On the Simpson index for the Moran process with random selection and immigration

Arnaud Personne, Arnaud Guillin, Franck Jabot

▶ To cite this version:

HAL Id: hal-01877501
https://hal.science/hal-01877501

ON THE SIMPSON INDEX FOR THE MORAN PROCESS WITH RANDOM SELECTION AND IMMIGRATION

ARNAUD GUILLIN $^\diamond$,   FRANCK JABOT $^\clubsuit$,
AND   ARNAUD PERSONNE $^\diamond$

$^\diamond$ UNIVERSITÉ CLERMONT-AUVERGNE
$^\clubsuit$ IRSTEA

ABSTRACT. Moran or Wright-Fisher processes are probably the most well known model to study the evolution of a population under various effects. Our object of study will be the Simpson index which measures the level of diversity of the population, one of the key parameter for ecologists who study for example forest dynamics. Following ecological motivations, we will consider here the case where there are various species with fitness and immigration parameters being random processes (and thus time evolving). To measure biodiversity, ecologists generally use the Simpson index, who has no closed formula, except in the neutral (no selection) case via a backward approach, and which is difficult to evaluate even numerically when the population size is large. Our approach relies on the large population limit in the "weak" selection case, and thus to give a procedure which enable us to approximate, with controlled rate, the expectation of the Simpson index at fixed time. Our approach will be forward and valid for all time, which is the main difference with the historical approach of Kingman, or Krone-Neuhauser. We will also study the long time behaviour of the Wright-Fisher process in a simplified setting, allowing us to get a full picture for the approximation of the expectation of the Simpson index.

Key words : Simpson index, multidimensional Wright-Fisher process, random selection, random immigration.

CONTENTS

1

1. Introduction

Community ecology has been deeply shaked by the book of Hubbell (2001) [26] that elaborated on the idea that the dynamics of ecological communities might be mainly shaped by random processes. A number of predictions made by this neutral theory of biodiversity have been indeed corroborated by empirical evidence (Hubbell [26], Condit et al.[5], Jabot and Chave [27]). This good performance of neutral models for reproducing empirical patterns has stimulated the mathematical study of neutral ecological models (Etienne [14], Fuk et al [19]), in connection with the rich body of work on evolutionary neutral models (Volkov et al. [39], Ewens [16], Muirhead and Wakeley [32]). More recent empirical evaluations of neutral predictions on tropical forest data have focused on the temporal dynamics of individual populations and have shown that the temporal variance of population sizes was actually larger than the one typically predicted by neutral models (Chisholm et al. [4]). These authors have suggested that this may be due to species-specific responses to the temporal variability of the environment. Subsequent modelling studies have elaborated on this idea (Kalyuzhny et al. [29], Jabot and Lohier [28], Krone and Neuhauser [30] and more mathematically-oriented contributions on this topic have since been made ([25, 38, 9, 10, 7, 8, 20, 11, 23, 6]). All these studies share the same idea that community dynamics is influenced by a species-specific selection coefficient and that this selection coefficient is temporally varying, so that good and bad periods are subsequently experienced by all species within the community.

The aim of this contribution is to provide a comprehensive study of a simple model encapsulating this kind of ideas within a weak selection framework. Subsequent works within a strong selection framework will complement the present study.

Indeed, the simplest model for the evolution of population is surely the Moran model (or its cousin the discrete Wright-Fisher model), in which in

a given population an individual is chosen to die (uniformly in the population) and then a child chooses his parent proportionally to the abundance in the previous population. One may also add immigration, i.e. a probability that the child comes from another community, and selection so that some species (or traits) have a selective advantage. Kalyuzhny et al [29], to bypass the neutrality of the model, chosed to consider immigration and selection as random processes (independent of the Moran system), but which preserves neutrality "in mean". One of the main goal is of course to study the effect of these models on biodiversity and its evolution. There are many ways to measure biodiversity. We will focus here on the Simpson index [36], usually considered in neutral model [15]: it measures the probability that two individuals uniformly chosen may be of the same species. More precisely denoting X^i the number of individuals of species i, S the number of species and J the toal size of the population, the Simpson index is given by

$$\mathcal{S} = \sum_{i=1}^{S} \frac{X^i(X^i - 1)}{J(J - 1)}$$

thus varying (roughly) from 0 to 1, from maximal to minimal diversity. Using backward approach, and Kingman's coalescent, an explicit formula may be given for the (asymptotic in time) Simpson index in the neutral case with immigration as $\frac{1-m}{1+J-2m}$, as otherwise it is 1 as a particular specie will almost surely invade all the population. Note that a closed formula (even for the expectation) of the Simpson index at a given time, is usually not reachable. We will consider in this paper this Moran model with immigration and selection as general random process. As said previously such models were recently considered for example by Griffiths [24, 25], Kalyuznhy et al [29] for a simulation study, but no theoretical framework towards the Simpson index. The backward approach constitutes the works of Krone and Neuhauser [30, 34] leading to new coalescent type processes which are however quite difficult to study and may not give a closed formula for the Simpson index. A recent work by Grieshammer [23] considers a forward in time approach but he does not focus on the Simpson index. Our approach is only forward here. As is often done in population genetics, we will consider the large population approximation. Our first task is then to justify this asymptotic to a Wright-Fisher diffusion process in random environment in the weak immigration and selection case. As a flavour, with only two species, the evolution of the proportion of one species is given by

$$dX_t = m_t(p_t - X_t)dt + s_t X_t(1 - X_t)dt + \sqrt{2X_t(1 - X_t)}dB_t$$

where m_t is the immigration process, p_t the probability that this species is chosen, and s_t the selection advantage. It will be done by the usual martingale method. Another quantitative approach will be considered in [21]. The Simpson index is then a quadratic form involving the proportion of each species, and by Itô's formula it involves higher order term. The equation for the expectation of the Simpson index is therefore not closed. We will

then introduce a quantitative approximation procedure for the expectation of the Simpson index, in the quenched case (corresponding to a given random environment) and in some particular case in the annealed case for two or more species. We will also study in some simple case (constant parameter) the long time behaviour of the Simpson index. It is reminiscent with the very recent work of Coron, Méléard and Villemonais [6] (discovered while finishing this work). Very schematically our approach is the following

(1) approximate the true discrete process by a SDE, i.e. Wright-Fisher process;
(2) approximate the expectation of the Simpson index for the Wright-Fisher process by a deterministic ODE;
(3) approximate the infinite time expectation of the Simpson index by the finite time, through evaluation of the speed of convergence towards equilibrium of the Wright-Fisher process.

In Section 2, we introduce the Moran model in random environment and prove its convergence in the large population limit. In Section 3, we focus on the two species case where the approximation of the Simpson index is studied in the quenched case as well as its long time behaviour. Section 4 generalizes to a large number of species and also considers the annealed case when the selection parameter has a particular form (a variant of a Wright-Fisher process). The last section contains technical proofs or recall some known results for the Wright-Fisher process.

2. THE MORAN MODEL IN RANDOM ENVIRONMENT AND ITS APPROXIMATION IN LARGE POPULATION

2.1. Discrete model with selection and immigration.

In this section we describe in detail the discrete model, i.e. the Moran process, which is the basic of our study. One may also consider here the Wright-Fisher discrete process with adequat change. The Moran process is an evolution of population model, in which a single event occurs at each time step. More precisely each event corresponds to the death of an individual and the birth of another who replaces it.

We consider a population, whose size is constant over time equal to J, composed of $S + 1$ species . The proportion of the i species at the nth event is denoted X_n^i, $i \in \mathbb{S} = \{1, ..., S + 1\}$, $n \in \mathbb{N}$.

As usual one we know $(X_n^i)_{i=1,...,S}$, we deduce the proportion for the last species, $X_n^{S+1} = 1 - \sum_{i=1}^{S} X_n^i$

We note X_n the species vector or abundance vector having for coordinate i, X_n^i. The dynamics of evolution follows the following pattern at the step n:

(1) The individual designated to die is chosen uniformly among the community.
(2) The one which replaces it, chooses his parent in the community with probability $1 - m_n$ (filiation) or a parent from the immigration process

with probability m_n (immigration). The quantity m_n varies between 0 and 1, it can be random and time dependent.
(3) If there is immigration the chosen parent is from the species i, with probability p_n^i $i \in \mathbb{S}$. The p_n^i verify $\sum_{i \in \mathbb{S}} p_n^i = 1$ and can be time dependent and random. We note p for the vector having for coordinate i, p_n^i.
(4) In a filiation, the chosen parent is of the species i with probability
$$\frac{X_n^i(1+s_n^i)}{1+\sum_{k=1}^{S+1} X^k s_n^k}.$$
The $s_n^i, i \in \mathbb{S}$ are the selection parameters , they may be time dependent and random. Furthermore, we assume $s_n^{S+1} = 0$. Indeed, we can obtain it from any configuration by changing all the coefficients by $s_n^i = \frac{\tilde{s}_n^i - \tilde{s}_n^{S+1}}{1+\tilde{s}_n^{S+1}}$.

We will assume throughout this work that m_n, p_n, s_n are autonomous, in the sense that their evolution do not depend on $(X_n)_{n \geq 0}$. We will further assume that $(m_n, p_n, s_n)_{n \geq 0}$ is a Markov chain. Note also that m_n, p_n, s_n may also depend in some sense of the size of the population J, but we do not add another superscript to get lighter notations.

This model therefore describes a Markovian dynamic in which selection and immigration play an important role. Immigration already introduced by Hubbell [26] avoids the definitive invasion of the community by a species. Selection changes the dynamics of a species related to the neutral model([29]). The temporal evolution of the population could be simulated numerically from the transition matrix of the Markov system. Let us describe precisely these transition probabilities for the evolution of proportions. The assumption for the dynamics for the immigration and selection will be given later on.
Let x be the vector having for coordinate i, x^i and suppose m_n ,p_n known. Denote $\Delta = \frac{1}{J}$, so for the i species:

$$
\begin{aligned}
P_{x^i+} &= \mathbb{P}(X_{n+1}^i = x + \Delta | X_n = x) \\
&= (1 - x^i)\left(m_n p_n^i + (1 - m_n)\frac{x^i(1 + s_n^i)}{1 + \sum_{k=1}^{S+1} x^k s_n^k}\right),
\end{aligned}
$$

$$P_{x^i-} = \mathbb{P}(X_{n+1}^i = x - \Delta | X_n = x))$$

$$= x^i \left(m_n(1 - p_n^i) + (1 - m_n)\left(1 - \frac{x^i(1 + s_n^i)}{1 + \sum\limits_{k=1}^{S+1} x^k s_n^k}\right)\right),$$

$$P_{x^i+x^j-} = \mathbb{P}(\{X_{n+1}^i = x^i + \Delta\} \cap \{X_{n+1}^j = x^j - \Delta\} | X_n = x)$$

$$= x^j \left(m_n(1 - p_n^i) + (1 - m_n)\frac{x^i(1 + s_n^i)}{1 + \sum\limits_{k=1}^{s+1} x^k s_n^k}\right).$$

With the dynamics of (m_n, p_n, s_n) given, one may of course simulate exactly the vector of proportion X_n and thus evaluate the expectation of the Simpson index, which is what we will do to validate our approximation procedure, but when the population size J is very large, it may be computationally too costly (and even impossible). Thus we will approach the dynamics of this model by a stochastic differential system continuous in time.

2.2. **To a limit in large population.**
In this section, we explain how approaching the dynamics of the preceding model by a diffusion, and associated process for the immigration and selection processes, when J goes to infinity.

We need to define a J dependent time scale. Indeed, when J goes to infinity, the time scale has to change, expectation and variance are about $\frac{1}{J}$ and $\frac{1}{J^2}$, and goes to 0 when J goes to infinity. It corresponds to considering a large number of event for the Markov chain, to obtain a non-trivial convergence of our discrete process towards a limit process, i.e. not look at the event-by-event evolution as we did before but in packets of several events.
Several choices for scales are possible, each one leads to study a different process. We choose to study a continuous multidimensional diffusion in time.

2.2.1. *Diffusion approximation.*
In a general framework, the limiting process we obtain is characterized by the first moment and the covariance matrix of the infinitesimal variation of abundance.
More precisely if we note Δ_t the infinitesimal variation in time (which depends on the scale) and $\Delta X_t = X_{t+\Delta_t} - X_t$ the infinitesimal variation of abundance, the diffusion process is characterized by the quantities:

$$b(x) = \lim_{\Delta_t \Rightarrow 0} \frac{E[\Delta X_t | X_t = x]}{\Delta_t}$$

$$\sigma_{i,j}(x) = \lim_{\Delta_t \Rightarrow 0} \frac{\text{Cov}[\Delta X_{t+\Delta_t}(i), \Delta X_{t+\Delta_t}(j)|X_t = x]}{\Delta_t} \quad i,j \in \mathbb{S}$$

The following property characterizes the order (relative to J) of the expectation, variance, and covariance of the abundance variation of a species during an event:

Proposition 1.   (1) $\mathbb{E}[X_{n+1}^i - X_n^i | X_n = x] = \Delta(P_{x^i+} - P_{x^i-})$
    (2) $Var[X_{n+1}^i - X_n^i | X_n = x] = \Delta^2((P_{x^i+} + P_{x^i-}) - (P_{x^i+} - P_{x^i-})^2)$

The proof is standard calculus and thus omitted. The last property shows that the expectation is of the order of $\frac{1}{J}$ whereas the variance is of the order of $\frac{1}{J^2}$. The choice we make to preserve a stochastic part in our limit equation is to consider the infinitesimal time variation is of the order of $\frac{1}{J^2}$ .Other choices would have led to a Piecewise Deterministic Markov process in which only the parameters s, m, p brings randomness. It will be left for further study.

