - \widehat{L}^1)], \quad r^i(\delta_i \otimes \nu) = 0, \quad \text{and} \quad r^{i+1 \pmod{3}}(\delta_i \otimes \nu) = \mathbb{E}[\log(1 + \widehat{W}^1)].$$

A simple inequality calculation reveals that there are weights p^i such that $\sum_i p^i r^i (\delta_j \otimes v) > 0$ for all j if and only if

$$\mathbb{E}[\log(1 + W_t^1)] + \mathbb{E}[\log(1 - L_t^1)] > 0. \quad (5.2)$$

When this inequality is met, Theorem 1 implies all three strategies stochastically persist. Alternatively, there are weights p^i such that $\sum_i p^i r^i (\delta_j \otimes v) < 0$ for all j if and only if

$$\mathbb{E}[\log(1 + W_t^1)] + \mathbb{E}[\log(1 - L_t^1)] < 0. \quad (5.3)$$

When this inequality is met, Theorem 2 implies that Z_t asymptotically approaches \mathcal{S}_0 with high probability whenever one of the strategies is at low frequency. Hence, the conditions for robust stochastic persistence (5.2) and exclusion (5.3) are sharp except for the degenerate case in which the left hand side of (5.2) equals zero.

So what effect does environmental stochasticity have on maintaining a polymorphism? When $\text{Var}[W_t^1] + \text{Var}[L_t^1] > 0$, Jensen's inequality implies

$$\mathbb{E}[\log(1 + W_t^1)] + \mathbb{E}[\log(1 - L_t^1)] < \log(1 + \mathbb{E}[W_t^1]) + \log(1 - \mathbb{E}[L_t^1]).$$

Biological implication. *Environmental fluctuations in the payoffs inhibit coexistence of the rock, paper, scissors strategies. In particular, unlike the deterministic games, even if the average winning payoff is greater than the average losing payoff, the three strategies may not coexist (Fig. 1).*

5.2 Stochastic Lotka–Volterra difference equations

Hofbauer et al. (1987) introduced a discrete-time counterpart to the Lotka–Volterra differential equations. These equations consider n interacting species whose dynamics in the absence of environmental stochasticity are

$$X_{t+1}^i = X_t^i \exp\left(\sum_j a^{ij} X_t^j + b^i\right) \quad i = 1, 2, \dots, n \quad (5.4)$$

where a^{ij} describes the per-capita effect of species j on species i 's per-capita growth rate and b^i corresponds to the intrinsic rates of growth of species i . Hofbauer et al. (1987) showed that these difference equation share many mathematical properties of the Lotka–Volterra differential equations. Most importantly, as we describe below for their stochastic counterpart, they exhibit an averaging property that allows one to verify permanence or exclusion by solving linear inequalities.

Here, we consider a stochastic counterpart to these equations by allowing the interaction strengths a^{ij} and intrinsic per-capita growth rates b^i to vary stochastically. Specifically, we assume that for each i, j the sequence $a_1^{ij}, a_2^{ij}, a_3^{ij}, \dots$ are i.i.d. and for each i the sequence $b_1^i, b_2^i, b_3^i, \dots$ are i.i.d. Then the dynamics become

$$X_{t+1}^i = X_{t+1}^i \exp \left(\sum_j a_{t+1}^{ij} X_t^j + b_{t+1}^i \right) \quad i = 1, 2, \dots, n. \tag{5.5}$$

This system of stochastic difference equations generalizes the one introduced by Schreiber et al. (2011) who only considered fluctuations in the b_t^i . Then $\xi_t = (a_t^{ij}, b_t^i)_{1 \leq i, j \leq n}$ for these models.

To ensure the dynamics of (5.5) remain in a compact set, we follow Hofbauer et al. (1987) and define the interaction strengths a_t^{ij} to be *hierarchically ordered* if there exists a reordering of the indices such that $a_t^{ii} < 0$ for all i and $t \geq 0$, and $a_t^{ij} \leq 0$ whenever $i \leq j$ and $t \geq 0$. While this assumption excludes mutualistic interactions, it allows for many types of predator–prey or competitive interactions. The following lemma shows that hierarchically ordered systems with bounded interaction strengths and intrinsic per-capita growth rates exhibit bounded dynamics.

Lemma 1 *If stochastic Lotka–Volterra model (5.5) is hierarchically ordered and there exist $\alpha, \beta > 0$ such that $a_t^{ii} \leq -\alpha$ and $|a_t^{ij}|, |b_t^i| \leq \beta$ for all i, j and $t \geq 0$, then there exists $K > 0$ such that $X_t \in [0, K]^n$ for $t \geq n + 1$.*

Proof Following Hofbauer et al. (1987) observe that

$$X_{t+1}^1 \leq X_t^1 \exp(-\alpha X_t^1 + \beta) \quad \text{for all } t$$

as $a_t^{1j} \leq 0$ for all $j \geq 2$ and $t \geq 0$. Hence, $X_t^1 \in [0, K_1]$ for $t \geq 2$ where $K_1 = \exp(\beta - 1)/\alpha$.

Assume that there exist $K_1, \dots, K_{j-1} > 0$ such that $X_t^i \in [0, K_i]$ for $i \leq j - 1$ and $t \geq i + 1$. We will show that there exists K_j such that $X_t^j \in [0, K_j]$ for $t \geq j + 1$. Indeed, by the hierarchically ordered assumption and our inductive assumption,

$$X_{t+1}^j \leq X_t^j \exp(-\alpha X_t^j + \beta + \sum_{i < j} \beta K_i)$$

for $t \geq j$. Hence, $X_t^j \leq K_j$ for $t \geq j + 1$ where $K_j = \exp(\beta + \beta \sum_{i < j} K_i - 1)/\alpha$. Defining $K = \max K_j$ completes the proof. □

The following lemma shows that verifying the conditions for Theorems 1, 2, and 3 reduces to a linear algebra problem. In particular, this lemma implies that many of the permanence and exclusion criteria developed by Hofbauer et al. (1987); Hofbauer and Sigmund (1998) for hierarchal systems extends to the stochastic Lotka–Volterra difference equations by replacing the deterministic terms a^{ij} and b^i in (5.4) with the expectations $\mathbb{E}[a_t^{ij}]$ and $\mathbb{E}[b_t^i]$ in (5.5).

Lemma 2 *Let μ be an ergodic measure for (5.5). If there exists a unique non-negative solution \hat{x} to*

$$\sum_j \mathbb{E}[a_t^{ij}] \hat{x}^j + \mathbb{E}[b_t^i] = 0 \quad \text{for } i \in S(\mu) \text{ and } \hat{x}^i = 0 \text{ for } i \notin S(\mu)$$

then

$$r^i(\mu) = \begin{cases} 0 & \text{if } i \in S(\mu) \\ \sum_j \mathbb{E}[a_t^{ij}] \widehat{x}^j + \mathbb{E}[b_t^i] & \text{otherwise.} \end{cases}$$

Proof Let μ be an ergodic measure. Assertion (iii) of Proposition 1 implies that

$$0 = r^i(\mu) = \sum_j \mathbb{E}[a_t^{ij}] \int x^j \mu(dx) + \mathbb{E}[b_t^i]$$

for all $i \in S(\mu)$. Since we have assumed there is a unique non-negative solution \widehat{x}^i to this system of linear equations with the additional condition $\widehat{x}^i = 0$ for $i \notin S(\mu)$, it follows that $\int x^i \mu(dx) = \widehat{x}^i$ for all i . Thus, for any $i \notin S(\mu)$,

$$r^i(\mu) = \sum_j \mathbb{E}[a_t^{ij}] \int x^j \mu(dx) + \mathbb{E}[b_t^i] = \sum_j \mathbb{E}[a_t^{ij}] \widehat{x}^j + \mathbb{E}[b_t^i]$$

as claimed. □

To illustrate the use of the lemma, let's consider a predator-prey system with a fluctuating carrying capacity K_t of the prey. In this system, X_t^1 and X_t^2 correspond to the prey and predator densities, respectively. Let ρ be the intrinsic rate of growth of the prey, a the attack rate of the predator on the prey, b the conversion of prey to predator, c the predator's intraspecific competition term, and d the predator's per-capita density-independent death rate. The resulting hierarchically ordered model is given by

$$\begin{aligned} X_{t+1}^1 &= X_t^1 \exp\left(\rho(1 - X_t^1/K_{t+1}) - aX_t^2\right) \\ X_{t+1}^2 &= X_t^2 \exp\left(bX_t^1 - cX_t^2 - d\right). \end{aligned} \tag{5.6}$$

As long as K_t take values in a compact subset of $(0, \infty)$, Lemma 1 implies that X_t enters and remains in a compact subset \mathcal{S} of \mathbb{R}_+^2 . Let $\delta_{(0,0)}$ be the Dirac measure at the origin. Then $r^1(\delta_{(0,0)}) = \rho > 0$ and $r^2(\delta_{(0,0)}) = -d$. Any other ergodic measure μ supported by \mathcal{S}_0 must be supported by prey axis $(0, \infty) \times \{0\}$. For such an ergodic μ , Proposition 1(iii) implies that

$$0 = r^1(\mu) = \rho - \underbrace{\rho \int x^1 \mu(dx^1)}_{=:\widehat{x}^1} \mathbb{E}[1/K_t].$$

Hence, $\widehat{x}^1 = 1/\mathbb{E}[1/K_t]$ i.e. the harmonic mean of K_t . On the other hand,

$$r^2(\mu) = b\widehat{x}^1 - d = b/\mathbb{E}[1/K_t] - d.$$

If $r^2(\mu) > 0$, then Theorem 1 with $p^1 = 2d/\rho$, $p^2 = 1$ implies (5.6) is persistent. On the other hand, if $r^2(\mu) < 0$, then Theorem 3 (with the prey subsystem being the persistent subsystem) implies that asymptotic extinction of the predator occurs with high likelihood whenever the predator density is sufficiently low.