A scale in $\frac{1}{J^2}$ and the weak selection and immigration.
Let $\Delta_t = \frac{1}{J^2}$ et $n = tJ^2$.
In this case, $\mathbb{E}[X_{t+\Delta_t} - x | X_t = x] = O(J)$ and the limit would be infinite. To hope for a finite term and thus to observe the influence of s and m in our limit DSE, we must assume that s and m are inversely proportional to J. We now assume that migration and speciation are weak.

Proposition 2. *Let $\Delta_t = \frac{1}{J^2}$ and $n = tJ^2$, note $m'(t) = m(t) \times J$ and $s'(t) = s(t) \times J$. So when J goes to infinity:*

(1) $\mathbb{E}[X_{t+\Delta_t}^i - x^i | X_t = x] = \left(m'_t(p_t^i - x^i) + x^i(s_t'^i - \sum_{k \in \mathbb{S}} x^k s_t'^k) \right) \times \Delta_t + o(\Delta_t)$

(2) $\text{Var}[X_{t+\Delta_t}^i - x^i | X_t = x] = 2x^i(1 - x^i)\Delta_t + o(\Delta_t)$

(3) $\text{Cov}[X_{t+\Delta_t}^i, X_{t+\Delta_t}^l | X_t = x] = -2x^i x^l \Delta_t + o(\Delta_t) \quad \forall i \neq l.$

We will now introduce notations and assumptions. Remark that a very recent work by Bansaye et al [3] considered a very general framework for population model convergence in random environment, but in their work the environment is usually i.i.d. whereas we are in a Markovian setting. To get shorter statement and proofs, we will make considerable simplifications for our main assumption.

Assumption (A).
- the process (p_n) is assumed to be constant, which corresponds to a non evolving pool of immigration;
- the process (m^J) is an autonomous Markov chain, and consider its rescaled piecewise linear extension $\tilde{m}_t^J = J\, m_{\lfloor tJ^2 \rfloor}^J$, which is assumed to take values in a finite set E_m and uniformly bounded (in J). Let

denote $P_{m^J}(m, m', t)$ its transition probabilities and assume for all $m \neq m'$,

$$\lim_{J \to \infty} P_{m^J}\left(m, m', \frac{1}{J^2}\right) \times J^2 = Q(m, m')$$

is well defined;

- the process (s^J) is an autonomous Markov chain, and consider its rescaled piecewise linear extension $\tilde{s}_t^J = J\, s_{\lfloor tJ^2 \rfloor}^J$, which is assumed to take values in a finite set E_s and uniformly bounded (in J). Let denote $P_{s^J}(s, s', t)$ its transition probabilities and assume for all $s \neq s'$,

$$\lim_{J \to \infty} P_{s^J}\left(s, s', \frac{1}{J^2}\right) \times J^2 = Q^s(s, s')$$

is well defined.

- The processes (m_n) and (s_n) are supposed to be independent.

These assumptions about the limits of transitions probabilities are intended to ensure the convergence in law of s and m towards Markovian jump processes when J goes to infinity.

Denote again $U_t^J = \begin{pmatrix} X_t^J \\ s_t^J \\ m_t^J \end{pmatrix}$ taking values in $E = E_x \times E_s \times E_m$ a compact

set of $\mathbb{R}^{2S+1}$ and for $\Gamma \in E$, we define for $k \in \mathbb{N}$, $t \in [\frac{k}{J^2}, \frac{k+1}{J^2}[$, $\pi_J(U_t, \Gamma) = \pi_J(U_{\frac{k}{J^2}}, \Gamma) = \mathbb{P}(U_{\frac{k+1}{J^2}}^J \in \Gamma | U_{\frac{k}{J^2}}^J)$.

We are now in position to state our diffusion approximation result.

Theorem 3. *Assume* **(A)** *then when J goes to infinity the sequence of processes (X_t^J) converges in law to the process (X_t) whose coordinates are solutions of the following stochastic differential equation*

$$\begin{pmatrix} dX_t^1 \\ \vdots \\ dX_t^S \end{pmatrix} = \begin{pmatrix} m_t \times (p_t^1 - X_t^1) + X_t^1 \left(s_t^1 - \sum_{k \in \mathbb{S}} X^k s_t^k\right) \\ \vdots \\ m_t \times (p_t^S - X_t^S) + X_t^S \left(s_t^S - \sum_{k \in \mathbb{S}} X^k s_t^k\right) \end{pmatrix} dt + \sigma(X_t)dB_t \quad (1)$$

where σ is such that $\sigma.\sigma^* = a$ with $a_{i,j} = -2x^i x^j$ if $i \neq j$ and $a_{i,i} = 2x^i(1-x^i)$ and where $s_t = \lim s_t^J J$ and $m_t = \lim m_t^J J$ are the Markovian jump processes with generators Q^s and Q^m and for initial conditions $s_0 = \lim J s_0^J$ and $m_0 = \lim J m_0^J$.

Proof. The proof is in Section 5.1 and relies on the usual Martingale Problem Method. □

Let us give some remarks

(1) We may also consider the proportions p_i random but their law would be J dependent (through the change of time) and it is in disharmony with the biological model which assumes the pool independent of the community size.
(2) Other types of processes for s would have led to similar results, for example we could take a diffusion, with obvious modifications.
(3) It is possible to give an upper bound of the error made by performing the diffusion approximation, by a direct approach not relying on the martingale problem method. It will be the purpose of [21].

2.2.2. *Simpson index.*
Our main object to quantify the biodiversity will be the Simpson index :

$$\mathscr{S}_t = \sum_{i=1}^{S+1} (x_t^i)^2 \qquad (2)$$

with $x_t^{S+1} = 1 - \sum_{i=1}^{S} x_t^i$.

Notice that this expression is the limits of the discrete Simpson index when J goes to the infinity. Its dynamic is given by a non autonomous stochastic differential equation.

Proposition 4. *Denote as usual* $p_t^{S+1} = 1 - \sum_{i=1}^{S} p_t^i,$ *and* $X_t^{S+1} = 1 -$
$\sum_{i=1}^{S} X_t^i.$
So the Simpson index is solution of the following equation :

$$d\mathscr{S}_t = 2(1 - \mathscr{S}_t) - 2 \sum_{i=1}^{S} s_t^i X_t^i (\mathscr{S}_t - X_t^i) + 2m_t \left(\sum_{i=1}^{S+1} p^i X_t^i - \mathscr{S}_t \right) dt + dM_t$$

where M_t is a martingale.
The drift is composed of three terms, the first is the drift in the neutral case without immigration (autonomous equation), the second is a term due to the presence of selection only and the third to the presence of immigration.

This proposition follows from Itô's calculus and details may be found in Section 5. As we will consider mainly the evolution of the expectation of the Simpson index, we do not describe the martingale term.
Now that the large limit SDE is established, we may consider the approximation of the Simpson index. We will first consider a simplified case, but which contains all the main difficulties: the two species case.

3. Approximation of the Simpson index in the quenched or deterministic case: the two species case.

In this part, we study a population of only two species. The equations obtained are in dimension one and the quantities are easier to calculate. It is a basic example to understand the dynamic in greater dimension.

In all this section we will suppose that the *selection and immigration processes are deterministic*, which also amounts to consider the quenched case, i.e. we fix the random environment (immigration and selection), as our goal will be to give a numerical method to approximate $\mathbb{E}[\mathscr{S}_t]$ at a lower cost. We will see in the next section how to consider the annealed case for a very particular case of selection and immigration. In a second part, we will look for constant selection and immigration the behaviour of the process in long time.

The one dimensional Simpson index is

$$\mathscr{S}_t = X_t^2 + (1 - X_t)^2$$

Following the result of the previous section we are thus interested in the case where X_t and $\mathscr{S}_t$ dynamics are given by

$$dX_t = m_t(p_t - X_t)dt + s_t X_t(1 - X_t)dt + \sqrt{2X_t(1 - X_t)}dB_t \qquad (3)$$

$$d\mathscr{S}_t = 4X_t(1 - X_t) \times \left(1 + s_t(X_t - \frac{1}{2})\right) + 2m_t(p_t - X_t)(2X_t - 1)dt$$
$$+ 2(2X_t - 1)\sqrt{2X_t(1 - X_t)}dB_t \qquad (4)$$

where m_t and s_t are the rescaled limit immigration and selection processes.

Comment 1. There is a first interesting feature when analysing the instantaneous behaviour of the dynamic of $\mathscr{S}_t$ in the case where there is no immigration (and thus the Simpson index will tend to 1): when $|s_t| < 2$, the drift of $\mathscr{S}_t$ is always positive so that a even variability if the selection is small does not change the trend to non diversity. At the opposite, if the selection is sufficiently strong it may change locally the behaviour of the Simpson index, and we may thus imagine that change of fitness may lead to oscillation of the Simpson index. We will illustrate this phenomenon numerically.

We will now concentrate on a method to approximate all the moments of X_t, and thus an approximation of $\mathbb{E}[\mathscr{S}_t]$.

3.1. **Approximation of the moments of X_t.**

3.1.1. *The approximation theorem.*
As remarked earlier the momentum of $\mathscr{S}_t$ can not be calculated directly, as the equation of $\mathbb{E}[\mathscr{S}_t]$ is not autonomous(3). However we only need to evaluate the first two moments of X_t. We will see that it will be more difficult than planned. Indeed, taking expectation in (4) (recalling that m_t and s_t are considered deterministic), we get:

$$d\mathbb{E}[X_t] = m_t p_t - m_t \mathbb{E}[X_t] + s_t \big(\mathbb{E}[X_t] - \mathbb{E}[X_t^2]\big)dt \tag{5}$$

By analyzing (5), we cannot express the first momentum of X_t without the second moment and when considering the second moment, the third will appear and so on. It is thus impossible to express the momentum of X_t as the solution of an autonomous equation (except for the trivial case where $s_t = 0$) Nevertheless, the following theorem gives a way of obtaining an approximation of the first moments of X_t by solving a differential system whose size is all the greater that one wishes to be precise.

Theorem 5. *Denote $\tilde{A}_t^n$ the tridiagonal matrix whose coefficients are given by $\tilde{a}_{i+1,i} = (i + p_t m_t)(i + 1)$, $\tilde{a}_{i,i} = i(s_t - i + 1 - m_t)$, $\tilde{a}_{i,i+1} = -is_t$ for i in $\{1, ..., n-1\}$ and $\tilde{a}_{n,n} = -n(n - 1 + m_t)$*
Let consider the following system of ordinary differential equations

$$d\tilde{M}_t = \tilde{A}_t^n \times \tilde{M}_t dt + \mathscr{C}_t dt$$

where $\mathscr{C}_t = (m_t p_t, 0, ..., 0)^t$.

So for j fixed, the jth coordinate of the solution $\tilde{M}_t$ converges when n (the size of the differential system) tends to infinity towards j^{th} momentum of X_t. The error committed is at most

$$\frac{\sqrt{n}\|s\|_\infty^{n-1}}{(n-1)!}.$$

As seen from the upper bound, the convergence is quite fast and even enable to approximate the Laplace transform of X_t quite efficiently. It is mainly due to the fact that we have only two species here. We will see in the next section what happens for three species and will explain how it deteriorates with the number of species.

3.1.2. *Proof of the theorem.* We will begin by considering the error between the solution of our approaching system and the real solution. For practical reasons, each coordinate of the error is multiplied by a coefficient independent of the system size. Let us begin by giving the (non autonomous) system of ordinary differential equations verified by the moments of X_t

Proposition 6. *Let M_t being the vector having for coordinate i the $\mathbb{E}[X_t^i]$ (up to n^{th} moment). M_t is solution of*

$$dM_t = \tilde{A}_t^n \times M_t dt + \mathscr{C}_t dt + B_t dt$$

where $B_t = (0, ..., 0, ns_t\mathbb{E}[X_t^n(1 - X_t)])^T$

Proof.
It is of course a simple consequence of Itô's Formula X_t^i:

$$dX_t^i = iX_t^{i-1}\left((m_t(p_t - X_t) + s_t X_t(1 - X_t)) \times (i - 1)(1 - X_t)\right)dt + d\mathscr{M}_t$$
$$= -is_t X_t^{i+1} + i\left(s_t - (i - 1) - m_t\right)X_t^i + i(i - 1 + m_t p_t)X_t^{i-1}dt + d\mathscr{M}t_t$$

where $\mathscr{M}$ is a martingale and by taking the expectation

$$d\mathbb{E}[X_t^i] = -is_t\mathbb{E}[X_t^{i+1}] + i(s_t - (i-1) - m_t)\mathbb{E}[X_t^i] + i(i-1+m_tp_t)\mathbb{E}[X_t^{i-1}]dt$$

to recover the coefficients of the matrix $\tilde{A}_t^n$.