Biological implication. As the d/b can be interpreted as the predator’s break even point (i.e. the prey density at which its per-capita growth rate is zero in the absence on interference), we have shown that predator–prey persistence requires that the harmonic mean of the carrying capacity exceeds the predator’s break even density. As the harmonic mean is less than the arithmetic mean, it follows that fluctuations in the prey’s carrying capacity inhibit predator–prey coexistence.

5.3 Trait evolution in a stationary environment

Following Lande and Shannon (1996), we consider a population with density X_t and a trait (e.g. log body size) z which is normally distributed in the population with mean \bar{z}_t in year t and constant variance σ^2 . The optimal trait value over the time interval $(t, t + 1]$ is θ_{t+1} . The fitness of a individual with trait z over this time step equals

$$W_{t+1}(z) = \exp(r_{\max} - \gamma/2(z - \theta_{t+1})^2 - a(\bar{z}_t)X_t)$$

where γ determines the strength of stabilizing selection and $a(\bar{z}) > 0$ corresponds to the strength of intraspecific competition as a function on the mean trait value. Integrating across the normal distribution of the population gives the mean fitness \bar{W}_{t+1} of the population over the time interval $(t, t + 1]$

$$\begin{aligned} \bar{W}_{t+1} &:= \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{\infty} W_{t+1}(z) \exp(-(z - \bar{z}_t)^2/(2\sigma^2)) dz \\ &= \frac{1}{\sqrt{\gamma\sigma^2 + 1}} \exp\left(r_{\max} - \frac{(\bar{z}_t - \theta_{t+1})^2\gamma}{2\sigma^2\gamma + 2} - a(\bar{z}_t)X_t\right). \end{aligned}$$

Lande (1976) has shown that if $h^2 \in (0, 1]$ is the heritability of the trait z , then the population and the mean trait dynamics can be approximated by

$$\begin{aligned} X_{t+1} &= X_t \bar{W}_{t+1} \\ \bar{z}_{t+1} &= \bar{z}_t + \sigma^2 h^2 \frac{\partial \log \bar{W}_{t+1}}{\partial \bar{z}_t} \\ &= (1 - \alpha)\bar{z}_t + \alpha\theta_{t+1} - \sigma^2 h^2 a'(\bar{z}_t)X_t \text{ where } \alpha = \frac{h^2\sigma^2\gamma}{\sigma^2\gamma + 1} < 1. \end{aligned} \tag{5.7}$$

To finalize the model, we model the optimal trait dynamics as a first-order autoregressive process:

$$\theta_{t+1} = \rho\theta_t + \sqrt{1 - \rho^2}\xi_{t+1} \tag{5.8}$$

where $\rho \in (-1, 1)$ determines the temporal autocorrelation, and $\xi_1, \xi_2, \xi_3, \dots$ is an i.i.d. sequence of random variables with mean 0, variance τ^2 , and compact support. The rescaling of ξ_{t+1} by $\sqrt{1 - \rho^2}$ ensures that the variance of the stationary distribution of θ_t equals τ^2 i.e. ρ only influences the autocorrelation between θ_t and θ_{t+1} but not the long-term variance of θ_t .

In the context of our general model formulation (2.1), $Y = (\bar{z}, \theta)$ is the auxiliary vector, the fitness f for the species is

$$f(X, Y, \xi) = \frac{1}{\sqrt{\gamma\sigma^2 + 1}} \exp\left(r_{\max} - \frac{(\bar{z} - \rho\theta - \sqrt{1 - \rho^2}\xi)^2\gamma}{2\sigma^2\gamma + 2} - a(\bar{z})X\right),$$

and the auxiliary function G is

$$G(X, Y, \xi) = \left((1 - \alpha)\bar{z} + \alpha(\rho\theta + \sqrt{1 - \rho^2}\xi) - \sigma^2 h^2 a'(\bar{z})X, \rho\theta + \sqrt{1 - \rho^2}\xi \right).$$

Assume there exists a compact set $\mathcal{S} \subset [0, \infty) \times (-\infty, \infty)^2$ such that Z_t enters and remains in \mathcal{S} for sufficiently large t . For example, this holds when $a(\bar{z})$ is a fixed positive constant. The extinction set $\mathcal{S}_0 = \{z = (0, y) \in \mathcal{S}\}$ corresponds to extinction of the population. On this extinction set, the auxiliary dynamics Y_t are a first-order, bivariate auto-regressive process:

$$\begin{pmatrix} \bar{z}_{t+1} \\ \theta_{t+1} \end{pmatrix} = \underbrace{\begin{pmatrix} 1 - \alpha & \rho\alpha \\ 0 & \rho \end{pmatrix}}_{=:A} \begin{pmatrix} \bar{z}_t \\ \theta_t \end{pmatrix} + \underbrace{\begin{pmatrix} \alpha\sqrt{1 - \rho^2}\xi_{t+1} \\ \sqrt{1 - \rho^2}\xi_{t+1} \end{pmatrix}}_{:=b_{t+1}}. \tag{5.9}$$

As the spectral radius of A equals $\max\{1 - \alpha, \rho\}$ which is strictly less than one, the dynamics of Y_{t+1} on the extinction set converge to a unique stationary solution $\widehat{Y}_t = (\widehat{\bar{z}}_t, \widehat{\theta}_t)$ (see, e.g., Diaconis and Freedman (1999)). The covariance matrix $\text{Cov}[\widehat{Y}_t]$ of this stationary solution is given by (see, e.g., Schreiber and Moore 2018)

$$\text{vec}(\text{Cov}[\widehat{Y}_t]) = (\text{Id} - A \otimes A)^{-1} \text{vec}(\text{Cov}[b_t])$$

where vec denotes the vec operator that concatenates the columns of the matrix into a column vector and \otimes denotes the Kroenker product. The entries of this covariance matrix are given by

$$\begin{aligned} \text{Var}[\widehat{\theta}_t] &= \tau^2, \\ \text{Var}[\widehat{\bar{z}}_t] &= \tau^2 \frac{\alpha(\alpha - 1)\rho - \alpha}{(\alpha^2 - 3\alpha + 2)\rho + \alpha - 2}, \text{ and} \\ \text{Cov}[\widehat{\bar{z}}_t, \widehat{\theta}_t] &= \tau^2 \frac{\alpha}{(\alpha - 1)\rho + 1}. \end{aligned}$$

Furthermore, as $\mathbb{E}[\widehat{\theta}_t, \widehat{\theta}_{t+1}] = \rho V$, we have

$$\text{Cov}[\widehat{z}_t, \widehat{\theta}_{t+1}] = \tau^2 \frac{\rho\alpha}{(\alpha - 1)\rho + 1},$$

and

$$\text{Var}[\widehat{z}_t - \widehat{\theta}_{t+1}] = \frac{1 - \rho}{1 - (1 - \alpha)\rho} \frac{2\tau^2}{2 - \alpha}.$$

Let μ be the law of the stationary solution $(0, \widehat{Y}_t)$ which is the unique ergodic probability measure on the extinction set \mathcal{S}_0 . Then

$$\begin{aligned} r(\mu) &= r_{\max} - \frac{\gamma}{2} \left(\sigma^2 + \text{Var}[\widehat{z}_t - \widehat{\theta}_{t+1}] \right) \\ &= r_{\max} - \frac{\gamma}{2} \left(\sigma^2 + \frac{1 - \rho}{1 - (1 - \alpha)\rho} \frac{2\tau^2}{2 - \alpha} \right). \end{aligned} \tag{5.10}$$

Theorem 1 implies if $r(\mu) > 0$ then the population stochastically persists. Conversely if $r(\mu) < 0$, then Theorem 2 implies that the population goes extinct asymptotically with positive probability whenever X_0 is sufficiently low. Furthermore, if

$$\mathbb{P} \left[\frac{\gamma(\widehat{z}_t - \widehat{\theta}_{t+1})^2}{2\sigma^2\gamma + 2} > r_{\max} \right] > 0$$

and $r(\mu) < 0$, then \mathcal{S}_0 is accessible and Corollary 1 implies that $\lim_{t \rightarrow \infty} X_t = 0$ with probability one.

Biological implications. Equation (5.10) provides several insights about the effects of trait evolution in a fluctuating environment on population persistence. First, $r(\mu)$ decreases with the strength γ of stabilizing selection and the environmental variation τ^2 in the optimal trait value. Intuitively, both of these quantities increase the long-term genetic load (i.e. the “reduction in the mean fitness of a population relative to a population composed entirely of individuals having optimal genotypes” (Whitlock and Davis 2011)) and, consequently, the chance for persistence. Second, $r(\mu)$ increases with the temporal autocorrelation ρ of the optimal trait in the changing environment. Intuitively, the stronger the temporal autocorrelation, the more time trait evolution has to track the optimal trait (compare Fig. 2a,b to c,d). Indeed, in the limit $\rho \uparrow 1$, there is no effect of τ^2 on population persistence. Finally, the speed of evolution, which is determined by α , plays an important role on persistence. In the limit of no evolution (i.e. $\alpha \downarrow 0$), the genetic variance and environmental variance contributing equally to a reduction in the realized per-capita growth rate i.e. $r(\mu) = r_{\max} - \frac{\gamma}{2} (\sigma^2 + 2\tau^2)$. Alternatively, if the speed of evolution is maximized (i.e. $\alpha \uparrow 1$), then in this limit $r(\mu) = r_{\max} - \frac{\gamma}{2} (\sigma^2 + 2(1 - \rho)\tau^2)$. Hence, if environmental fluctuations are nearly perfectly autocorrelated (i.e. $\rho \approx 1$), then trait evolution eliminates the environmental reduction in the per-capita growth rate. In contrast, if the environmental fluctuations are strongly negatively autocorrelated (i.e. $\rho \approx -1$), then fast evolution causes a twofold reduction in the realized per-capita growth rate. Intuitively, in negatively

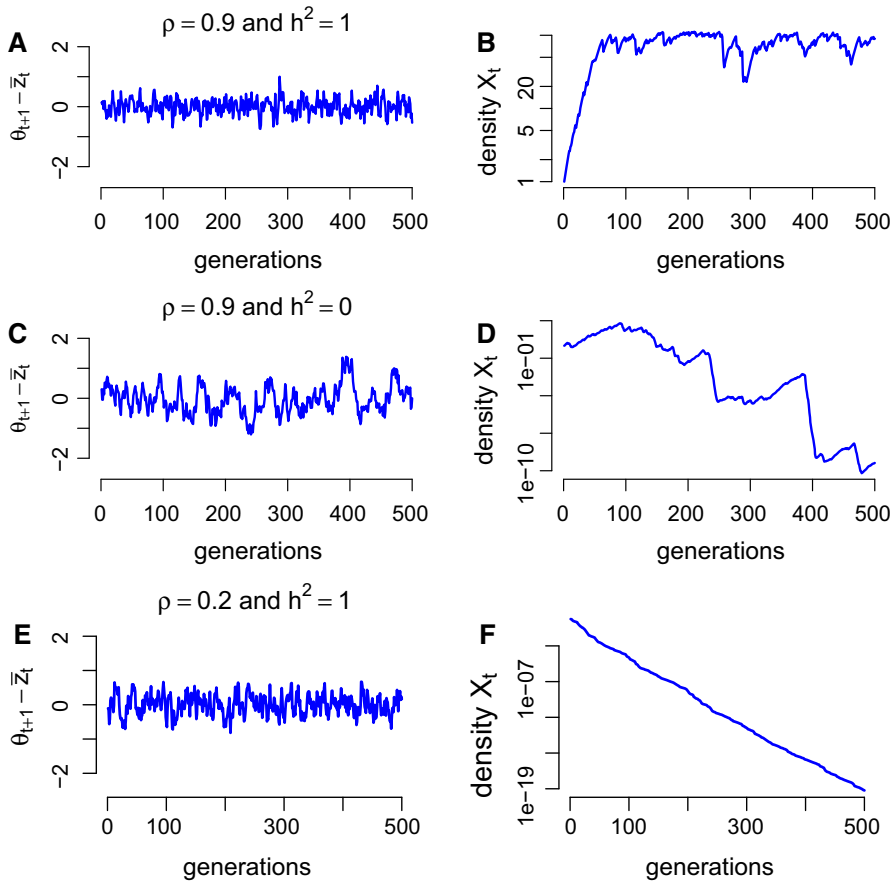


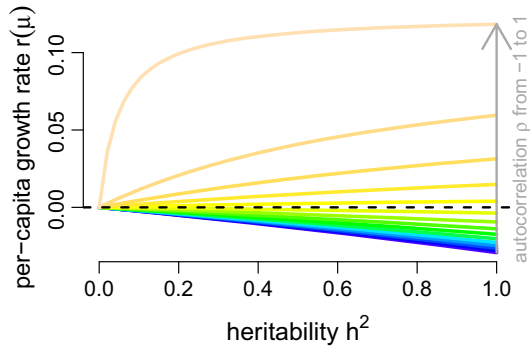
Fig. 2 Evolutionary rescue more likely in autocorrelated environments for (5.7)–(5.8). Left hand panels plot the difference between the mean population trait \bar{z}_t and the optimal trait value θ_{t+1} . The absolute value of this difference corresponds to the genetic load. Right hand panels plot the population density X_t . In (A)–(B), heritability and autocorrelation is high and the population persists i.e. $r(\mu) > 0$. In (C)–(D), the population doesn't evolve (i.e. $h^2 = 0$) and autocorrelation is high, but the population tends toward extinction i.e. $r(\mu) < 0$ and S_0 is accessible. In (E)–(F), heritability is high but autocorrelation is low, but the population tends toward extinction. Parameters: $\gamma = 1$, $\sigma = 0.5$, $r_{\max} = 0.25$, h^2 , ρ as shown in panels, $\tau = 0.5$, and $a = 0.001$

autocorrelated environments, passing on the optimal trait for the parents to the offspring provides the offspring with a suboptimal trait. Hence, in strongly negatively autocorrelated environments, non-evolving populations may be more likely to persist than evolving populations. Several of these conclusions are illustrated in Fig. 3.

5.4 A structured SIS disease model

To illustrate how auxiliary variables can account for population structure, we analyze a spatially structured disease model. A closely related model in continuous time was

Fig. 3 The effects of heritability h^2 and temporal autocorrelation on the realized per-capita growth rate $r(\mu)$. With low autocorrelation, greater heritability decreases $r(\mu)$ and can lead to extinction. With higher autocorrelation, $r(\mu)$ increases with h^2 and fast evolution can rescue the population. Parameter values as in Fig. 2



analyzed by Benaim and Strickler (2019). In this model, there are k subpopulations of individuals with densities S_t^i and I_t^i of susceptible and infected individuals, respectively, in the i -th subpopulation. The subpopulation sizes are constant with a density of $N^i = S_t^i + I_t^i$ individuals in the i -th subpopulation. The length of time between updates corresponds to the infection duration. After this infectious period, infected individuals become susceptible again. Individuals in subpopulation i experience a mortality rate of μ_{t+1}^i over the time interval $(t, t + 1]$. Thus, the fraction of individuals surviving in this subpopulation is $\exp(-\mu_{t+1}^i)$. The susceptible individuals that survive in subpopulation i become infected with a net force of infection $\sum_{j=1}^n \beta^{ij} I_t^j$ where β^{ij} is the per-capita force of infection from subpopulation j to subpopulation i . If infections are Poisson distributed, the fraction of susceptible individual escaping infection is $\exp(-\sum_{j=1}^k \beta^{ij} I_t^j)$. Under these assumptions, the SIS dynamics are given by the following k equations:

$$I_{t+1}^i = (N^i - I_t^i)e^{-\mu_{t+1}^i} \left(1 - \exp\left(-\sum_{j=1}^k \beta^{ij} I_t^j\right) \right) \quad i = 1, 2, \dots, k. \quad (5.11)$$

To complete the model, we assume that the mortality rates are given by a first order auto-regressive process

$$\mu_{t+1} = A\mu_t + \xi_{t+1} \quad (5.12)$$

where $\mu_t = (\mu_t^1, \dots, \mu_t^k)$ is the vector of mortality rates, A is a non-negative matrix with a spectral radius < 1 and ξ_t are i.i.d. random, non-negative vectors taking values in a compact set.

To place (5.11)–(5.12) into our framework, define X_t to be the total density of infected and F_t^i to be the fraction of total infected in subpopulation i :

$$X_t = \sum_{i=1}^k I_t^i \quad \text{and} \quad F_t^i = I_t^i / X_t.$$

If we define $\phi(x) = (1 - e^{-x})/x$, then for $X_t > 0$ the dynamics of (5.11)–(5.12) are equivalent to

$$\begin{aligned} X_{t+1} &= \sum_{i=1}^k (N^i - F_t^i X_t) e^{-\mu_{t+1}^i} \left(\sum_{j=1}^k \beta^{ij} F_t^j X_t \right) \phi \left(\sum_{j=1}^k \beta^{ij} F_t^j X_t \right) \\ F_{t+1}^i &= (N^i - F_t^i X_t) e^{-\mu_{t+1}^i} \left(\sum_{j=1}^k \beta^{ij} F_t^j X_t \right) \phi \left(\sum_{j=1}^k \beta^{ij} F_t^j X_t \right) / X_{t+1} \\ \mu_{t+1} &= A\mu_t + \xi_{t+1}. \end{aligned} \tag{5.13}$$

As $\lim_{x \rightarrow 0} \phi(x) = 1$, for $X_t = 0$, the dynamics extend continuously to

$$\begin{aligned} X_{t+1} &= 0 \\ F_{t+1}^i &= N^i e^{-\mu_{t+1}^i} \left(\sum_{j=1}^n \beta^{ij} F_t^j \right) / \left(\sum_{j,\ell=1}^k e^{-\mu_{t+1}^j} N^j \beta^{j\ell} F_t^\ell \right) \\ \mu_{t+1} &= A\mu_t + \xi_{t+1}. \end{aligned} \tag{5.14}$$

Thus, (5.11)–(5.12) can be written in the desired form (2.1) with $Y_t = (F_t, \mu_t)$. The state space for these dynamics is $\mathcal{S} = [0, \sum_i N^i] \times \Delta$ where $\Delta = \{x \in [0, 1]^k : \sum_{i=1}^k x^i = 1\}$ is the probability simplex. The extinction set is $\mathcal{S}_0 = \{0\} \times \Delta$.