$\square$

We define the vector of errors Δ_t^n by

$$\Delta_t^n(i) = (s_t)^{i-1} \times \frac{M_t^i - \tilde{M}_t^{i}}{(i-1)!},$$

and we introduce also $\tilde{\mu}_t^n = (0, ..., 0, \mu_t^n)^T$ with

$$\mu_t^n = \frac{(s_t)^{n-1}}{(n-1)!} \times ns_tE[X_t^n(1-X_t)].$$

Note that when we multiply, for i fixed, the difference $M_t^i - \tilde{M}_t^i$ by a coefficient independent of n, the speed of convergence of the ith coordinate of the error change but not the fact that this quantity tends towards 0. On the other hand, it forces the dependent coordinates of n to tend to 0. So we just need to prove that $\|\Delta_t^n\|$ goes to 0, which is not the case for $M_t^i - \tilde{M}_t^i$.

Proposition 7. Δ_t^n *is the solution of the following differential system:*

$$\begin{cases} d\Delta_t^n = A_t^n \times \Delta_t^n dt + \tilde{\mu}_t^n dt \\ \Delta_0^n = 0 \end{cases} \tag{6}$$

and

$$A_t^n = \begin{pmatrix} a_t^1 & b_t^1 & & & & (0) \\ c_t^1 & a_t^2 & b_t^2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & & c_t^{n-2} & a_t^{n-1} & b_t^{n-1} \\ (0) & & & & c_t^{n-1} & a_t^n \end{pmatrix}$$

with

$$a_t^i = i\big(s_t - (i-1) - m_t\big), \quad i < n$$
$$a_t^n = -n\big((n-1) + m_t\big)$$
$$b_t^i = -i^2 \quad i < n$$
$$c_t^i = s_ti(1 - \frac{m_tp_t)}{i-1}) \quad i < n$$

Proof.

Indeed, by substraction,

$$d(M_t^n - \tilde{M}_t^n) = \tilde{A}_t^n \times (M_t^n - \tilde{M}_t^n) + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ s_t n E[(X_t)^n(1 - X_t)] \end{pmatrix}.$$

This involves for all $k < n$

$$d(M_t^k - \tilde{M}_t^k) = k(k - 1 + p_t m_t)(M_t^{k-1} - \tilde{M}_t^{k-1}) \\ + k(s_t - (k - 1) - m_t)(M_t^k - \tilde{M}_t^k) - ks_t(M_t^{k+1} - \tilde{M}_t^{k+1}).$$

Next, we multiply by $\frac{(s_t)^{k-1}}{(k-1)!}$,

$$\begin{aligned} d(\Delta_t^{k,n}) &= \frac{(M_t^{k-1} - \tilde{M}_t^{k-1})s_t^{k-2}}{(k-2)!} \frac{s_t k(k - 1 + p_t m_t)}{k - 1} \\ &\quad + \frac{ks_t^{k-1}}{(k-1)!}\left(s_t - (k - 1) - m_t)(M_t^k - \tilde{M}_t^k)\right) \\ &\quad - \frac{s_t^k(M_t^{k+1} - \tilde{M}_t^{k+1})}{(k)!}k^2 \\ &= ks_t\left(1 - \frac{m_t p_t)}{k - 1}\right)\Delta_t^{k-1,n} + k(s_t - (k - 1) - m_t)\Delta_t^{k,n} - k^2\Delta_t^{k+1,n}. \end{aligned}$$

Now if k=n,

$$\begin{aligned} d\Delta_t^{n,n} &= n\left[s_t\left(1 - \frac{m_t p_t}{n - 1}\right)\Delta_t^{n-1,n} - (m_t + n - 1)\Delta_t^{n,n} \right. \\ &\quad \left. + \frac{s_t^{i-1}}{(n-1)!}s_t \mathbb{E}[X_t^n(1 - X_t)]\right] \end{aligned}$$

We thus find the coefficients of the previous equation system.   □

We may now provide the solution to the system of equation for the error.

Proposition 8. *The solution of* (6) *can be written*

$$\Delta_t^n = \int_0^t \exp\left(\int_q^t A_u^n\right)\tilde{\mu}_q^n dq$$

Proof. First, the solution of (6) can be written as the solution of the homogeneous system and a particular solution.

$$\Delta_t^n = \mathcal{K}\exp\left(\int_q^t A_u^n du\right) + \exp\left(\int_0^t A_u^n du\right)\int_0^t \exp\left(-\int_0^q A_u^n du\right)\tilde{\mu}_t^n(q)dq$$

As $\Delta_0^n = 0$ necessarily $\mathcal{K} = 0$ and $\Delta_t^n = \int_0^t \exp\left(\int_q^t A_u^n du\right)\tilde{\mu}_t^n(q)dq$.   □

We will now show that, for a fixed time interval, $\|\Delta_t^n\|_2$ is uniformly bounded by a quantity which goes to 0 when n goes to infinity. In the following, $\|.\|_F$ stands for the Frobenius norm. Thanks to the formula of the previous proposition:

$$
\begin{aligned}
\|\Delta_t^n(t)\|_2^2 &= \left\| \int_0^t \exp\left(\int_q^t A_u^n du \right) \tilde{\mu}_q^n dq \right\|_2^2 \\
&= \sum_{k=1}^n \left(\int_0^t \left(\exp\left(\int_q^t A_u^n du \right) \tilde{\mu}_q^n \right) dq \right)_k^2 \\
&\leqslant t \times \sum_{k=1}^n \int_0^t \left(\exp\left(\int_q^t A_u^n du \right) \tilde{\mu}_q^n \right)_k^2 dq \\
&\leqslant t \int_0^t \left\| \exp\left(\int_q^t A_u^n du \right) \tilde{\mu}_q^n \right\|_2^2 dq \\
&\leqslant t \int_0^t \left\| \exp\left(\int_q^t A_u^n du \right) \right\|_F^2 \|\tilde{\mu}_q^n\|_2^2 dq \\
&\leqslant t^2 \times \sup_{x \in [0,t]} \left\| \exp\left(\int_x^t A_u^n du \right) \right\|_F^2 \times \sup_{x \in [0,t]} |\mu_x^n|^2
\end{aligned}
$$

The following easy lemma gives us an upper bound for the two previous norms:

Lemma 9.
If $X \in [0,1]$ then $sX^n(1-X) = O(\frac{1}{n})$ and there is a constant c_1 independent of n such that $\sup_{x \in [0,t]} |\mu_x^n| \leqslant \sup_x \in [0,t] s_x^{n-1} \frac{c_1}{(n-1)!}$

Proof. The maximum of $X^n(1-X)$ on $[0,1]$ is achieved in $\frac{n}{n+1}$ and it's worth $(1 - \frac{1}{n+1})^n \times \frac{1}{n+1}$.
This quantity is of the order of $\frac{1}{n}$ when n goes to infinity.
The upper bound of μ_x^n follow. $\square$

Next,

$$
\begin{aligned}
\left\| \exp\left(\int_x^t A_u^n du \right) \right\|_F^2 &= Tr\left(\exp\left(\int_x^t A_u^n du \right)^T \times \exp\left(\int_x^t A_u^n du \right) \right) \\
&= Tr\left(\exp\left(\int_x^t A_u^n + (A_u^n)^T du \right) \right) \\
&= \sum_{\lambda_i} \exp\left(\lambda_i^x \right) \\
&\leqslant n \times \max_{i \in \{1...n\}} \left(\exp(\lambda_i^x) \right)
\end{aligned}
$$

where the λ_i^x are the eigenvalues of $\int_x^t A_u^n + (A_u^n)^T du$.

Then if $\max\limits_{i \in \{1...n\}} \exp(\lambda_i^x)$ is independent of n, there is a constant c_2 independent of n verifying

$$\sup_{x \in [0,t]} \left\| \exp\left(\int_x^t A_u^n du \right) \right\|_F \leqslant \sqrt{n} \times c_2$$

And then there exist a constant C such as

$$\|\Delta_t^n(t)\|_2 \leqslant \sqrt{n} c_2 \times \frac{c_1 \|s\|_\infty^{n-1}}{(n-1)!} \leqslant \frac{C\sqrt{n}\|s\|_\infty^{n-1}}{(n-1)!}$$

It remains to show that the eigenvalues of $\int_x^t A_u^n + (A_u^n)^T du$ (which are real) have an upper bound independent of n.

Proposition 10. *Let* $A_t^n + (A_t^n)^T =$
$$\begin{pmatrix} a_t^1 & b_t^1 & & & & \\ b_t^1 & a_t^2 & b_t^2 & & & \\ & \ddots & \ddots & \ddots & & \\ & & b_t^{n-2} & a_t^{n-1} & b_t^{n-1} \\ & & & b_t^{n-1} & a_t^n \end{pmatrix}$$ *with*

$$a_t^i = 2i\big(s_t - (i-1) - m_t\big), \quad i < n$$
$$a_t^n = -n\big((n-1) + m_t\big)$$
$$b_t^i = s_t i (1 - \frac{m_t p_t)}{i-1}) - (i-1)^2$$

then the eigenvalues of $\int_x^t (A_u^n)^T + A_u^n du$ *are uniformly bounded in n .*

Proof. To prove this result we need to use Gershgorin's disk. The eigenvalues of $\int_x^t (A_u^n)^T + A_u^n du$ are included in the union of the disk D_i whose centers are the ith term on the diagonal $(\int_x^t a_u^i du)$ and for radius the sums of the coefficients norms on the ith line except the diagonal term $(|\int_x^t b_u^{i-1} du| + |\int_x^t b_u^i du|)$. It is important to consider the forms of the discs in our case. As we have to show that the eigenvalues have a upper bound independently of n. We just need to look at the shape of the discs for i and n big enough. From the matrix if $i < n$, their centers are

$$2i \left(\int_x^t s_u du - (i-1)(t-x) - \int_x^t m_u du \right)$$

and their radius are equal to

$$\left|\int_x^t b_u^{i-1}du\right| + \left|\int_x^t b_u^i\right| = \left|\int_x^t s_u i\left(1 - \frac{m_u p_u}{i-1}\right) - (i-1)^2 du\right|$$

$$+ \left|\int_x^t s_u(i+1)\left(1 - \frac{m_u p_u}{i}\right) - i^2 du\right|$$

$$= \left|i\int_x^t s_u du - (i^2 - 2i)(t-x) + i\int_x^t m_u du\right|$$

$$+ \left|i\int_x^t s_u du - i^2(t-x) - i\int_x^t m_u du\right| + O(1)$$

$$= 2i\left((i-1)(t-x) - \int_x^t s_u du + \int_x^t m_u du\right) + O(1).$$

So the maximum value for an eigenvalue of $\int_x^t (A_u^n)^T + A_u^n du$ belonging to D_i is

$$\int_x^t a_u^i du + \left|\int_x^t b_u^{i-1}du\right| + \left|\int_x^t b_u^i du\right|$$

$$= 2i\left(\int_x^t s_u du - (i-1)(t-x) - \int_x^t m_u du\right)$$

$$+ 2i\left((i-1)(t-x) - \int_x^t s_u du + \int_x^t m_u du\right) + O(1)$$

$$= O(1)$$

as soon as i is big enough. If $i = n$ the same reasoning is still working. So the eigenvalues of $(A^n)^T + A^n$ have an upper bound independent of n. $\qquad\square$

This property concludes the proof and we get the following upper bound

$$\|\Delta_t^n(t)\|_\infty \leqslant \|\Delta_t^n(t)\|_2 \leqslant \frac{C\sqrt{n}\|s\|_\infty^{n-1}}{(n-1)!}$$

The constant C depends on time (exponentially) and therefore this algorithm will be less accurate if we look at the behavior of the process in long time. This result gives a satisfactory approximation of the $\mathscr{S}_t$'s moment. Convergence is very fast, and the algorithm boils down to solving a linear differential system. To ensure the interest of this method we can compare the expectation of Simpson's index obtained by a Monte Carlo method from the discrete model to that obtained with this approximation. The Figure 1 presents such an approximation.

3.2. Numerical applications.
The simulations presented in this part are obtained from the previous theorem. The values of s, m, p are those of the large population approximation and not that of the discrete model. The size of the approaching system will be usually between 80 and 144 depending on the needed precision.
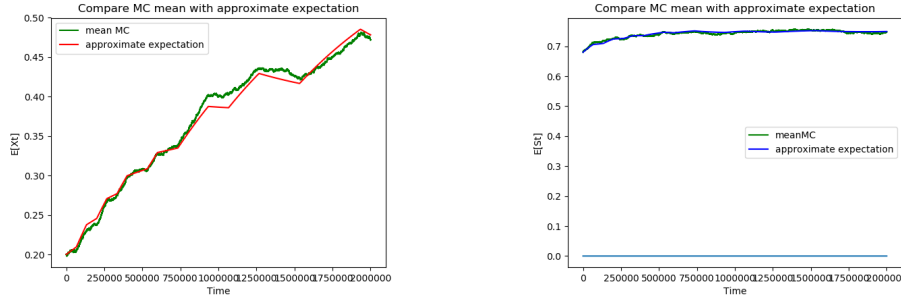
FIGURE 1. Are plotted the approximate values of $E[\mathscr{S}_t]$ and $E[X_t]$ by the precedent method from the approximation in large population and by MC method from the discreet model. The number of simulated trajectories for MC mean is 500 (red and blue), $J = 1000$, $m = 2$, $p = 0.5$, $X_0 = 0.2$, s switches between 2 and -2 at regular time intervals, the size of the approaching linear system is 100.

3.2.1. *Influence of s on Simpson Index.*

In this part, $p = 0.5$. Now we know how to approximate the expectation of $\mathscr{S}_t$, so we can check the influence of s on this quantity. Let us make precise a statement enounced when deriving $\mathscr{S}_t$.

Proposition 11.
If $\|s\|_\infty$ is smaller than 2 and if $m = 0$, $E[\mathscr{S}_t]$ is increasing.