The F_t dynamics on the extinction set \mathcal{S}_0 correspond to projection of the following stochastic linear difference equation onto the simplex (i.e. $F_t = V_t / \sum_i V_t^i$)

$$V_{t+1} = \text{diag}(e^{-\mu_{t+1}^i} N^i) B V_t \tag{5.15}$$

where $B = (\beta^{ij})_{i,j}$ and $\text{diag}(a_i)$ denotes a diagonal matrix with entries a_1, \dots, a_n . If B is a primitive matrix, then the random Perron Frobenius Theorem (see, e.g., Roth and Schreiber 2014b, Proposition 8.3) implies that F_t converges to a unique stationary distribution \widehat{F} . Let μ be the product of the Dirac measure at $\{0\}$ and the law of \widehat{F} . Then $r(\mu)$ corresponds to the dominant Lyapunov exponent of (5.15). If $r(\mu) > 0$, Theorem 1 implies that the disease persists in probability and almost surely. Conversely if $r(\mu) < 0$, then Theorem 2 implies that the disease is unlikely to establish whenever arriving at low densities.

In general, finding explicit expression for this dominant Lyapunov exponent $r(\mu)$ is challenging. However, in the special case, of perfect mixing (i.e. $B = \beta J/n$ where J is the matrix of ones and $\beta > 0$), work of Metz et al. (1983) implies that

$$r(\mu) = \mathbb{E} \left[\log \left(\frac{1}{k} \sum_{i=1}^k \beta N^i e^{-\mu^i} \right) \right]. \tag{5.16}$$

In the special case of a single subpopulation (say i), $r(\mu) = \mathbb{E}[\log(\beta N^i e^{\mu^i})]$ and persistence requires $\log(\beta N^i) > \mathbb{E}[\mu^i]$. Even if persistence within any subpopulation

isn't possible, the coupling of the subpopulations by infection can allow for persistence. To see why, applying Jensen's inequality to the inner and outer averages of (5.16) and assuming there is variation in the μ_t^i , we get

$$\log \left(\frac{1}{k} \sum_{i=1}^k \beta N^i \mathbb{E} \left[e^{-\mu_t^i} \right] \right) > r(\mu) > \frac{1}{k} \sum_{i=1}^k \mathbb{E} \left[\log \left(\beta N^i e^{-\mu_t^i} \right) \right]. \quad (5.17)$$

The second equality in this equation implies that $r(\mu)$ for the structured population is strictly greater than the averaged realized per-capita growth rate for the isolated subpopulations.

Biological implications. *The first inequality in (5.17) implies that disease persistence is possible for the entire population despite it being unable to persist within any subpopulation. In a fully mixed population this only occurs if at least one of arithmetically averaged reproductive numbers $\beta N_i \mathbb{E}[e^{-\mu_t^i}]$ is greater than one. However, even if all the arithmetically averaged reproductive numbers are less than one, then an approximation developed in Schreiber (2010) implies that imperfect mixing and positive temporal autocorrelations in the mortality rates μ_t^i can allow for disease persistence.*

6 Discussion

We developed criteria for stochastic persistence and exclusion for stochastic, discrete-time multispecies models that allow for internal and external variables. Our criteria are based on species realized per-capita growth rates when rare (Turelli 1981; Chesson 1982; Chesson and Ellner 1989; Chesson 1994) evaluated at ergodic probability measures supporting a subset of species. For the non-supported species, these realized per-capita growth rates correspond to Lyapunov exponents (Ferriere and Gatto 1995) that characterize the rate of growth of infinitesimal additions of the missing species. As pioneered by Hofbauer (1981) for ordinary differential equations, our stochastic persistence criterion requires the existence of positive weights associated with each species such that this weighted combination of realized per-capita growth rates, what we call the realized community per-capita growth rate, is positive for all ergodic probability measures supporting a subset of species. When this occurs, we show there is persistence from the ensemble point of view (i.e. persistence in probability) and the typical trajectory point of view (i.e. almost-sure persistence). These results extend prior results (Schreiber et al. 2011; Schreiber 2012; Roth and Schreiber 2014b) for stochastic difference equations in four ways. First, we provide explicit lower bounds for the fraction of time trajectories spend near the boundary. Second, we prove persistence in probability for models with population structure without the additional accessibility assumption made in Schreiber et al. (2011) for unstructured population models. Moreover, we provide explicit lower bounds for the probability of any species being below a low density far into the future. Third, our results provide explicit estimates for the time required to escape a neighborhood of the extinction set. This time to escape decreases with the realized community per-capita growth rate. Finally, our

results account for more types of internal variables than the multispecies, non-linear matrix models of Roth and Schreiber (2014b) as we illustrated with our example of trait evolution.

No understanding of persistence is complete without sufficient conditions for asymptotic exclusion of one or more species. Prior results about exclusion for stochastic difference equations have been limited to scalar models (Chesson 1982; Gyllenberg et al. 1994; Roth and Schreiber 2014b, a) or monotone competing species models in i.i.d. environments (Chesson and Ellner 1989). Here we introduced two types of extinction results. The first result is the natural complement of our persistence theorem: if the community per-capita growth rate is negative for all ergodic probability measures supporting a subset of species, then the trajectories of communities starting near the extinction set are likely to asymptotically converge to this extinction set. To our knowledge, this is the first time for stochastic models such a criterion has been proven. This criterion is most useful for showing the entire extinction set is a stochastic attractor, as illustrated with the evolutionary game of rock–paper–scissors. For many other systems, however, the extinction set is not a stochastic attractor despite between some species being extinction prone. For these models, we proved a discrete-time analog of an extinction result due to Hening and Nguyen (2018a) for unstructured Kolmogorov stochastic differential equation models. This result requires identifying a subset of species which satisfy the Hofbauer condition for stochastic persistence but for which all missing species have negative realized per-capita growth rates. As with our persistence results, we provide explicit lower bounds for the probability of extinction in terms of the realized per-capita growth rates.

While our persistence and exclusion results are applicable to many types of models, several major mathematical challenges remain. First and foremost, as with the deterministic theory, there is a gap between the conditions for stochastic persistence (Theorem 1) and exclusion (Theorems 2 and 3). In the deterministic persistence theory, this gap has been partially closed by considering Morse decomposition of the extinction set (i.e. the dynamics are gradient-like when collapsing a finite number of invariant sets to points) and allowing the species weights $p^i > 0$ to be chosen separately for each of the invariant sets of the Morse decomposition (Schreiber 2000; Garay and Hofbauer 2003; Hofbauer and Schreiber 2010; Patel and Schreiber 2018). Having a similar result for stochastic models would further increase the applicability of the persistence and exclusion theory. Second, our results assume the dynamics remain on a compact set. While this assumption is biologically realistic, theoretical population biologists do build models that violate this assumption e.g. using lognormal random variables on $[0, \infty)^n$. Hence, extending our results to this setting would be useful and should be possible with the methods developed by Hening and Nguyen (2018a); Benaïm (2018). Third, often evaluating the persistence or extinction criteria for particular models can be exceptionally challenging. Hence, it will be useful to identify special classes of models beyond the stochastic Lotka–Volterra difference equations for which they can be explicitly evaluated. Moreover, the further development of approximation methods, such as the small-noise approximations of Chesson (1994), is needed. Fourth, in deterministic theory, coexistence can be equated with the existence of a positive attractor (Schreiber 2006) that need not be globally attracting i.e. need not be a permanent system (Hofbauer and Sigmund 1998). Developing meth-

ods to identify when there are positive stochastic attractors potentially coexisting with stochastic attractors on the extinction set will require going beyond realized per-capita growth rates (see, e.g., Roth and Schreiber (2014a) for single species models).

Acknowledgements This work was supported in part by United States National Science Foundation Grant DMS-1716803 to SJS and Swiss National Science Foundation Grant 2000021-175728 to MB. We thank three anonymous reviewers for comments on an earlier draft of their paper.

7 Proofs

7.1 Proof of Proposition 1

We begin with the following fundamental lemma. Recall ξ_t take values in a polish space \mathcal{E} and have a common law $m(d\xi)$.