Proof.
If we refer to the equation of $dE[\mathscr{S}_t]$ (cf (4)), we see that if $\|s\|_\infty < 2$, the quantity $4X_t(1 - X_t) \times \left(1 + s_t(X_t - \frac{1}{2})\right)$ is positive whatever the initial condition, so Simpson's index mean is always growing . $\qquad\square$

In other words, selection alone can not bring about a renewal of biodiversity. On the other hand if for some t, $s_t > 2$, $E[\mathscr{S}_t]$ can decrease. In this case the more the selection is important, more the decay is pronounced. We will see in the last part that this phenomenon can be generalized to a larger number of species. Figure 2 present these different behaviours with respect to s and Figure 3 the combination of initial parameters and selection under which the Simpson's index is decreasing.

3.2.2. *Approximation of T_1, T_0, $T_{1,0}$.*

In the special case where $m_t = 0$, a species inevitably invaded the community in a finite time. We define by T_1, (respectively T_0), the smallest time from which the process X_t reaches 1 (respectively 0) and $T_{1,0} = min(T_1, T_0)$.
Thanks to the approximation of moments we can obtain an approximation of the T_1 distribution function . In fact we know X_t^n for n big enough and so since $\lim_{n\to\infty} E[(X_t)^n] = 1 \times \mathbb{P}(T_1 < t)$ we obtain $P(T_1 < t)$. The same
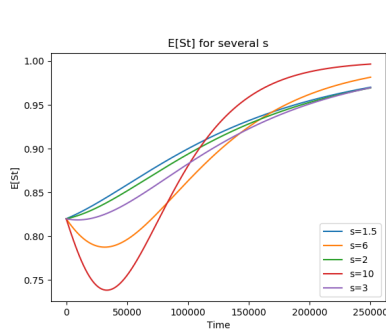
FIGURE 2. Several trajectories of $E[\mathscr{S}_t]$ are drawn for different s. $X_0 = 0.1$, m=0.
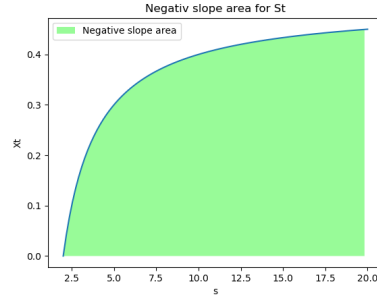

FIGURE 3. The colored area represents the pairs (X, s) for which the slope of the Simpson's index will be negative.

way with $1 - X_t$ gives $\mathbb{P}(T_0 < t)$. Figure 4 gives an approximation of the distribution function of T_1 when there is no immigration.
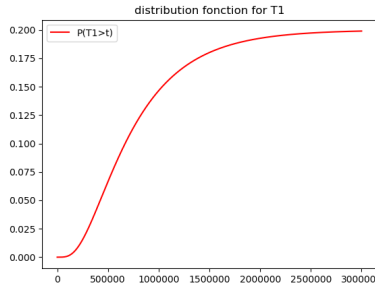

FIGURE 4. Distribution function of T_1 for m=0,p=0.5,$X_0 = 0.2$, $s = 2$. The size of the approaching linear system is 100

We can use the same method to obtain $T_{\mathscr{S}_t}$ the probability that $\mathscr{S}_t$ is equal to 1 at time t.

3.3. Long time behavior.

Two cases are distinguished in this part, the case $m = 0$ and the case $m \neq 0$. In this first case there is no immigration and necessarily a species invade the community. Invasion times and the probability that the species with a selective advantage will invade the community are calculated. In the second case the system admits an invariant measure, we explain it and we specify the speeds of convergence towards this measure. More details about the behavior of stochastic processes in long time can be found in [33] and [22].

3.3.1. *The case without immigration ($m = 0$): absorption.*
The results in this section are partially well known and we include them only
to get a full picture of the behavior of the Simpson index.
Recall that $\mathscr{S}_t$ satisfies the equation

$$d\mathscr{S}_t = 4X_t(1 - X_t) \times \left(1 + s_t(X_t - \frac{1}{2})\right)dt + dM_t. \tag{7}$$

If there is no more immigration the states 0 and 1 are absorbing, and it is
then well known that the process reaches them in finite time almost surely.
For more details about this refer to [33]. Let T_1 et T_0 the hit times of 1 and
0 for the random variable X_t et $T_1 \wedge T_0 = T_{1,0}$.

Proposition 12. *Suppose $s_t = s \in \mathbb{R}$,*

  (1) $T_{1,0} < \infty$ *almost surely, for all initial condition X_0.*

  (2) *Let g be the solution of*

  $$x(1 - x)g''(x) + sx(1 - x)g'(x) = -1 \quad et \quad g(0) = g(1) = 0 \tag{8}$$

  then $E_{X_0}[T_{1,0}] = g(X_0)$.

  (3) $P(T_1 < T_0) = \frac{e^{-sX_0} - 1}{e^{-s} - 1}$.

The proof is given in section 5.2
Let us consider some particular case which illustrates that the same behavior
may be obtained with varying selection. Suppose $T > 0$ and s_t is a constant
function on intervals $[kT, (k+1)T]$, $k \in \mathbb{N}$ which can take the values s_0 or $-s_0$
for $s_0 > 0$, randomly. We will establish a result similar to the constant case.
Let us begin by the following lemma which asserts that without selection
one may reach the boundary at any time.

Lemma 13. *Consider the following process*

$$dX_t = \sqrt{2X_t(1 - X_t)}dB_t \tag{9}$$

Note $T_1 = \inf\{t, X_t = 1\}$, an initial condition $x \in]0, 1]$ and a time t .
*Then $\mathbb{P}_x(T_1 < t) > 0$, in other words, 1 is accessible for X_t from any non-zero
initial condition and in a as little time as one wants.*

Proof. Assume that $t_0 = \inf\{t, \mathbb{P}_x(T_1 < t) > 0\} > 0$.
t_0 is well defined. Remark that $\mathbb{E}[X_t]$ is constant in time because X_t is a
bounded martingale. Then $0 < x = \lim_{t \to \infty} E[X_t] = 1 \times \mathbb{P}_x(T_1 < T_0) + 0$ as
$\mathbb{E}_x[T_{1,0}] < \infty$ and so $\mathbb{P}_x(T_1 < \infty) > 0$. Then, let $\Delta_t > 0$ be such that
$t_0 - 2\Delta_t > 0$. We will show then that there is y such that $\mathbb{P}_y(T_1 < \Delta_t) > 0$.

If that was not the case then $\forall y \in [0,1], \mathbb{P}_y(T_1 < \Delta_t) = 0$ and

$$
\begin{aligned}
\mathbb{P}_x(T_1 < t_0) &= \mathbb{E}_x[\mathbb{1}_{T_1 < t_0}] = \mathbb{E}_x[\mathbb{1}_{T_1 < t_0} \times (\mathbb{1}_{T_1 < \Delta_t} + \mathbb{1}_{T_1 > \Delta_t})] \\
&= \mathbb{E}_x[\mathbb{1}_{T_1 < \Delta_t}] + \mathbb{E}_x[\mathbb{1}_{T_1 < t_0} \times \mathbb{1}_{T_1 > \Delta_t}] \\
&= \mathbb{E}_x[\mathbb{1}_{T_1 < t_0} \times \mathbb{1}_{T_1 > \Delta_t}] = \mathbb{E}_x[\mathbb{1}_{T_1 > \Delta_t} E[\mathbb{1}_{T_1 < t_0} | X_{\Delta_t}]] \\
&= \mathbb{E}_x[\mathbb{1}_{T_1 > \Delta_t} \times \mathbb{P}_{X_{\Delta_t}}(T_1 < t_0 - \Delta_t)] \\
&= \mathbb{E}_x[\mathbb{1}_{T_1 > \Delta_t} \mathbb{E}_{X_{\Delta_t}}[\mathbb{1}_{T_1 > \Delta_t} \cdots \mathbb{E}_{X_{n\Delta_t}}[\mathbb{1}_{T_1 > \Delta_t} \\
&\qquad\qquad \times \mathbb{P}_{(n+1)\Delta_t}(T_1 < t_0 - (n+1)\Delta_t)]]].
\end{aligned}
$$

Let us choose n such that $t_0 - (n+1) < \Delta_t$. Then $\mathbb{1}_{T_1 > \Delta_t} \times \mathbb{P}_{(n+1)\Delta_t}(T_1 < t_0 - (n+1)\Delta_t) \leqslant \mathbb{1}_{T_1 > \Delta_t} \mathbb{P}_{(n+1)\Delta_t}(T_1 < \Delta_t) = 0$ and $\mathbb{P}_x(T_1 < t_0) = 0$ which is contrary to the assumptions. Now, we show that $\mathbb{P}_x(T_1 < t_0 - \Delta_t) > 0$:

$$
\begin{aligned}
\mathbb{P}_x(T_1 < t_0 - \Delta_t) &= \mathbb{E}_x[\mathbb{1}_{T_1 < t_0 - \Delta_t} \times (\mathbb{1}_{T_y < t_0 - 2\Delta_t} + \mathbb{1}_{T_y > t_0 - 2\Delta_t})] \\
&\geqslant \mathbb{E}_x[\mathbb{1}_{T_1 < t_0 - \Delta_t} \times \mathbb{1}_{T_y < t_0 - 2\Delta_t}] \\
&\geqslant \mathbb{E}_x[\mathbb{1}_{T_y < t_0 - 2\Delta_t} E[\mathbb{1}_{T_1 < t_0 - \Delta_t} | T_y]] \\
&\geqslant \mathbb{E}_x[\mathbb{1}_{T_y < t_0 - 2\Delta_t} \mathbb{P}_y(T_1 < t_0 - \Delta_t - T_y)] \\
&\geqslant \mathbb{E}_x[\mathbb{1}_{T_y < t_0 - 2\Delta_t} \mathbb{P}_y(T_1 < \Delta_t)] \\
&\geqslant \mathbb{P}_x(T_y < t_0 - 2\Delta_t) \times \mathbb{P}_y(T_1 < \Delta_t).
\end{aligned}
$$

But we know that $\mathbb{P}_y(T_1 < \Delta_t) > 0$ by the previous calculation and the local uniform ellipticity of diffusion also ensures us that $\mathbb{P}_x(T_y < t_0 - 2\Delta_t) > 0$. So we obtain $\mathbb{P}_x(T_1 < t_0 - \Delta_t) > 0$ which is contrary to the fact that $t_0 > 0$. So $\forall t > 0, \mathbb{P}_x(T_1 < t) > 0$ which concludes the proof of the lemma. $\qquad\square$

Of course, this result may be adapted for the process $dX_t = sX_t(1 - X_t)dt + \sqrt{2X_t(1 - X_t)}dB_t$. Indeed if $s > 0$ the drift goes in the right direction. Else, if $s < 0$, we obtain a symmetric result by replacing in the previous reasoning T_1 by T_0.

Proposition 14.
Let $s_0 > 0$, $T > 0$ and s_t a constant function on the intervals $[kT, (k+1)T]$, $k \in \mathbb{N}$ which can take the values s_0 or $-s_0$ values randomly.
Let consider the process

$$dX_t = s_t X_t(1 - X_t)dt + \sqrt{2X_t(1 - X_t)}dB_t \qquad (10)$$

Finally $T_{1,0} = \inf\{t, X_t = 0 \quad ou \quad X_t = 1\}$ then $\mathbb{P}_x(T_{1,0} < \infty) = 1$.

Proof. The idea is to show that for each time interval of size T, the probability of reaching 1 or 0 is non-zero and independent of the position where the process is located. So we compare the probability that our process reaches 1 or 0 to a geometric law. So let us first show that $\forall t \in]0, T], \exists \alpha > 0$, such that $\forall x \in [0,1], \mathbb{P}_x(T_{1,0} < t) > \alpha$. Suppose $s > 0$ on $[0, T]$ and denote $g_i(x) = \mathbb{P}_x(T_i < t)$, $i \in \{0, 1\}$. Both functions are continuous, g_0 is decreasing and $g_0(0) = 1$, $g_0(1) = 0$, whereas g_1 is increasing and $g_1(1) = 1$,

$g_1(0) = 0$. Then there exist a $x_0 \in]0, 1[$ such as $g_0(x_0) = g_1(x_0) = \alpha_+$. And by the previous lemma since s does not vary on $]0, T]$ we have that $\alpha_+ > 0$. A symmetric reasoning for $s < 0$ guarantees us the existence of a $\alpha_- > 0$. Let $\alpha = 2 \min(\alpha_-, \alpha_+)$, α is then strictly positive and $\forall x \in [0, 1]$, $\mathbb{P}_x(T_{1,0} < t) = \mathbb{P}_x(T_1 < t) + \mathbb{P}_x(T_0 < t) \geqslant \alpha$. Then, for $t \in [nT, (n+1)T]$, using previous inequality:

$$\begin{aligned}
\mathbb{P}_x(T_{1,0} < t) \geqslant & \mathbb{P}_x(T_{1,0} < nT) \\
\geqslant & \mathbb{E}_x[\mathbb{1}_{T_{1,0}<(n-1)T}] + \mathbb{E}_x[\mathbb{1}_{T_{1,0}>(n-1)T}\mathbb{P}_{X_{(n-1)T}}(T_{1,0} < T)] \\
\geqslant & \mathbb{P}_x(T_{1,0} < (n-1)T) + \alpha \times \mathbb{E}_x[1 - \mathbb{1}_{T_{0,1}<(n-1)T}] \\
\geqslant & (1 - \alpha) \times \mathbb{P}_{X_{(n-1)T}}(T_{1,0} < (n-1)T) + \alpha \\
\geqslant & \alpha + \alpha(1 - \alpha) + ... + \alpha(1 - \alpha)^{n-1} \geqslant 1 - (1 - \alpha)^n
\end{aligned}$$

We obtain when n goes to infinity $\mathbb{P}_x(T_{1,0} < \infty) = 1$. $\qquad\square$

Even with frequent changes of fitness, a species always ends up invading the community if m is zero. We may then consider the process with immigration. Note that while preparing this paper, comparable (and even more general) results were obtained (in the multi-allelic case) by Coron et al [6].