Lemma 3 *Let $g : \mathcal{S} \times \mathcal{E} \mapsto \mathbb{R}$ be a measurable map such that $\sup_{z \in \mathcal{S}} \int g(z, \xi)^2 m(d\xi) < \infty$. Define $\bar{g}(z) = \int g(z, \xi) m(d\xi)$. Then*

(i) *For all $z \in \mathcal{S}$ and $Z_0 = z$*

$$\lim_{t \rightarrow \infty} \frac{\sum_{s=0}^{t-1} g(Z_s, \xi_{s+1}) - \sum_{s=0}^{t-1} \bar{g}(Z_s)}{t} = 0 \text{ with probability one.}$$

(ii) *Let μ be an invariant (respectively ergodic) probability measure for (Z_t) , then there exists a bounded measurable map \widehat{g} such that with probability one and for μ -almost every z*

$$\lim_{t \rightarrow \infty} \frac{\sum_{s=0}^{t-1} g(Z_s, \xi_{s+1})}{t} = \lim_{t \rightarrow \infty} \frac{\sum_{s=0}^{t-1} \bar{g}(Z_s)}{t} = \widehat{g}(z) \text{ when } Z_0 = z.$$

Furthermore

$$\begin{aligned} \int \bar{g}(z) \mu(dz) &= \int \widehat{g}(z) \mu(dz) \text{ (respectively } \widehat{g}(z) \\ &= \int \bar{g}(z) \mu(dz) \text{ } \mu - \text{almost surely)}. \end{aligned}$$

Proof To prove the first assertion, let $C = \sup_{z \in \mathcal{S}} \int g(z, \xi)^2 m(d\xi)$ and define the martingale

$$M_t = \sum_{s=0}^{t-1} (g(Z_s, \xi_{s+1}) - \bar{g}(Z_s)).$$

with respect to the sigma-algebra \mathcal{F}_t generated by $(Z_0, \xi_0), \dots, (Z_t, \xi_t)$. M_t has square integrable martingale differences as $\mathbb{E}[(M_t - M_{t-1})^2 | \mathcal{F}_{t-1}] \leq 2C^2$. Define the previsible increasing process $\langle M \rangle_t$ by $\langle M \rangle_0 = 0$ and $\langle M \rangle_t = \langle M \rangle_{t-1} +$

$\mathbb{E}[(M_t - M_{t-1})^2 | \mathcal{F}_{t-1}]$. $\langle M \rangle_t$ is known as the angle-brackets process (Williams 1991, Sect. 12.12). By construction, we have $\langle M \rangle_t \leq 2C^2t$. Define $\langle M \rangle_\infty = \lim_{t \rightarrow \infty} \langle M \rangle_t$, which exists (possibly infinite) as $\langle M \rangle_t$ is increasing. On the event $\langle M \rangle_\infty < +\infty$, Williams (1991, Theorem 12.13a) implies that $\lim_{t \rightarrow \infty} M_t$ exists and is finite. In particular, $\lim_{t \rightarrow \infty} M_t/t = 0$ on the event $\langle M \rangle_\infty < +\infty$. On the event $\langle M \rangle_\infty = +\infty$, Williams (1991, Theorem 12.14a) implies that $\lim_{t \rightarrow \infty} M_t/\langle M \rangle_t = 0$. In particular, as $\langle M \rangle_t \leq 2C^2t$, $\lim_{t \rightarrow \infty} M_t/t = 0$ on the event $\langle M \rangle_\infty = +\infty$. Thus, $\lim_{t \rightarrow \infty} M_t/t = 0$ with probability one which completes the proof of the first assertion.

The second assertion follows from Birkhoff’s ergodic theorem applied to stationary Markov Chains (see Meyn and Tweedie (2009), Theorem 17.1.2) □

Equation (3.2) and the second assertions of Proposition 1 follow directly from the preceding lemma applied to $g(z, \xi) = \log f^i(z, \xi)$. Assumption **A4** implies that $\log f^i(z, \xi_t)$ are uniformly integrable (UI). Therefore, equation (3.3) follows from (3.2) and the UI convergence theorem. The first half of the third assertion follows from ergodicity. To prove the second half of the third assertion, let $i \in S(\mu)$. By assertion (i) of Proposition 1

$$\lim_{t \rightarrow \infty} \frac{\log X_t^i}{t} = \widehat{r}^i(z) \quad \text{where } Z_0 = z$$

for μ -almost $z \in \{(x, y) \in \mathcal{S} : x^i > 0\}$. Let $\mathcal{S}^{i,\eta} = \{(x, y) \in \mathcal{S} : x^i \geq \eta\}$ and $\eta^* > 0$ be such that $\mu(\mathcal{S}^{i,\eta}) > 0$ for all $\eta \leq \eta^*$. By Poincaré Recurrence Theorem, for μ almost all $z \in \mathcal{S}^{i,\eta}$

$$\mathbb{P}_x[Z_t \in \mathcal{S}^{i,\eta} \text{ infinitely often}] = 1$$

for $\eta \leq \eta^*$. Thus $\widehat{r}_i(z) = 0$ for μ -almost all $z \in \mathcal{S}^{i,\eta}$ with $\eta \leq \eta^*$. Hence $\widehat{r}^i(z) = 0$ for μ -almost all $z \in \bigcup_{\eta \in \mathbb{N}} \mathcal{S}^{i,1/\eta} = \{(x, y) \in \mathcal{S} : x^i > 0\}$. This proves assertion (iii).

7.2 Proof of Proposition 2

We begin by showing there is $T \geq 1$ and $\alpha \in (0, 1)$ such that inequality (3.5) holds i.e. $P_T V(z) - V(z) \leq -\alpha$ for all $z \in \mathcal{S}_0$. Suppose to the contrary. Then there exists an increasing sequence of times $t_k \uparrow \infty$, a decreasing sequence of positive reals $\alpha_k \downarrow 0$ and a sequence of points z_k in \mathcal{S}_0 such that $P_{t_k} V(z_k) - V(z_k) \geq -\alpha_k$ for all k . Define a sequence of Borel probability measures μ_k on \mathcal{S}_0 by

$$\int h(z)\mu_k(dz) = \frac{1}{t_k} \mathbb{E}_{z_k} \left[\sum_{\tau=0}^{t_k-1} h(Z_\tau) \right] \text{ for any continuous } h : \mathcal{S}_0 \rightarrow \mathbb{R}.$$

Let μ be a weak* limit point of the μ_i , which exists as \mathcal{S}_0 is compact. By a standard argument due to Khasminskii (see, e.g., Kifer (1988, Theorem 1.1)), μ is an invariant

measure for the Markov chain and by weak* compactness, $\sum_i p^i r^i(\mu) \leq 0$. By the ergodic decomposition theorem, there exists an ergodic measure ν supported on \mathcal{S}_0 such that inequality (3.4) is violated; a contradiction.

Let $\phi(x) = e^x - 1 - x$. For any real C ,

$$|\phi(-\theta C)| \leq \sum_{k \geq 2} \frac{|\theta C|^k}{k!} \leq \theta^2 e^C \quad \text{whenever } |\theta| \leq 1.$$

Our assumption that $\sup_{z, \xi} |\log f^i(z, \xi)| < \infty$ implies that there exists $C > 0$ such that

$$\left| \sum_{\tau=0}^{T-1} \sum_i p^i \log f^i(z_\tau, \xi_\tau) \right| \leq C$$

for any $z_0, \dots, z_{T-1} \in \mathcal{S}_0$ and $\xi_1, \dots, \xi_T \in \mathcal{E}$. Thus, for $z \in \mathcal{S} \setminus \mathcal{S}_0$

$$\begin{aligned} P_T V_\theta(z) &= \mathbb{E}_z \left[\exp\left(-\theta \sum_i p^i \log X_T^i\right) \right] \\ &= \mathbb{E}_z \left[\exp\left(-\theta \sum_i p^i \left(\sum_{\tau=0}^{T-1} \log f^i(Z_\tau, \xi_{\tau+1}) + \log x^i\right)\right) \right] \\ &= \mathbb{E}_z \left[\exp\left(-\theta \sum_i p^i \sum_{\tau=0}^{T-1} \log f^i(Z_\tau, \xi_{\tau+1})\right) \right] V_\theta(z) \\ &= \mathbb{E}_z \left[1 - \theta \sum_i p^i \sum_{\tau=0}^{T-1} \log f^i(Z_\tau, \xi_{\tau+1}) \right. \\ &\quad \left. + \phi\left(-\theta \sum_i p^i \sum_{\tau=0}^{T-1} \log f^i(Z_\tau, \xi_{\tau+1})\right) \right] V_\theta(z) \\ &\leq (1 - \theta\alpha + \theta^2 e^C) V_\theta(z). \end{aligned}$$

Choosing $\theta = \alpha e^{-C}/2$ and $\rho = 1 - \theta\alpha/2$ completes the proof of the proposition.

7.3 Proof of Theorem 1

In this section, we provide the details of the proof of Theorem 1 that are not presented in the main text. Through out this section, we assume that (3.4) holds. Let $\theta > 0$, $T \geq 1$, $\rho \in (0, 1)$, and $\beta > 0$ be as given in Proposition 2.

We begin proving persistence in probability for the general case of $T \geq 1$. Given any $z_0 \in \mathcal{S} \setminus \mathcal{S}_0$, any integer $t = \ell T + s \geq 1$ with $\ell \geq 0$ and $0 \leq s \leq T - 1$, and $\eta > 0$, Proposition 2 implies

$$\begin{aligned} \mathbb{P}_{z_0} [Z_t \in \mathcal{S}_\eta] \min_{z \in \mathcal{S}_\eta \setminus \mathcal{S}_0} V_\theta(z) &\leq \int_{\mathcal{S}} V_\theta(z) (\delta_{z_0} P_{\ell T+s})(dz) \\ &\leq \rho^\ell \int_{\mathcal{S}} V_\theta(z) (\delta_{z_0} P_s)(dz) + \frac{\beta}{1-\rho} \\ &\leq \rho^\ell \max_{0 \leq \tau \leq T-1} \int_{\mathcal{S}} V_\theta(z) (\delta_{z_0} P_\tau)(dz) + \frac{\beta}{1-\rho} \end{aligned}$$

As $\rho^\ell \rightarrow 0$ as $t \rightarrow \infty$, inequality (3.7) implies

$$\limsup_{t \rightarrow \infty} \mathbb{P}_{z_0} [Z_t \in \eta] \leq a\eta^b \text{ where } a = \frac{\beta}{a_0(1-\rho)}$$

which completes the proof of persistence in probability.

To complete the proof of almost-sure persistence presented in the main text, we need the following lemma which follows the strategy of proof of Schreiber et al. (2011, Lemma 6).

Lemma 4 *For all $z \in \mathcal{S} \setminus \mathcal{S}_0$, the weak* limit points μ of Π_t almost surely satisfy $\mu(\mathcal{S}_0) = 0$.*

Proof The process $\{Z_t\}_{t=0}^\infty$ being a (weak) Feller Markov chain over a compact set \mathcal{S} implies that the set of weak* limit points of $\{\Pi_t\}_{t=0}^\infty$ is almost surely a non-empty compact subset of probability measures supported by \mathcal{S} . Almost sure invariance of these weak* limit points follows from Lemma 3 (i).

Assertion (i) of Lemma 3 applied to $g(z, \xi) = \sum_i p^i \log f^i(z, \xi)$ gives we have

$$\lim_{t \rightarrow \infty} \frac{\sum_i p^i \log X_t^i - \sum_i p^i \log x^i - \sum_{s=0}^{t-1} \sum_i p^i \overline{\log f^i}(Z_s)}{t} = 0$$

where $\overline{\log f^i}(z) = \int \log f^i(z, \xi) m(d\xi)$. Since $\limsup_{t \rightarrow \infty} \frac{1}{t} \sum_i p^i (\log X_t^i - \log x^i) \leq 0$ almost surely, we get that

$$\sum_i p^i r^i(\mu) \leq 0 \tag{7.1}$$

almost surely for any weak* limit point μ of $\{\Pi_t\}_{t=0}^\infty$.

Since \mathcal{S}_0 and $\mathcal{S} \setminus \mathcal{S}_0$ are invariant, there exists $\alpha \in (0, 1]$ such that $\mu = (1 - \alpha)v_0 + \alpha v_1$ where v_0 is an invariant probability measure with $v_0(\mathcal{S}_0) = 1$ and v_1 is an invariant probability measure with $v_1(\mathcal{S}_0) = 0$. By Proposition 1, $\sum_i p^i r^i(v_1) = 0$. Thus, $(1 - \alpha) \sum_i p^i r^i(v_0) \leq 0$. Since by assumption $\sum_i p^i r^i(v_0) > 0$, α must be 1. \square

It follows that with probability one, the weak* limit points μ of Π_t (given $Z_0 = z \in \mathcal{S} \setminus \mathcal{S}_0$) are invariant probability measures satisfying $\mu(\mathcal{S}_0) = 0$. To complete, the proof of almost-sure persistence, we need to provide uniform upperbounds to the amount of weight that invariant probability measures μ with $\mu(\mathcal{S}_0) = 0$ place near \mathcal{S}_0 .

To this end, let μ be an invariant probability measure with $\mu(\mathcal{S}_0) = 0$. As, in general, we can not assume that $\int V_\theta(z)\mu(dz)$ is well-defined and finite, we present a slightly longer argument than shown in the main text, to deal with this issue. This argument follows Hairer (2006, Proposition 4.24). Let $M > 0$ be any positive real. For two real numbers a, b , let $a \wedge b$ denote $\min\{a, b\}$. Then invariance, Jensen's inequality and Proposition 2 imply

$$\begin{aligned} \int (V_\theta \wedge M)(z)\mu(dz) &= \int (P_T(V_\theta \wedge M))(z)\mu(dz) \\ &= \int \mathbb{E}_z[V_\theta(Z_T) \wedge M]\mu(dz) \leq \int \mathbb{E}_z[V_\theta(Z_T)] \wedge M\mu(dz) \\ &\leq \int (\rho V_\theta(z) + \beta) \wedge M\mu(dz). \end{aligned}$$

Iterating this inequality $k \geq 1$ times yields

$$\int (V_\theta \wedge M)(z)\mu(dz) \leq \int ((\rho)^k V_\theta(z) + \beta/(1 - \rho)) \wedge M\mu(dz). \tag{7.2}$$

By the dominated convergence theorem, taking the limit $k \rightarrow \infty$ yields

$$\int (V_\theta \wedge M)(z)\mu(dz) \leq \frac{\beta}{1 - \rho}. \tag{7.3}$$

By the dominated convergence theorem, taking the limit $M \rightarrow \infty$ yields

$$\int V_\theta(z)\mu(dz) \leq \frac{\beta}{1 - \rho}. \tag{7.4}$$

Thus, as in the main text, for any $\eta > 0$,

$$\mu(\mathcal{S}_\eta) \min_{z \in \mathcal{S}_\eta} V_\theta(z) \leq \int V_\theta(z)\mu(dz) \leq \frac{\beta}{1 - \rho}.$$

Inequality (3.7) implies that

$$\mu(\mathcal{S}_\eta) \leq a(\eta)^b \text{ for all } \eta \leq 1 \text{ where } a = \frac{\beta}{a_0(1 - \rho)}$$

which completes the proof of almost-sure persistence.

7.4 Proof of Theorem 2 and Corollary 1

The strategy of this proof is based on the proof of Theorem 3.3 by Benaïm and Lobry (2016). Define $V(x, y) = \sum_i p^i \log x^i$. As in the proof of Theorem 1, assumption (4.1)

and Proposition 1 imply there exists $T \geq 1, \eta > 0$ and $\alpha > 0$ such that

$$P_T V(z) - V(z) \leq -\alpha \quad \text{for all } z \in \mathcal{S}_\eta. \tag{7.5}$$

Moreover, using the same argument as found in Proposition 2, there exists $\rho \in (0, 1)$ and $\theta > 0$ such that (choosing $\eta > 0$ to be smaller if necessary)

$$P_T V_\theta(z) \leq \rho V_\theta(z) \quad \text{for all } z \in \mathcal{S}_\eta \text{ where } V_\theta(z) = \exp(\theta V(z)). \tag{7.6}$$

Define the stopping time $\tau = \inf\{k : Z_{kT} \notin \mathcal{S}_\eta\}$ and the event $\mathcal{A} = \{\limsup_{t \rightarrow \infty} V(Z_t)/t \leq -\alpha\}$. We will show that there exists a function $q(\varepsilon) \in (0, \eta)$ such that $\mathbb{P}_z(\mathcal{A}) \geq q(\varepsilon)$ for $z \in \mathcal{S}_\varepsilon$ and $q(\varepsilon) \uparrow 1$ as $\varepsilon \downarrow 0$. To this end, define $W_k = V_\theta(Z_{kT})$. Equation(7.6) implies that $W_{k \wedge \tau}$ is a super martingale. Hence, if we define $C(\varepsilon) = \sup_{z \in \mathcal{S}_\varepsilon} \exp(V(z))$ for any $\varepsilon > 0$ (note that $C(\varepsilon) \downarrow 0$ as $\varepsilon \downarrow 0$), then

$$\mathbb{E}[W_{k \wedge \tau} \mathbf{1}_{\tau < \infty}] \leq W_0 \leq C(\varepsilon)^\theta \quad \text{whenever } Z_0 = z \in \mathcal{S}_\varepsilon.$$

Taking the limit as $k \rightarrow \infty$ and defining $D = \min_{z \in \mathcal{S} \setminus \mathcal{S}_\eta} \exp(V(z)) > 0$, the dominated convergence theorem implies that

$$\mathbb{P}_z[\tau < \infty] D^\theta \leq \mathbb{E}_z[W_\tau \mathbf{1}_{\tau < \infty}] \leq C(\varepsilon)^\theta \quad \text{whenever } z \in \mathcal{S}_\varepsilon.$$

Thus,

$$\mathbb{P}_z[\tau = \infty] \geq 1 - \left(\frac{C(\varepsilon)}{D}\right)^\theta =: q(\varepsilon) \quad \text{whenever } z \in \mathcal{S}_\varepsilon$$

and where $q(\varepsilon) \uparrow 1$ as $\varepsilon \downarrow 0$.

Next, consider the martingale, $M_n = \sum_{\ell=1}^n V(\ell T) - P_T V((\ell - 1)T)$. By the strong law for martingales and inequality (7.5), $\limsup_{n \rightarrow \infty} V(nT)/n \leq -\alpha$ on the event $\tau = \infty$. As we have assumed that $|\log f^i|$ is uniformly bounded on $\mathcal{S} \times \mathcal{E}$, $\limsup_{t \rightarrow \infty} V(t)/t \leq -\alpha$ on the event $\tau = \infty$. Thus, as claimed, we have shown that

$$\mathbb{P}_z[\mathcal{A}] \geq q(\varepsilon) \quad \text{for } z \in \mathcal{S}_\varepsilon.$$

As $V(Z_t) \downarrow -\infty$ implies that $\text{dist}(Z_t, \mathcal{S}_0) \downarrow 0$, the proof of Theorem 2 is complete.