3.3.2. *The case with immigration ($m \neq 0$): invariant measure.*
Assume now s, m and p are constants. The long time behavior for varying selection and immigration is far more complicated and may lead to interesting behavior that will be considered in another paper. We thus consider the following process:

$$dX_t = m(p - X_t) + sX_t(1 - X_t)dt + \sqrt{2X_t(1 - X_t)}dB_t$$

and will consider the long time behavior in a quantitative way, i.e. not using Meyn-Tweedie's theory, but rather via a Poincaré inequality.
Our process X_t is Markovian and evolves in a range bounded by 0 and 1. At first we will ask ourselves what is the behaviour of our process in the neighbourhood of 0 and 1, by considering the criterion given by Feller cf[17, 18]. According to the values of m, p and s, our process will have different behaviours in the neighbourhood of 1 and 0. We have already seen that if $m = 0$ then 0 and 1 are absorbing states reached by the process in a finite time almost surely.(The same hold if m is non-zero and if p is 0 or 1.) Now if m and p are not trivial, 0 and 1 are no longer absorbing. In other words, immigration prevents the invasion of the community by a species. Moreover, for some values of m and p these two states are not accessible, i.e the process can not reach them in a finite time.

Proposition 15. *The state 1 (respectively 0) is accessible by the process X_t if and only if $m(1 - p) < 1$ (respectively $mp < 1$) and regular otherwise.*

The proof will be given in section 5.3. In the case of inaccessible or reflective boundaries (which is our the case), the law of the process X_t admit a density

and converge in long time to an invariant measure. This measure has a density, denoted π. In addition π is a solution of the Fokker-Planck equation:

$$0 = \frac{\partial}{\partial t}\pi = -\frac{\partial \pi}{\partial y}\big(m(p-y) + sy(1-y)\big) + \frac{\partial^2 \pi}{\partial^2 y}y(1-y)$$

The solution of this equation is:

$$\pi(y) = c \times y^{mp-1} \times (1-y)^{m(1-p)-1} \times \exp(sy) \qquad (11)$$

The constant c is chosen so that $\int_0^1 \pi(y)dy = 1$.

The following Figures 5 and 6 show the influence of the parameters on the expectation and the variance of Simpson's equilibrium index.
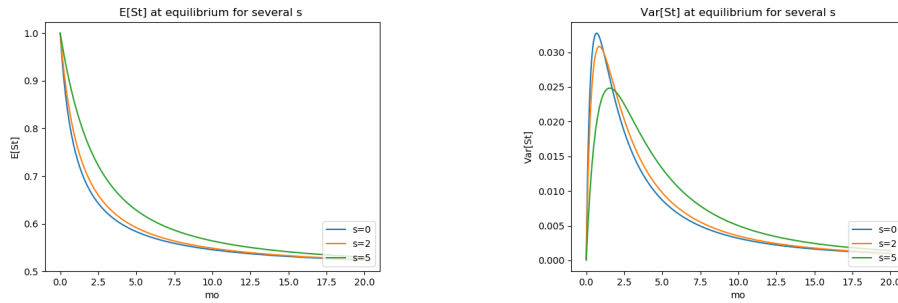

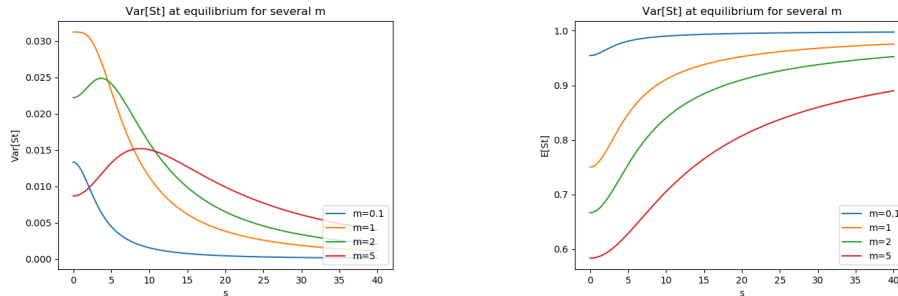FIGURE 5. Expectation and variance of the Simpson index at equilibrium against m for several values of s, p=0.5

FIGURE 6. Here are plotted the variance and the expectation of the simpson index at equilibrium against s for several values of m. p=0.5, size of linear approaching system is 100.
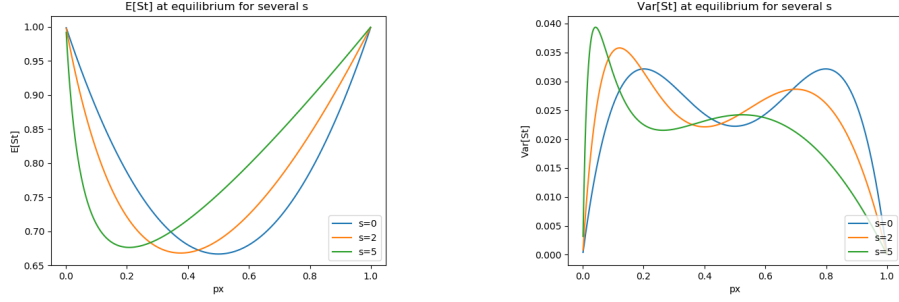
FIGURE 7. Here are plotted the variance and the expectation of the Simpson index at equilibrium against p for several values of s. m=2, size of linear approaching system is 100.

Let us now quantify the convergence to equilibrium. Recall at first that the process X_t has for generator $\mathscr{L}$ and for invariant measure π. Denote $P_t f(x) = \mathbb{E}[f(X_t)|f(0) = x]$ the associated semigroup. In fact, when $s = 0$, the full spectrum is known, see for example Shimakura [35] which provides a spectral gap value m. It will imply an exponential convergence to equilibrium in L_π^2.

Proposition 16. *Let us suppose that $s > 0$. The following Poincaré inequality is valid, i.e. for every smooth function f*

$$Var_\pi(f) \leqslant \min\left(\frac{e^s}{m}, \frac{8e^{(1-M)s}}{m}\right) \int_0^1 f'^2(x) \times x(1-x)d\pi(x)$$

where M is a median of π. As a consequence π_t converge to π in L_π^2 exponentially:

$$Var_\pi(P_t f) \leqslant e^{-2\max\left(\frac{e^s}{m}, \frac{8e^{(1-M)s}}{m}\right)} Var_\pi(f).$$

Proof. Using usual Holley-Stroock's perturbation argument we easily deduce, that the Poincaré constant is at most e^s/m. We will now use Hardy's type condition (see for example [1]) for Poincaré condition, that we recall now

Lemma 17. *Let π and ν be two measures and M the median of π. Let*

$$B_M^+ = \sup_{x > M} \int_x^1 d\pi \int_M^x \frac{1}{\nu(t)}dt, \qquad B_M^- = \sup_{x < M} \int_0^x d\pi \int_x^M \frac{1}{\nu(t)}dt.$$

If B_M^- et B_M^+ are bounded, then the following Poincaré inequality holds

$$Var_\pi(f) \leq c_P \int f'^2 d\nu.$$

In addition, the optimal constant c_P verifies

$$\frac{1}{2}\max(B_M^+, B_M^-) \leqslant c_P \leqslant 4\max(B_M^+, B_M^-)$$

.

We apply the lemma to π and $\nu = x(1-x)\pi$ using both sides of the estimates. Denote π_0 and ν_0 the case where $s = 0$ and the Poincaré constant is m. Then

$$
\begin{aligned}
B_M^{+,s} &= \sup_{x>M} \int_x^1 d\pi \int_M^x \frac{1}{\nu(t)} dt \\
&= \sup_{x>M} \int_x^1 e^{st} d\pi_0(t) \int_M^x \frac{e^{-st}}{\nu_0(t)} dt \\
&\leq e^{(1-M)s} B_M^{+,0}
\end{aligned}
$$

The same reasoning shows that $B_M^{-,s} \leq B_M^{-,0}$.   □

Of course, one can do easily the same for $s < 0$ using a symmetric reasoning. If the order is good with respect to the immigration parameter, as the case $s = 0$ is optimal, it is an open question to look at the dependence with respect to the selection parameter. We may also consider a convergence in entropy, via the logarithmic Sobolev inequality without selection established by Stannat [37] or Miclo [31] and the same line of proof using Holley-Stroock perturbation argument or the Hardy type condition for logarithmic Sobolev inequality (see again [1]). Note that the convergence in entropy entails a convergence in total variation via Cszisar-Pinsker-Kullback inequality, but the constant involved are less explicit so we omit the details.

This quantitative long time behaviour enables us to give an error while approximating the asymptotic Simpson index (being a smooth function of the species). As usual, an L^2 decay will enable us to consider long time behaviour for initial measures whose density with respect to the invariant measure is bounded, which in could prevent starting from a Dirac measure. However due to regularization, and so waiting a time t_0, enables (loosing on the constants in the decay) to start from a Dirac measure. See for example [2].

4. Generalization to a larger number of species or in random environment

In this section we provide extensions of the two species case to 1) finite number of species, 2) two species case in a particular random environment, namely Wright-Fisher diffusion environment.

4.1. Expectation approximation for three species.

In fact we will give the main ideas for $S = 2$. Extension to a larger number of species is only technically involved and requires no further arguments. Denote X_t and Y_t the proportions of the two main species, s_t^x and s_t^y their selection parameters and p_t^x and p_t^y their proportions in the pool. The immigration parameter will still be denoted m_t. The method presented for $S = 1$ in the previous section can be generalized to a larger number of species. It will have of course some limitations: greater the number of species is, larger

will be the size of the approaching linear system. In fact, the derivative of the expectation of order n involves only the expectation of lower and higher order in the case of two species. Now with 3 species we also need to know the expectation of the form $\mathbb{E}[X_t^n Y_t^k]$ for k, n in $\mathbb{N}$. We will thus need a system of size N^2. We present here the extension of our approximation for 3 species.

$$\begin{pmatrix} dX_t \\ dY_t \end{pmatrix} = \begin{pmatrix} m_t(p_t^x - X_t) + X_t(s_t^x - X_t s_t^x - Y_t s_t^y) \\ m_t(p_t^y - Y_t) + Y_t(s_t^y - X_t s_t^x - Y_t s_t^y) \end{pmatrix} dt + \sigma(X_t, Y_t) dB_t \quad (12)$$

where σ verifies $\sigma.\sigma^*(x,y) = a(x,y)$ with

$$a(x,y) = 2 \begin{pmatrix} x(1-x) & -xy \\ -xy & y(1-y) \end{pmatrix}$$

We have to calculate $d(X_t^n Y_t^k)$ with Itô's formula :

$$\begin{aligned}
d(X_t^n Y_t^k) =& \big(m_t(p_t^x - X_t) + X_t(s_t x - X_t s_t^x - Y_t s_t^y)\big) n X_t^{n-1} Y_t^k \\
&+ \big(m_t(p_t^y - Y_t) + Y_t(s_t^y - X_t s_t^x - Y_t s_t^y)\big) k X_t^n Y_t^{k-1} \\
&+ n(n-1)(1-X_t)X_t^{n-1}Y_t^k + k(k-1)(1-Y_t)X_t^n Y_t^{k-1} \\
&- 2nk X_t^n Y_t^k + d\mathscr{M}_t \\
=& X_t^{n-1}Y_t^k n(m_t p_t^x + n - 1) \\
&+ X_t^n Y_t^{k-1} k(m_t p_t^y + k - 1) \\
&+ X_t^n Y_t^k \big(-m_t(n+k) - 2kn - k(k-1) - n(n-1) + n s_t^x + k s_t^y\big) \\
&- X_t^{n+1}Y_t^k s_t^x(n+k) \\
&- X_t^n Y_t^{k+1} s_t^y(n+k) \\
&+ d\mathscr{M}_t \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (13)
\end{aligned}$$

here $\mathscr{M}_t$ is a martingale. Then $d\mathbb{E}[X_t^n Y_t^k]$ is expressed in terms of 4 other quantities which complicates the one dimensional calculations. Moreover we must define what are the neglected expectations on which we will make an approximation, that is how to close the system. We can decide we make an approximation to the order N then that we neglect all the terms of higher order in the expression of $dE[X_t^n Y_t^k]$ where $max(n,k) = N$.
Suppose we want to get the expectation up to order N we need exactly $\sum_{k=1}^{N} 2k+1$ quantities. And so the size of the approaching differential linear system will be of order $(N+1)^2$.
The following figure represents the complexity of the problem, for example $d\mathbb{E}[X_t Y_t]$ is expressed as a function of the expectations of the quantities to

which the blue arrows point.

$$X\rightarrow \qquad \leftarrow\!XY\rightarrow \qquad Y$$

$$X^2 \qquad X^2Y \qquad X^2Y^2 \qquad Y^2X \qquad Y^2$$
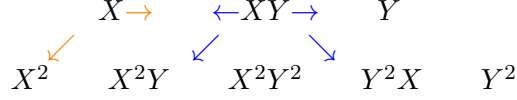
Algorithmically it is not very difficult to build the matrix approaching the expectations of the diffusion. We must begin by giving a vector composed of the different expectation of size $(N+1)^2$. For that, let us define an application that transforms the expectation of order (n,k) that is to say $\mathbb{E}[X_t^n Y_t^k]$ into an integer which corresponds to its coordinate in the expectation vector. Next we build the matrix A_{N^2} as in the case of two species from the coefficients calculated in 13. We consider as error each expectation (n,k) with $max(n,k) > N$ in the Itô formula. So that the error is composed of $2N+1$ terms. And the approximation boils down to solving numerically a linear system. As an example consider the case $N = 1$, we therefore involve four expectations which are $\mathbb{E}[X^n Y^k], k, n \in \{0,1\}$. The order imposed by ϕ is therefore $(1, \mathbb{E}[X], \mathbb{E}[Y], \mathbb{E}[XY])$. Three terms compose the error: $\mathbb{E}[s_t^x X^2], \mathbb{E}[s_t^y Y^2], 2\mathbb{E}[(s_t^x + s_t^y)XY^2]$. We can also, as in the case of two species, prove the convergence of this algorithm by following exactly the same pattern as in the one species case. The renormalizing coefficients of the expectation (n,k) then become $\frac{(s_t^x)^n (s_t^y)^k}{(n-1)!(k-1)!}$. It can similarly be shown that the error is at most of the order of $\frac{N^2 \max(s_t^x, s_t^y)^N}{(N-1)!}$.