To prove Corollary 1, define the event

$$\mathcal{E} = \left\{ \lim_{t \rightarrow \infty} \text{dist}(Z_t, \mathcal{S}_0) = 0 \right\}.$$

Choose $\varepsilon > 0$ such that $q(\varepsilon) > 1/2$. Define the stopping time

$$\tilde{\tau} = \inf\{t \geq 0 : Z_t \in \mathcal{S}_\varepsilon\}.$$

Since \mathcal{S}_0 is accessible, there exists $\gamma > 0$ such that $\mathbb{P}_z[\tilde{\tau} < \infty] > \gamma$ for all $z \in \mathcal{S}$. The strong Markov property implies that for all $z \in \mathcal{S}$

$$\begin{aligned} \mathbb{P}_z[\mathcal{E}] &= \mathbb{E}_z[\mathbb{P}_{Z_{\tilde{\tau}}}[\mathcal{E}]\mathbf{1}_{\{\tilde{\tau} < \infty\}}] + \mathbb{E}_z[\mathbb{P}_{Z_{\tilde{\tau}}}[\mathcal{E}]\mathbf{1}_{\{\tilde{\tau} = \infty\}}] \\ &= \mathbb{E}_z[\mathbb{P}_{Z_{\tilde{\tau}}}[\mathcal{E}]\mathbf{1}_{\{\tilde{\tau} < \infty\}}] \geq \gamma/2. \end{aligned}$$

Let \mathcal{F}_t be the σ -algebra generated by $\{Z_1, \dots, Z_t\}$. The Lévy zero-one law implies that for all $z \in \mathcal{S}$, $\lim_{t \rightarrow \infty} \mathbb{E}_z[\mathbf{1}_{\mathcal{E}}|\mathcal{F}_t] = \mathbf{1}_{\mathcal{E}}$ almost surely. On the other hand, the Markov property implies that $\mathbb{E}_z[\mathbf{1}_{\mathcal{E}}|\mathcal{F}_t] = \mathbb{E}_z[\mathbb{P}_{Z_t}[\mathcal{E}]] \geq \gamma/2$ for all $z \in \mathcal{S}$. Hence $\mathbb{P}_z[\mathcal{E}] = 1$ for all $z \in \mathcal{S}$.

7.5 Proof of Theorem 3

This proof follows the strategy of Theorem 2, but using a V function introduced by Hening and Nguyen (2018a).

Let $I \subset \{1, 2, \dots, n\}$ be the subset of species and $\{p^i\}_{i \in I}$ the set of positive reals such that $\sum_{i \in I} p^i r^i(\mu) > 0$ for every ergodic μ supported on $\mathcal{S}_0^I = \{z = (x, y) \in \mathcal{S}^I : \prod_{i \in I} x^i = 0\}$ where $\mathcal{S}^I = \{z = (x, y) \in \mathcal{S} : x^i = 0 \text{ for all } i \notin I\}$. Assume that $r^i(\mu) < 0$ for all $i \notin I$ and ergodic μ such that $\mu(\mathcal{S}_+^I) = 1$ where $\mathcal{S}_+^I = \mathcal{S}^I \setminus \mathcal{S}_0^I$. Choose $\delta > 0$ and $\alpha > 0$ such that $-\sum_{i \in I} p^i r^i(\mu) + \delta \max_{i \notin I} r^i(\mu) \leq -2\alpha$ for all ergodic probability measures μ such that $\mu(\mathcal{S}^I) = 1$.

Define $V(x, y) = -\sum_{i \in I} p^i \log x^i + \delta \max_{i \notin I} \log x^i$ and

$$V_\theta(x, y) = e^{\theta V(x, y)} = \left(\prod_{i \in I} (x^i)^{-\theta p^i} \right) \max_{i \notin I} (x^i)^{\theta \delta}.$$

As in the proof of Theorem 1 and Proposition 1, there exists $T \geq 1, \eta \in (0, 1]$ such that

$$P_T V(z) - V(z) \leq -\alpha \text{ for all } z \in K_\eta \tag{7.7}$$

where $K_\eta = \{(x, y) \in \mathcal{S} : \max_{i \notin I} x^i \leq \eta\}$. Moreover, using the same argument as found in Proposition 2, there exists $\rho \in (0, 1)$ and $\theta > 0$ such that (choosing $\eta > 0$ to be smaller if necessary)

$$P_T V_\theta(z) \leq \rho V_\theta(z) \text{ for all } z \in K_\eta \setminus \mathcal{S}_0^I. \tag{7.8}$$

Choose $\tilde{\eta} > 0$ such that $\{z : V_\theta(z) \leq \tilde{\eta}\} \subset K_\eta$. Define the stopping time $\tau = \{k : V_\theta(Z_{kT}) \geq \tilde{\eta}\}$ and the event $\mathcal{A} = \{\limsup_{t \rightarrow \infty} V(Z_t)/t \leq -\alpha\}$. Define $W_k = V_\theta(Z_{kT})$ and $\tilde{W}_k = \tilde{\eta} \wedge W_k$. Equation (7.8) implies that $W_{k \wedge \tau}$ is a super martingale. Thus, for any k , concavity of $t \mapsto \delta \wedge t$ and Jensen's inequality implies that

$$\begin{aligned} \mathbb{E}[\tilde{W}_{k \wedge \tau} \mathbf{1}_{\tau < \infty}] &\leq \tilde{\eta} \wedge \mathbb{E}[W_{k \wedge \tau} \mathbf{1}_{\tau < \infty}] \\ &\leq \tilde{\eta} \wedge W_0 = \tilde{\eta} \wedge \frac{\max_{i \notin I} (x^i)^{\theta \delta}}{\prod_{i \in I} (x^i)^{\theta p^i}} =: C(z) \\ &\text{whenever } Z_0 = z = (x, y) \in K_\eta \setminus \mathcal{S}_0^I. \end{aligned}$$

Taking the limit as $k \rightarrow \infty$, the dominated convergence theorem implies that

$$\mathbb{P}_z[\tau < \infty] \tilde{\eta} \leq \mathbb{E}_z[\tilde{W}_\tau \mathbf{1}_{\tau < \infty}] \leq C(z) \quad \text{for all } z \in K_\eta \setminus \mathcal{S}_0^I.$$

Thus,

$$\mathbb{P}_z[\tau = \infty] \geq 1 - \frac{C(z)}{\tilde{\eta}} \quad \text{for all } z \in K_\eta \setminus \mathcal{S}_0^I.$$

Next, consider the martingale, $M_n = \sum_{\ell=1}^n V(\ell T) - P_T V((\ell-1)T)$. By the strong law for martingales and (7.7), $\limsup_{n \rightarrow \infty} V(nT)/n \leq -\alpha$ on the event $\tau = \infty$. As we have assumed that $|\log f^i|$ is uniformly bounded on $\mathcal{S} \times \mathcal{E}$, $\limsup_{t \rightarrow \infty} V(t)/t \leq -\alpha$ on the event $\tau = \infty$. Thus, we have shown that

$$\mathbb{P}_z[\mathcal{A}] \geq 1 - \frac{C(z)}{\tilde{\eta}} \quad \text{for all } z \in K_\eta \setminus \mathcal{S}_0^I.$$

Let $M = \sup_{(x,y) \in \mathcal{S}} \max_i x^i > 0$. As $\limsup_{t \rightarrow \infty} -\frac{\log X_t^i}{t} \geq \limsup_{t \rightarrow \infty} -\frac{\log M}{t} = 0$ with probability one whenever $X_0^i > 0$,

$$\begin{aligned} -\alpha &\geq \limsup_{t \rightarrow \infty} \frac{V(Z_t)}{t} = \limsup_{t \rightarrow \infty} \frac{1}{t} \left(-\sum_{i \in I} p^i \log X_t^i + \delta \max_{i \notin I} \log X_t^i \right) \\ &\geq \limsup_{t \rightarrow \infty} \frac{\delta}{t} \max_{i \notin I} \log X_t^i \end{aligned}$$

almost surely on the event \mathcal{A} whenever $Z_0 = z \in K_\eta \setminus \mathcal{S}_0^I$. Hence, on this event, $\lim \text{dist}(Z_t, \mathcal{S}_0^I) \downarrow 0$ and the proof of Theorem 3 is complete.