The extension to a larger number is straightforward and will entail an error of the order $\frac{N^{\frac{S}{2}+2}\|s_t\|_\infty^N}{N!}$, and it will still be reasonable but requires computations of a system of size N^S which may be prohibitive for large S.

Numerical applications. We can easily program such an algorithm and check that the results obtained are in agreement with quantities obtained by Monte Carlo method. See following figures:

Basic example

We consider here a case with no immigration and constant selection parameter. The number of simulated trajectories for MC mean is 1000, $J = 1000$, $m = 0$, $X_0 = 0.5$, $Y_0 = 0.3$, $s_y = 2$, $s_x = 1$, the size of the approaching linear system is 144. Figure 8 plots approximate values of $\mathbb{E}[\mathscr{S}_t]$ and $\mathbb{E}[X_t]$ by the precedent method from the approximation in large population and by MC method from the discrete model.

Time dependent parameter case

In this second example, we consider once again a case without immigration and time dependent selection parameter. The number of simulated trajectories for MC mean is 1000, $J = 1000$, $m = 0$, $X_0 = 0.5$, $Y_0 = 0.3$, $s_y = 2$, s_x is piecewise constant taking two values 4 and -4 at regular time intervals, the size of the approaching linear system is 144. Figure 9 plot the approximate
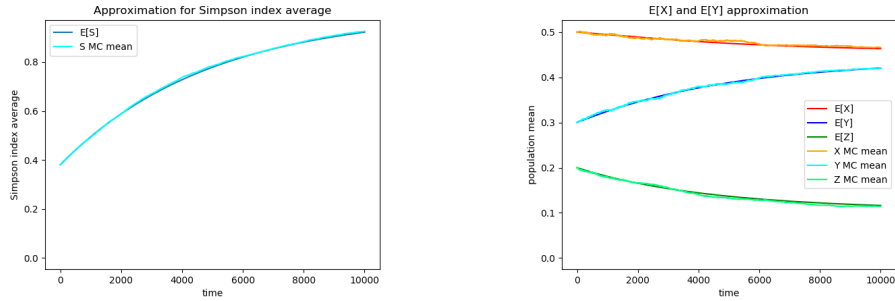
FIGURE 8. On the left hand side expectation of the Simpson index and Monte Carlo mean, and on the right hand side expectations of each species and their Monte Carlo means
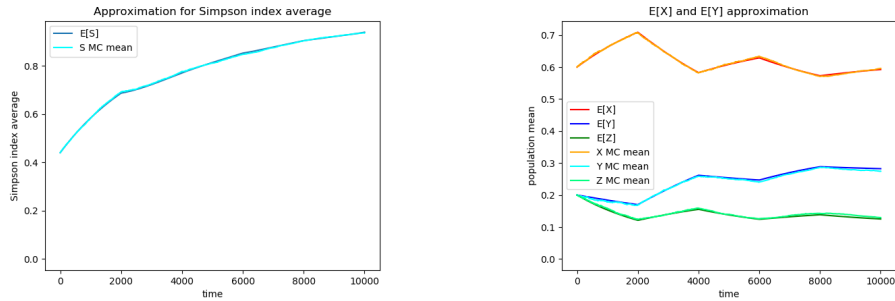

FIGURE 9. Left hand side Simpson index, right hand with the expectations of the three species.

values of $E[\mathscr{S}_t]$ and $E[X_t]$ by the precedent method from the approximation in large population and by MC method from the discrete model.

4.2. **When the selection is a diffusion.** In the third Section we gave a method to get the moments of X_t, and thus $\mathbb{E}(\mathscr{S}_t)$ for a time dependent immigration/selection parameter. If these parameters are random but autonomous, it gives a way to approximate the expectation of the Simpson index by doing a Monte Carlo mean with respect to the environment, passing from quenched to annealed. It would be however more interesting to evaluate directly the expectation of the Simpson index without further Monte Carlo simulations. It seems quite impossible to give a general algorithm for every environment but we will give in this section an efficient approximation method in a particular case. We consider for the selection parameter s_t a rescaled Wright-Fisher diffusion, whose leading Brownian motion is independent of the one leading the SDE for the species evolution. This choice assures us that s_t is a diffusion evolving in a bounded set and the choice of the different parameters leads to a wide choice of a Moran process with immigration.

4.2.1. *The expectation approximation.*

Let us just give first the diffusion approximation result for this particular case, whose proof is even simpler as it relies on usual approximation diffusion for Markov chains.

Theorem 18. *Assume that $(v_n^J)_{n \in \mathbb{N}}$ is a Moran process without selection with size J and the parameters m^s et p^s. Let c and b two constants such as $s_n^J = cv_n^J - b \quad \forall n \in \mathbb{N}$, and assume that X_n follow a Moran process with size J and parameters $(m_n)_{n \in \mathbb{N}}$, $(p_n)_{n \in \mathbb{N}}$ et $(s_n^J)_{n \in \mathbb{N}}$ describe in the first part. Let U_n^J be the process having for coordinates X_n^J et s_n^J.*
Then when J goes to infinity, the process $U_{tJ^2}^J$ converge in law to the process U_t which coordinates are solutions of the following stochastic differential equation:

$$\begin{pmatrix} dX_t \\ dv_t \end{pmatrix} = \begin{pmatrix} m_t'(p_t - X_t) + s_t' X_t (1 - X_t) \\ m_t'^s (p_t^s - v_t) \end{pmatrix} dt + \begin{pmatrix} \sqrt{2X_t(1 - X_t)} \\ \sqrt{2v_t(1 - v_t)} \end{pmatrix} dB_t \quad (14)$$

where $s_t' = \frac{s_t}{J} = cv_t - b$, $m_t' = \frac{m_t}{J}$, $m_t'^s = \frac{m_t^s}{J}$.

To approach the expectation of X_t we use the method describe previously for three species, here v play the same role as a third species. However the dynamics is not exactly the same, the Itô formula gives us:

$$\begin{aligned} d(X_t^n v_t^k) = & X_t^{n-1} v_t^k n(m_t' p_t + n - 1) \\ & + X_t^n v_t^{k-1} k(m_t'^s p_t^s + k - 1) \\ & + X_t^n v_t^k \left(-(m_t' + b)n - km_t'^s - k(k-1) - n(n-1) \right) \\ & + X_t^{n+1} v_t^k nb \\ & + X_t^n v_t^{k+1} cn \\ & - X_t^{n+1} v_t^{k+1} nc \qquad\qquad\qquad\qquad\qquad (15) \\ & + d\mathcal{M}_t \qquad\qquad\qquad\qquad\qquad\qquad\quad (16) \end{aligned}$$

with $\mathcal{M}_t$ a martingale. Then as previously we close our system, for a given N, and to do so to neglect all the terms of higher order in the expression of $d\mathbb{E}[X_t^n v_t^k]$ where $max(n,k) = N$. And now the algorithm is able to calculate all the expectations of the form $\mathbb{E}[X_t^n v_t^k]$ and so obtain the expectation of the Simpson index. The proof follows the same pattern. The renormalizing coefficients of the expectation (n,k) in the proof allow to control the eigenvalues of the matrix thanks to the Gershgorin disks as before. Many choices are possible and we take here the coefficient $\frac{1}{\sqrt[4]{n!k!}}$. This choice leads to a convergence speed at most of the order of $\frac{N^2}{\sqrt[4]{(N)!}}$.

4.2.2. *Comparison with the neutral model.*

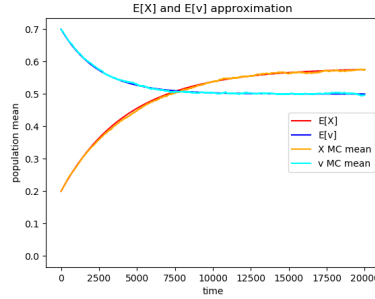In this part we compare the case where s is "neutral on average", to the

FIGURE 10. Are plotted the approximate values of $\mathbb{E}[v_t]$ and $\mathbb{E}[X_t]$ by the precedent method from the approximation in large population and by MC method from the discrete model. The number of simulated trajectories for MC mean is 5000, $J = 1000$, $X_0 = 0.2$, $v_0 = 0.7$, $m^s = 4$, $m = 2$, $p = p^s = 0.5$, $c = 3$, $b = 0.5$ the size of the approaching linear system is 144

neutral case with $s = 0$. Thanks to the previous method one can for example calculate the average Simpson index in the case where the selection expectation is 0. For it let's take $p^s = 1/2$, $v_0 = 1/2$ (this enforces $c = -2b$). The following figures show the results:
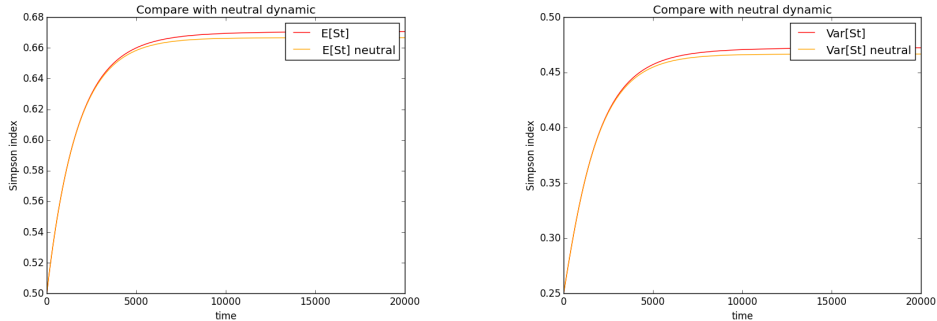

FIGURE 11. Comparison with neutral case. Approximate values of $\mathbb{E}[\mathscr{S}_t]$ and $\mathbb{E}[X_t]$ by the precedent method from the approximation in large population $X_0 = 0.5$, $m^s = 1$, $m = 2$, $p = 0.5$, $c = 3$, size of the approaching linear system is 144

We thus see that a selection even if neutral in mean, involves deeper mechanism which lead to a different behaviour than the neutral one. Of course the Simpson index involves not only the expectation of one species but also the moment of order two.

4.3. **Effect of selection on increase of biodiversity.**
We have already seen in the case of two species that selection alone could contribute to the decrease of the average Simpson index in the absence of
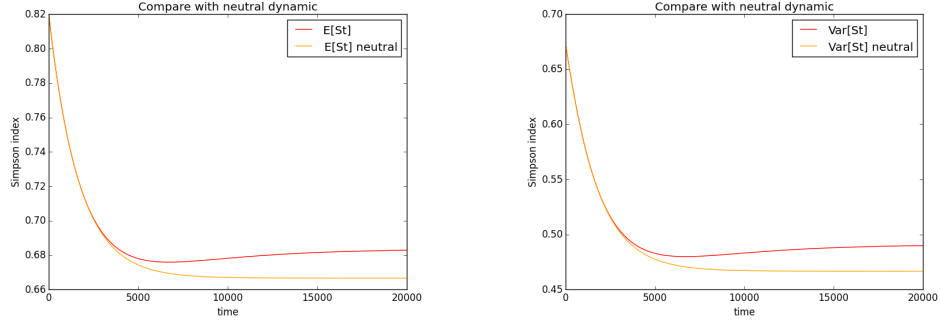
FIGURE 12. Comparison with the neutral case. Approximate values of $\mathbb{E}[\mathscr{S}_t]$ and $\mathbb{E}[X_t]$ by the precedent method from the approximation in large population $X_0 = 0.1$, $m^s = 0.5$, $m = 2$, $p = 0.5$, $c = 5$, size of the approaching linear system is 144.

immigration. There was however a threshold for s under which such a phenomenon could not occur. We sort of generalize it here to any number of species.

Proposition 19. *Note as previously $S + 1$ the number of species in the community with s_i the selection parameter for species i. Then if all $\|s_i\|_\infty$ are less than $\frac{1}{2}$ the Simpson index is increasing in the absence of immigration. In other words, the selection can not be the source of the diversity decreasing.*

Proof. Assume all the $s_t^i, \forall t$ are between α and $-\alpha$. First write:

$$
\begin{aligned}
1 - \mathscr{S}_t &= 2 \sum_{i=1}^{S} X_t^i (1 - X_t^i) - \sum_{j \neq i} X_t^i X_t^j \\
&= \sum_{i=1}^{S} X_t^i (1 - X_t^i) + \sum_{i=1}^{S} X_t^i \left(1 - \sum_{i=1}^{S} X_t^i\right) \\
&= \sum_{i=1}^{S+1} X_t^i (1 - X_t^i)
\end{aligned}
$$

Then,

$$
d\mathbb{E}[\mathscr{S}_t] = 2\mathbb{E}[1 - \mathscr{S}_t] - 2\mathbb{E}[\sum_{i=1}^{S} s_t^i X_t^i (\mathscr{S}_t - X_t^i)]dt
$$

$$
= 2\mathbb{E}[1 - \mathscr{S}_t - \sum_{i=1}^{S} s_t^i X_t^i (\mathscr{S}_t - 1) - \sum_{i=1}^{S} s_t^i X_t^i (1 - X_t^i)]dt
$$

$$
\geqslant 2\mathbb{E}[1 - \mathscr{S}_t - \sum_{i=1}^{S} s_t^i X_t^i (\mathscr{S}_t - 1) - \alpha \sum_{i=1}^{S} X_t^i (1 - X_t^i)]dt
$$

$$
\geqslant 2\mathbb{E}[(1 - \mathscr{S}_t)(1 + \sum_{i=1}^{S} s_t^i X_t^i) - \alpha \sum_{i=1}^{S} X_t^i (1 - X_t^i)]dt
$$

$$
\geqslant 2\mathbb{E}[(1 - \mathscr{S}_t)(1 - \alpha + \sum_{i=1}^{S} s_t^i X_t^i) + \alpha \sum_{i=1}^{S} X_t^i (1 - \sum_{i=1}^{S} X_t^i)]dt
$$

$$
\geqslant 2\mathbb{E}[(1 - \mathscr{S}_t)(1 - \alpha + \sum_{i=1}^{S} s_t^i X_t^i)]dt
$$

$$
\geqslant 2\mathbb{E}[(1 - \mathscr{S}_t)(1 - \alpha(1 + \sum_{i=1}^{S} X_t^i))]dt
$$

and so if $\forall i, \alpha \leqslant \frac{1}{2}$, $d\mathbb{E}[\mathscr{S}_t] \geqslant 0$ and $\mathbb{E}[\mathscr{S}_t]$ is increasing.

$\square$

Remark that this bound is certainly not optimal, as the two species case indicates but true for each S.

4.4. **Long time behaviour.**

We will once again assume in this part s, m, p are constants. If $m = 0$ then a species will still invade the community definitively. On the other hand, if $m \neq 0$, the law of the vector of abundance, converges in a long time to a unique invariant measure. Consider the generator of the diffusion (14) which is the generator of the Wright-Fisher diffusion with selection and mutation:

$$
\mathscr{L}f(x) = \sum_{i,j=1}^{S} x^i(\delta_{i,j} - x^j)\frac{\partial^2 f(x)}{\partial x^i \partial x^j} + \sum_{i=1}^{S}\left(m(p^i - x^i) + x^i\left(s^i - \sum_{i=1}^{S} x^i s^i\right)\right)\frac{\partial f(x)}{\partial x^i}
$$

A reversible and stationary measure for the diffusion (14) is given by (see for example [13, 40, 24]:

$$
\pi_S(dx) = C \times \exp\left(\sum_{i,j=1}^{S+1} s^i x^i x^j\right) \times (x^1)^{mp^1-1} \times ... \times (x^{S+1})^{mp^{S+1}-1}dx_1...dx_S
$$

Where $x^{S+1} = 1 - \sum_{i=1}^{S} x^i$ and $p^{S+1} = 1 - \sum_{i=1}^{S} p^i$, $s^{S+1} = 0$. C is a constant just like $\int \pi_S(dx) = 1$.

Of course, when s and m are time dependent, periodic for example, an invariant measure will not exist. The next figure presents the approximate values of $\mathbb{E}[\mathscr{S}_t]$ and $\mathbb{E}[X_t^i]$ for i in $\{1, 2, 3\}$ by the precedent method from the approximation in large population and by Monte Carlo method from thr discret model. The number of simulated trajectories for Monte Carlo mean is 5000, $J = 500$, m is a time dependant piecewise process, it takes alternatively the values of 3 and 0 at regular time intervals. $X_0 = 0.5$, $p_x = 0.33$, $Y_0 = 0.3$, $p_y = 0.33$, s_y et s_x are Markovian jump processes, the size of the approaching linear system is 144.
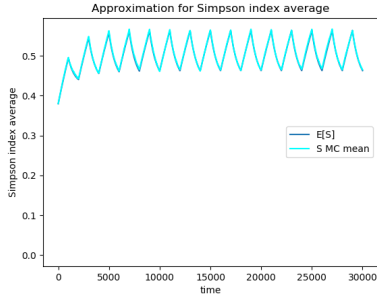
FIGURE 13. $\mathbb{E}[\mathscr{S}_t]$
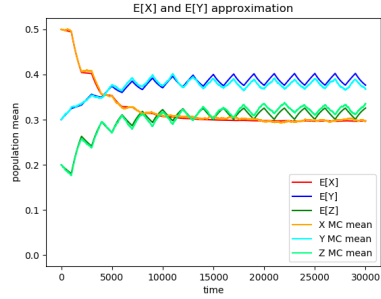
FIGURE 14. $\mathbb{E}[X_t], \mathbb{E}[Y_t]$

Concerning the long time behaviour, we may once again refer to [35] for the spectral gap which is $e^{(S+1)\sum_{1}^{S+1} s^i}/m$ by Holley-Stroock's perturbation argument. Unfortunately, it is not possible to refine this argument as there is no Hardy's type inequalities in this case. Once again it is also possible to derive a logarithmic Sobolev inequality, and thus convergence in entropy (and total variation) but constants are less explicit.

5. PROOFS

In this section we gather the proofs, technical or more or less well known.

5.1. **Proof of the diffusion approximation, Theorem 3.**

In the following proof we'll get back to a martingale problems. All the results used in this section can be funded in [41] p267-272.

For the sake of clarity, assume that $m = 0$, and that $S = 1$.

The multidimensional case is treated exactly the same way.

We can put $h = \frac{1}{J^2}$ and U means here (x, s) where $x \in E_x$ and $s \in E_s$.

Let $a(U) = x(1-x)$, $b(U) = sx(1-x)$ and $L_x f(U) = b(U) \times \frac{\partial f}{\partial x}(U) + a(U) \times$

$\frac{\partial^2 f}{\partial x^2}(U)$ the generator of the SDFE (14) and $L_s f(U) = \sum\limits_{y \in E_s} Q_{s,y} f(x, y)$ the generator of a Markovian jump process applied to a function depending of the population variable.

Let's start with the following lemma

Lemma 20. *([41] p268)*
Let f be a C^∞ function, note $A_J f(U) = \int_E f(y) - f(u) d\pi_J(U, dy)$ then $J^2 A_J f$ converge uniformly to $L_x f + L_s f$

Proof.

$$A_J f(U) = \int_E f(y) - f(u) d\pi_J(U, dy) = \int_E f(z, w) - f(x, s) d\pi_J(U = (x, s), dy)$$

$$= \int_E f(z, w) - f(x, s) d\pi_J(U, dy)$$

$$= \int_E f(z, w) - f(x, w) + f(x, w) - f(x, s) d\pi_J(U, dy)$$

$$= \int_E f(z, w) - f(x, w) d\pi_J(U, dy) + \int_E f(x, w) - f(x, s) d\pi_J(U, dy)$$

$$= \int_E f(z, w) - f(x, w) d\pi_J(U, dy) + \sum_{w \in E_s} f(x, w) - f(x, s) P_J^i(x, y, h).$$

Via Taylor's formula, we obtain

$$\int_E f(z, w) - f(x, w) d\pi_J(U, dy) = \mathbb{E}[f(X_{t+h}, s_{t+h}) - f(x, s_{t+h}) | U_t = (x, s)]$$

$$= \frac{\partial f}{\partial x}(x, s) \mathbb{E}[X_{t+h} - x | U_t] + \frac{\partial^2 f}{2\partial x^2}(x, s) \mathbb{E}[(X_{t+h} - x)^2 | U_t]$$

$$+ \frac{\partial^2 f}{\partial x \partial s}(x, s) \mathbb{E}[X_{t+h} - x | U_t] E[s_{t+h} - s | U_t] + O(\mathbb{E}[\|U_{t+h} - U_t\|^3 | U_t]).$$

we give the limits of the previous quantities,

- $\lim\limits_{h \to 0} \sup\limits_E \|\frac{1}{h} \mathbb{E}[X_{t+h}^J - x | U_t^J]\| = b(x, s)$ by the second propriety (2),
- $\lim\limits_{h \to 0} \sup\limits_E \|\frac{1}{h} \mathbb{E}[(X_{t+h}^J - x)^2 | U_t^J]\| = a(x, s)$ by the second propriety (2),
- $\lim\limits_{h \to 0} \sup\limits_E \|\frac{1}{h} \mathbb{E}[X_{t+h}^J - x | U_t^J] \mathbb{E}[s_{t+h}^J - s | U_t^J]\| = 0$ because

$$|\mathbb{E}[s_{t+h}^J - s | U_t]| = |\sum_{w \in E_s} (w - s_t) P_J^i(s_t, w, h)|$$

$$\leqslant \max_{x, y \in E_s} |x - y| \sum_{w \in E_s / s_t} P_J^i(s_t, w, h),$$

- $\lim\limits_{h \to 0} \sup\limits_E \mathbb{E}[\|U_{t+h}^J - U\|^3 | U_t^J] = 0$ because $\int_E \sup\limits_{i \leqslant S} |y_i - u_i|^3 \pi_J(U, dy) \to 0$

  and by the previous point.

Then, going to the limit in the previous expression,

$$\limsup_{h \to 0} {}_E |A_J f(U)\frac{1}{h} - L_x f(U) - L_s f(U)|$$

$$\leq \limsup_{h \to 0} {}_E |\frac{1}{h} \int_E f(z, w) - f(x, w) d\pi_J(U, dy) - L_x f(U)|$$

$$+ \limsup_{h \to 0} {}_E |\frac{1}{h} \sum_{w \in E_s} (f(x, w) - f(x, s)) P_J^i(x, y, h) - L_s f(U)|$$

$$= 0$$

And this expression conclude the proof.

$\square$

Now let f be $\mathscr{C}^\infty$, then

$$\mathbb{E}_U[f(U_t^J)] = f(U) + \mathbb{E}_U \left[\sum_{k=1}^{|tJ^2|-1} \mathbb{E}[f(U_{(k+1)h}^J) - f(U_{kh}^J)|U_{kh}^J] \right]$$

$$= f(U) + \mathbb{E}_U \left[\sum_{k=1}^{|tJ^2|-1} A_J f(U_{kh}) \right]$$

$$= f(U) + \mathbb{E}_U \left[\sum_{k=1}^{|tJ^2|-1} \frac{h}{h} A_J f(U_{kh}) \right]$$

and so

$$\mathbb{E}_U[f(U_{t+h}^J) - f(U) - \sum_{k=1}^{|tJ^2|-1} \frac{h}{h} A_J f(U_{kh})] = 0$$

i.e $f(U_t^J) - f(U) - \sum_{k=1}^{|tJ^2|-1} \frac{h}{h} A_J f(U_{kh})$ is a martingale for π^J.
Moreover, note the sum is a Riemann sum and the previous lemma ensures when J tends to infinity the convergence of

$$f(U_t^J) - f(U) - \sum_{k=1}^{|tJ^2|-1} \frac{h}{h} A_J f(U_{kh})$$

towards

$$f(U_t) - f(U) - \int_0^t L_x f(U_s) + L_s f(U_s) ds.$$

We need now to find a probability measure on the Borel sets of the canonical space $\mathscr{C}([0,1], R)$ verifying the martingale problem for $L_x + L_s$. Let us show

now that π^J admits an adherent value in the space of probability measure on the Borel of $\mathscr{C}([0,1], R)$ with the norm

$$\|\pi^J\| = \sup_{f \in C} \frac{|\int f d\pi^J|}{|f|_\infty}$$

(which is a norm since the π^J are supported in $[0,1]$).

Let note $\pi^J f = \int_E f d\pi^J$.

Let $(f_n)_\mathbb{N}$ be a dense sequence in the space of continuous functions then $(\pi^J f_n)_J$ is a sequence of $\mathbb{R}$ having an adherence value in $\mathbb{R}$ because it is uniformly bounded by $|f_n|_\infty$. Then by diagonal extraction, eventually for a subsequence, $\pi^J f_n$ converges to a certain ϕ_{fn} in $\mathbb{R}$ for all n. And by the uniformly continuous extension theorem, we define ϕ_f for all f of $\mathscr{C}$. And since ϕ_f is a linear form, the Riesz-Markov theorem ensures the existence of a unique measure μ such that $\int f d\mu = \phi_f$. Since this is true for all $f \in \mathscr{C}$, by considering the constant function equal to 1, we find $\mu(\Omega) = 1$ and μ is a probability. The convergence of $\|\pi^J - \mu\|$ to 0 is then immediate in view of the chosen norm.