References

Altizer S, Dobson P, Hosseini A, Hudson P, Pascual M, Rohani P (2006) Seasonality and the dynamics of infectious diseases. *Ecol Lett* 9:467–484

Axelrod R (1984) *The evolution of cooperation*. Basic Books, New York

Axelrod R, Hamilton W (1981) The evolution of cooperation. *Science* 211:1390–1396

Benaïm M (2018) Stochastic persistence, part I. arXiv preprint [arXiv:1806.08450](https://arxiv.org/abs/1806.08450)

Benaïm M, Lobry C (2016) Lotka–Volterra with randomly fluctuating environments or “how switching between beneficial environments can make survival harder”. *Ann Appl Probab* 26:3754–3785

Benaïm M, Schreiber SJ (2009) Persistence of structured populations in random environments. *Theor Popul Biol* 76:19–34

- Benaïm M, Strickler E (2019) Random switching between vector fields having a common zero. *Ann Appl Probab* 29:326–375
- Chesson P (1994) Multispecies competition in variable environments. *Theor Popul Biol* 45:227–276
- Chesson P (2018) Updates on mechanisms of maintenance of species diversity. *J Ecol* 106:1773–1794
- Chesson PL (1982) The stabilizing effect of a random environment. *J Math Biol* 15:1–36
- Chesson PL, Ellner S (1989) Invasibility and stochastic boundedness in monotonic competition models. *J Math Biol* 27:117–138
- Chesson PL (1985) Coexistence of competitors in spatially and temporally varying environments: a look at the combined effects of different sorts of variability. *Theor Popul Biol* 28:263–287
- Chesson PL (2000) General theory of competitive coexistence in spatially-varying environments. *Theor Popul Biol* 58:211–237
- Chesson PL, Warner RR (1981) Environmental variability promotes coexistence in lottery competitive systems. *Am Nat* 117:923–943
- Cuddington K, Wilson WG, Hastings A (2009) Ecosystem engineers: feedback and population dynamics. *Am Nat* 173:488–498
- Diaconis P, Freedman D (1999) Iterated random function. *SIAM Rev* 41:45–76
- Eltahir EAB (1998) A soil moisture-rainfall feedback mechanism: 1. Theory and observations. *Water Resour Res* 34:765–776
- Evans SN, Ralph P, Schreiber SJ, Sen A (2013) Stochastic growth rates in spatio-temporal heterogeneous environments. *J Math Biol* 66:423–476
- Ferriere R, Gatto M (1995) Lyapunov exponents and the mathematics of invasion in oscillatory or chaotic populations. *Theor Popul Biol* 48:126–171
- Garay BM, Hofbauer J (2003) Robust permanence for ecological differential equations, minimax, and discretizations. *SIAM J Math Anal* 34:1007–1039
- Gillespie J (1973) Polymorphism in random environments. *Theor Popul Biol* 4:193–195
- Gillespie JH (1978) A general model to account for enzyme variation in natural populations V. The SAS-CFF model. *Theor Popul Biol* 14:1–45
- Gillespie JH, Turelli M (1989) Genotype-environment interactions and the maintenance of polygenic variation. *Genetics* 121:129–138
- Gyllenberg M, Hognas G, Koski T (1994) Population models with environmental stochasticity. *J Math Biol* 32:93–108
- Hairer M (2006) Ergodic properties of Markov processes. Lectures given at The University of Warwick, Spring. <http://www.hairer.org/notes/Markov.pdf>
- Hening A, Nguyen DH (2018a) Coexistence and extinction for stochastic Kolmogorov systems. *Ann Appl Probab* 28:1893–1942
- Hening A, Nguyen DH (2018b) Stochastic Lotka–Volterra food chains. *J Math Biol* 77:135–163
- Hening A, Nguyen DH, Yin G (2018) Stochastic population growth in spatially heterogeneous environments: the density-dependent case. *J Math Biol* 76:697–754
- Hofbauer J (1981) A general cooperation theorem for hypercycles. *Monatshefte für Mathematik* 91:233–240
- Hofbauer J, Schreiber SJ (2010) Robust permanence for interacting structured populations. *J Differ Eq* 248:1955–1971
- Hofbauer J, Sigmund K (1998) *Evolutionary games and population dynamics*. Cambridge University Press, Cambridge
- Hofbauer J, Sigmund K (2003) *Evolutionary game dynamics*. *Bull Am Math Soc* 40:479–519
- Hofbauer J, Hutson V, Jansen W (1987) Coexistence for systems governed by difference equations of Lotka–Volterra type. *J Math Biol* 25(5):553–570
- Hutchinson GE (1961) The paradox of the plankton. *Am Nat* 95:137–145
- Jones CG, Lawton JH, Shachak M (1994) Organisms as ecosystem engineers. *Oikos* 69:373–386
- Jordan-Cooley WC, Lipcius RN, Shaw LB, Shen J, Shi J (2011) Bistability in a differential equation model of oyster reef height and sediment accumulation. *J Theor Biol* 289:1–11
- Kerr B, Riley M, Feldman M, Bohannan J (2002) Local dispersal promotes biodiversity in a real-life game of rock-paper-scissors. *Nature* 418:171–174
- Kifer Y (1986) *Ergodic theory of random transformations*. Birkhauser, New York
- Kifer Y (1988) Random perturbations of dynamical systems, volume 16 of progress in probability and statistics. Birkhäuser Boston Inc., Boston
- Kirkup B, Riley M (2004) Antibiotic-mediated antagonism leads to a bacterial game of rock–paper–scissors in vivo. *Nature* 428:412–414

- Kuang JJ, Chesson PL (2009) Coexistence of annual plants: generalist seed predation weakens the storage effect. *Ecology* 90:170–182
- Kucharski F, Zeng N, Kalnay E (2013) A further assessment of vegetation feedback on decadal sahel rainfall variability. *Clim Dyn* 40:1453–1466
- Lande R (1976) Natural selection and random genetic drift in phenotypic evolution. *Evolution* 30:314–334
- Lande R, Shannon S (1996) The role of genetic variation in adaptation and population persistence in a changing environment. *Evolution* 50:434–437
- Lewontin RC, Cohen D (1969) On population growth in a randomly varying environment. *Proc Natl Acad Sci USA* 62:1056–1060
- Maynard Smith J (1982) *Evolution and the theory of games*. Cambridge University Press, Cambridge
- Maynard Smith J, Price GR (1973) The logic of animal conflict. *Nature* 246:15–18
- McLaughlin JF, Hellmann JJ, Boggs CL, Ehrlich PR (2002) Climate change hastens population extinctions. *Proc Nat Acad Sci USA* 99:6070–6074
- Metz JAJ, de Jong TJ, Klinkhamer PGL (1983) What are the advantages of dispersing; a paper by Kuno extended. *Oecologia* 57:166–169
- Metz JAJ, Geritz SAH, Meszéna G, Jacobs EJA, van Heerwaarden JS (1996) Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. *Dynamical systems and their applications*. Elsevier, North Holland, pp 147–194
- Meyn SP, Tweedie RL (2009) *Markov Chains and stochastic stability*. Cambridge University Press, Cambridge
- Michod RE (2000) *Darwinian dynamics: evolutionary transitions in fitness and individuality*. Princeton University Press, Princeton
- Moore J, Puckett B, Schreiber SJ (2018) Restoration of eastern oyster populations with positive density dependence. *Ecol Appl* 28:897–909
- Moore JL, Lipcius RN, Puckett B, Schreiber SJ (2016) The demographic consequences of growing older and bigger in oyster populations. *Ecol Appl* 26:2206–2217
- Nahum JR, Harding BN, Kerr B (2011) Evolution of restraint in a structured rock–paper–scissors community. *Proc Nat Acad Sci* 108:10831–10838
- Patel S, Schreiber SJ (2018) Robust permanence for ecological equations with internal and external feedbacks. *J Math Biol* 77:79–105
- Roth G, Schreiber SJ (2014a) Pushed to brink: Allee effects, environmental stochasticity, and extinction. *J Biol Dyn* 8:187–205
- Roth G, Schreiber SJ (2014b) Persistence in fluctuating environments for interacting structured populations. *J Math Biol* 68:1267–1317
- Schoener TW (2011) The newest synthesis: understanding the interplay of evolutionary and ecological dynamics. *Science* 331:426–429
- Schreiber SJ (2000) Criteria for C^r robust permanence. *J Differ Eq* 162:400–426
- Schreiber SJ (2006) Persistence despite perturbations for interacting populations. *J Theor Biol* 242:844–852
- Schreiber SJ (2012) Persistence for stochastic difference equations: a mini-review. *J Differ Eq Appl* 18:1381–1403
- Schreiber SJ, Benaïm M, Atchadé KAS (2011) Persistence in fluctuating environments. *J Math Biol* 62:655–683
- Schreiber SJ (2010) Interactive effects of temporal correlations, spatial heterogeneity, and dispersal on population persistence. *Proc R Soc Biol Sci* 277:1907–1914
- Schreiber SJ (2019) When do factors promoting balanced selection also promote population persistence? a demographic perspective on Gillespie’s SAS-CFF model. arXiv preprint [arXiv:1902.03507](https://arxiv.org/abs/1902.03507)
- Schreiber SJ, Moore J (2018) The structured demography of open populations in fluctuating environments. *Methods Ecol Evol* 9:1569–1580
- Schreiber SJ, Patel S, TerHorst C (2018) Evolution as a coexistence mechanism: does genetic architecture matter? *Am Nat* 191:407–420
- Sinervo B, Lively C (1996) The rock-paper-scissors game and the evolution of alternative male strategies. *Nature* 380:240–243
- Staver AC, Levin SA (2012) Integrating theoretical climate and fire effects on savanna and forest systems. *Am Nat* 180:211–224
- Turelli M (1981) Niche overlap and invasion of competitors in random environments I. Models without demographic stochasticity. *Theoretical Population Biology* 20:1–56

- Vincent TL, Brown JS (2005) Evolutionary game theory, natural selection, and Darwinian dynamics. Cambridge University Press, Cambridge
- Whitlock MC, Davis B (2011) Genetic load. eLS. <https://doi.org/10.1002/9780470015902.a0001787.pub2>
- Williams D (1991) Probability with martingales. Cambridge University Press, Cambridge
- Zeng N, Neelin JD, Lau KM, Tucker CJ (1999) Enhancement of interdecadal climate variability in the sahel by vegetation interaction. *Science* 286:1537–1540

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Michel Benaïm¹ · Sebastian J. Schreiber²

¹ Institut de Mathématiques, Université de Neuchâtel, Rue Emile-Argand, 2000 Neuchâtel, Switzerland

² Department of Evolution and Ecology and Center for Population Biology, University of California, Davis, California 95616, USA