Thus our sequence π^J admits an accumulation point. So, there is at least one π and one X process that satisfy the martingale problem associated with $L_x + L_s$.

And so π verifies $\int f(U_t) - f(U) - \int_0^t L_x f(U_s) + L_s f(U_s) ds d\pi(U, U_t) = 0$.

So it exist at least a solution to the martingale problem for $L_x + L_s$.

If the uniqueness of this martingale problem is verified then the process converges in law to the diffusion process (our X_t) defined by a, b and thus and s the jump process of generator Q, since they are both solutions of the same problem of martingale. The proof of uniqueness is quite standard, following Ethier [12] when s, m, p constant. A straightforward modification allows to obtain the result for s, m, p random.

5.2. **Proof of Proposition 12.**

Let g be the solution of the differential equation (8). Let us first verify that g is well defined on $[0,1]$. It must be ensured that the solutions do not diverge in 0 and 1, in which case the second member of the equation is not defined. For that we can write the solution of this equation. So

$$g(x) = \int_0^x e^{-su} \left(K + \int_{\frac{1}{2}}^u \frac{e^{st}}{t(1-t)} dt \right) du + C$$

where C, K are constant. As e^{-su} is bounded on $[0,1]$, there are two positive constants B and D such that

$$\lim_{x \to 0} |g(x)| \leqslant \lim_{x \to 0} \int_a^x \int_a^u \frac{D}{t} dt du + B \leqslant \infty$$

as ln is integrable on a neighbourhood of 0. Thus g is well defined on [0,1] and bounded (because continue).

So we have $g(X_{t \wedge T_{1,0}}) = g(X_0) + \int_0^{t \wedge T_{1,0}} g'(X_u)\sqrt{2X_u(1 - X_u)}dB_u - t \wedge T_{1,0}$. But the process $\int_0^{t \wedge T_{1,0}} g'(X_u)\sqrt{2X_u(1 - X_u)}dB_u$ is a stopped martingale because $t \wedge T_{1,0}$ is a stopping time and $g'(X_u)\sqrt{2X_u(1 - X_u)}$ is adapted to the considered filtration. We deduce that $\mathbb{E}_{X_0}[g(X_{t \wedge T_{1,0}})] = g(X_0) - \mathbb{E}_{X_0}[t \wedge T_{1,0}]$ and the first property, i.e $\mathbb{E}_{X_0}[t \wedge T_{1,0}] \leqslant 2\sup_{[0,1]}(g) \leqslant \infty$, and thus the second point is shown.Now if $t \to \infty$, $\mathbb{E}_{X_0}[g(X_{t \wedge T_{1,0}})] \to 0$ because $g(0) = g(1) = 0$ and we find again $g(X_0) = E_{X_0}[T_{1,0}]$.

To prove the third point, consider $f(x) = e^{-sx} - 1$. Then f is solution of $f''(x) + sf'(x) = 0$ and $f(0) = 0$. By Itô's formula, we obtain $d(f(X_t) = f(X_0) + dMt$. As T_1 and T_0 are stopping times $T_1 \wedge T_0 = T_{1,0}$ is also a stopping time. So $f(X_{T_{0,1}}) = f(X_0) + dM'_t$ où dM'_t is still a martingale. By taking expectation we have $\mathbb{E}_{X_0}[f(X_{T_{0,1}})] = f(X_0) = f(1)\mathbb{P}(T_1 < T_0)$ and we deduce $\mathbb{P}(T_1 < T_0) = \frac{e^{-sX_0}-1}{e^{-s}-1}$.

5.3. **Proof of Proposition 15.**

Let us consider the speed measure and the scale function as in Feller [18].

$$m(y) = \frac{1}{2y(1 - y)}exp(\int_a^y s + \frac{m(p - x)}{x(1 - x)}dx) \quad \text{(speed measure)}$$
$$= c \times y^{mp-1} \times (1 - y)^{m(1-p)-1} \times \exp(sy), \quad c \in \mathbb{R}$$

$$\mu(t) = \int_a^t exp(-\int_a^y s + \frac{m(p - x)}{x(1 - x)}dx)dy \quad \text{(scale function)}$$
$$= c' \int_a^t y^{-mp} \times (1 - y)^{-m(1-p)} \times \exp(-sy)dy, \quad c' \in \mathbb{R}.$$

Then 1 is reachable if and only if $\mu(1) < \infty$ and $\int_{\frac{1}{2}}^1 \mu(1) - \mu(y)m(y)dy < \infty$. It is easily seen that $\mu(1) < \infty$ if and only if $m(1 - p) < 1$.

Next, $\int_{\frac{1}{2}}^1 (\mu(1) - \mu(y)m(y))dy < \infty$ if and only if $\mu(1) - \int \mu(y)m(y)$ is integrable on a neighborhood of 1. But

$$\mu(1) - \mu(y)m(y) \leqslant \int_y^1 x^{-mp}(1 - x)^{-m(1-p)}y^{mp-1}(1 - y)^{m(1-p)-1}e^{sy-sx}dx$$
$$\leqslant \int_y^1 (1 - x)^{-m(1-p)}dx \times (1 - y)^{m(1-p)-1}\frac{1}{y}$$
$$\leqslant a \times \frac{1}{y},$$

for some constant a. This quantity is well defined and integrable on a neighborhood of 1. So 1 is reachable if and only if $m(1-p) < 1$.

Now if $m(1-p) \leqslant 1$, 1 is not reachable, it is regular (reflective barriers) if and only if $m(y)$ is integrable. It is indeed the case here, $-21 < m(1-p)-1 \leqslant 0$. Of course the same holds for 0.

REFERENCES

[1] C. Ané, S. Blachère, D. Chafaï, P Fougères, I Gentil, F Malrieu, C. Roberto, and G Scheffer. *Sur les inégalités de Sobolev logarithmiques*, volume 10 of *Panoramas et Synthèses [Panoramas and Syntheses]*. Société Mathématique de France, Paris, 2000. With a preface by Dominique Bakry and Michel Ledoux.

[2] D. Bakry, P. Cattiaux, and A. Guillin. Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.*, 254(3):727–759, 2008.

[3] V. Bansaye, M-E Caballero, and S. Méléard. Scaling limits of general population process-wright-fisher and branching processes in random environment. february 8 2018.

[4] R.A. Chisholm, R. Condit, K.A. Rahman, P.J. Baker, S. Bunyavejchewin, Y-Y. Chen, G. Chuyong, H.S. Dattaraja, S. Davies, C.E.N. Ewango, et al. Temporal variability of forest communities: empirical estimates of population change in 4000 tree species. *Ecology letters*, 17(7):855–865, 2014.

[5] R. Condit, N. Pitman, E.G. Leigh, J. Chave, J. Terborgh, R.B. Foster, P. Núnez, S. Aguilar, R. Valencia, G. Villa, et al. Beta-diversity in tropical forest trees. *Science*, 295(5555):666–669, 2002.

[6] C. Coron, S. Méléard, and D. Villemonais. Impact of demography on extinction/fixation events, 2018. To appear in *Journal of Mathematical Biology*.

[7] M. Danino, D.A. Kessler, and N.M. Shnerb. Stability of two-species communities: drift, environmental stochasticity, storage effect and selection. *Theoretical population biology*, 119:57–71, 2018.

[8] M. Danino and N.M. Shnerb. Fixation and absorption in a fluctuating environment. *Journal of theoretical biology*, 441:84–92, 2018.

[9] M. Danino and N.M. Shnerb. Theory of time-averaged neutral dynamics with environmental stochasticity. *Physical Review E*, 97(4):042406, 2018.

[10] M. Danino, N.M. Shnerb, S. Azaele, W.E. Kunin, and D.A. Kessler. The effect of environmental stochasticity on species richness in neutral communities. *Journal of Theoretical biology*, 409:155–164, 2016.

[11] A. Depperschmidt, A. Greven, , and P. Pfaelhuber. Tree- valued fleming-viot dynamics with mutation and selection. *Ann. Appl.Probab.*, (22):2560–2615, february 2012.

[12] SN. Ethier. A class of degenerate diffusion processes occuring in population genetics. *Comm.Pure Appl. Math*, 29:483–493, 1976.

[13] SN. Ethier and TG. Kurtz. The infinitely many neutral alleles diffusion model. *Adv.Appl.Prob*, (13):429–452, 1981.

[14] R.S. Etienne. A new sampling formula for neutral biodiversity. *Ecology letters*, 8(3):253–260, 2005.

[15] R.S. Etienne and H. Olff. A novel genealogical approach to neutral biodiversity theory. *Ecology Letters*, 7(3):170–175, 2004.

[16] W. Ewens and J. Warren. *Mathematical population genetics. I*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer-Verlag, New York, second edition, 2004. Theoretical introduction.

[17] W. Feller. The parabolic differential equations and the associated semigroups of transformation. *Ann.Math*, 55:468–519, 1952.

[18] W. Feller. Diffusion processes in one dimension. *Trans.Am.Math.soc*, 77:1–30, 1954.

[19] T. Fung, JP. O'Dwyer, and RA. Chisholm. Species-abundance distributions under colored environmental noise. *J. Math. Biol.*, 74(1-2):289–311, 2017.

[20] T. Fung, J.P. O'Dwyer, and R.A. Chisholm. Species-abundance distributions under colored environmental noise. *Journal of mathematical biology*, 74(1-2):289–311, 2017.

[21] G. Gackou, A. Guillin, and A. Personne. Quantitative approximation of a moran model in random environment by wright-fisher process, 2018. in preparation.

[22] N.S. Goel and N. Richter-Dyn. *Stochastic Models in Biology*. Academic Press, 1st January 1974.

[23] M. Grieshammer. Genealogical distance under selection. *??*, 2018.

[24] R.C. Griffiths. On the distribution of allele frequencies in a diffusion model. *Theor. Pop.Biol.*, (15):140–158, 1979.

[25] R.C. Griffiths and S. Tavaré. Sampling theory for neutral alleles in a varying environment. *Philos.Trans.R.Soc.London*, (344):403–410, february 1994.

[26] S. Hubbell. *The Unified Neutral Theory of Biodiversity and Biogeography*. Springler, Princeton University Press, Princeton, NJ, 2001.

[27] F. Jabot and J. Chave. Inferring the parameters of the neutral theory of biodiversity using phylogenetic information and implications for tropical forests. *Ecology letters*, 12(3):239–248, 2009.

[28] F. Jabot and journal=Oikos-volume=125 number=12 pages=1733–1742 year=2016 publisher=Wiley Online Library Lohier, T. Non-random correlation of species dynamics in tropical tree communities.

[29] M. Kalyuzhny, R. Kadmon, and N.M. Shnerb. A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities. *Ecology letters*, 18(6):572–580, 2015.

[30] S.M. Krone and C. Neuhauser. ancestral process with selection. *Theoretical Population Biology*, 51:210–237, 1997.

[31] L. Miclo. About projections of logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXVI*, volume 1801 of *Lecture Notes in Math.*, pages 201–221. Springer, Berlin, 2003.

[32] C. Muirhead and J. Wakeley. Modeling multiallelic selection using a moran model. *Genetics*, 182(4):1141–1157, 2009.

[33] S. Méleard. *Modèle probabilistes en écologie et évolution*. Springer, 2016.

[34] C. Neuhauser and S.M. Krone. The genealogy of samples in models with selection. *Genetics*, 145:519–534, february 1997.

[35] N. Shimakura. équations différentielles provenant de la génétique des populations. *Tôhoku Math. J.*, 29(2):287–318, 1977.

[36] E.H. Simpson. Measurement of diversity. *Nature*, (163):688, april 30 1949.

[37] W. Stannat. On the validity of the log-Sobolev inequality for symmetric Fleming-Viot operators. *Ann. Probab.*, 28(2):667–684, 2000.

[38] M. Steinrucken, Y-X-R. Wang, and Y-S. Yun S.Song. An explicit transition density expansion for amulti allelic wright-fisher diffusion with general diploid selection. *Theor Popul Biol.*, (83):1–14, february 2013.

[39] I. Volkov, J-R. Banavar, S-P. Hubbell, and A. Maritan. Neutral theory and relative species abundance in ecology. *Nature.*, 424(22):1035–1037, August 2003.

[40] S. Wright. Evolution and the genetics of populations. *Univ. of Chicago Press*, Vol. 3(1977), vol. 4 (1978)(13):443–473 and 460–476, 1977-1978.

[41] D. W.Stroock and S. Varadhan. *Multidimensional Diffusion Process*. Springler, 1997.

Arnaud GUILLIN, Laboratoire de Mathématiques Blaise Pascal, CNRS UMR 6620, Université Clermont-Auvergne, avenue des Landais, F-63177 Aubière.
E-mail address: `arnaud.guillin@uca.fr`

Franck JABOT, Laboratoire d'Ingéniérie pour les Systèmes Complexes, IRSTEA, Campus des Cézeaux 9, avenue Blaise Pascal - CS 20085 63178 Aubière
E-mail address: `franck.jabot@irstea.fr`

Arnaud PERSONNE, Laboratoire de Mathématiques Blaise Pascal, CNRS UMR 6620, Université Clermont-Auvergne, avenue des Landais, F-63177 Aubière.
E-mail address: `arnaud.personne@uca.fr`