

UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Fisica e Astronomia “Galileo Galilei”

Corso di Laurea in Fisica

Tesi di Laurea

Stochastic Modelling of Ecosystem dynamics

Relatore

Prof. Samir Suweis

Laureando

Davide Coldebella

Anno Accademico 2018/2019

«...perché a vent'anni è tutto ancora intero, perché a vent'anni è tutto chi lo sa ...»

Guccini, *Eskimo*

# Contents

<b>Introduction</b>	<b>v</b>
<b>1 Mathematical and Computational concepts</b>	<b>1</b>
1.1 Stochastic differential equations . . . . .	1
1.1.1 The jump process and the master equation (ME) . . . . .	2
1.1.2 The diffusion process and the Fokker-Planck equation (FPE) . . . . .	3
1.1.3 Deterministic evolution and Liouville's equation . . . . .	3
1.2 Ergodicity and expected values . . . . .	4
1.2.1 Stationary process . . . . .	4
1.2.2 Homogeneity . . . . .	4
1.2.3 Ergodicity . . . . .	5
1.2.4 Expected values . . . . .	5
1.2.5 Generating functions . . . . .	6
1.3 Algorithms and pseudocode . . . . .	6
1.3.1 Finite difference method . . . . .	6
1.3.2 Gillespie algorithm . . . . .	7
<b>2 Ecological concepts</b>	<b>9</b>
2.1 Neutral Theory . . . . .	9
2.2 The 1D Grand-Canonical Ensemble . . . . .	11
2.3 The Ecological Ensemble . . . . .	13
2.4 Results . . . . .	17
2.5 The 2D Lotka-Volterra Equation . . . . .	21
<b>3 Persistence times Analysis</b>	<b>27</b>
3.1 Empirical Persistence-time distribution . . . . .	27
3.2 Theoretical Persistence-time distribution . . . . .	27
3.3 Simulated Persistence-time distribution . . . . .	29
<b>Bibliography</b>	<b>31</b>



# Introduction

Ecology is a branch of science which was born as far as the 4th century BC with Aristotle, had a rapid growth in the 18th and 19th centuries, but established itself only in the last century [18]. In fact, Darwin's "Origin of Species" - which is among the most famous works in biology, was only published in 1909. On the modelling side, the first mathematical model of inter-species competition - known as Lotka-Volterra's equations - were published in 1925-6 [10, 17]. Moreover, even if scientists began systematically gathering data during the last fifty years [1], there is still a lack of big and long-term macro-ecological dataset due to the fact that field ecological campaigns are time-consuming and expensive process, that may even have to overcome both political obstacles and technical difficulties [3]. For example the most studied and well-known ecological site - the Barro Colorado Island - although studied for more than eighty years, had its species abundances systematically censused only from the 80's [19]. All these considerations explain why there is not always enough data to have statistical comparisons able to completely reject a theory or falsify an hypothesis.

Another critical aspect is the complexity of the system we want to model. In fact, we deal with systems such as forests where plants, insects and birds coexist together interacting, migrating, dying; all subjected to the laws of physics and chemistry and, at last, all influenced by random external events. Thus there is little hope to have an unified model describing every detail for every ecosystem. Instead, there are many successful models limited to the description of specific observables. For example, in the case of inter-species competition, which is in itself a vast and multifaceted phenomenon, various models are used to explain the observations. As we already said, historically the Lotka-Volterra model of prey-predator interaction was the first attempt to describe the deterministic evolution of the coupled prey-predator species, oscillating in their abundance -as shown by empirical observations. Another break-through work was *The theory of island bio-geography*, published in 1967 by MacArthur and Wilson [11]. In this book, a new stochastic oriented approach in modelling ecosystem dynamics was introduced. Following this seminal work, the Neutral Theory of Ecology was developed in the year 2001 by Hubbell [8]. Both deterministic and stochastic approaches are valid and give useful insights on the dynamics of the species abundance in a given ecological community.

Another remarkable achievement has been the discovery of common emergent macro-ecological patterns [16]. In fact, although the diversity of shapes and forms, from deserts to forests to oceans, some common properties in ecosystems are ubiquitously found and they are called emergent patterns.

In fact, emergence is a common property of *complex systems*, which are systems comprising a great number of parts which interact together and act collectively. In the realm of ecology, a lot of results have been achieved and some emergent patterns common across the most diverse communities have been universally recognized. We are interested in two of them: the RSA (relative species abundance) and the persistence time [16]. The first one is a description of the relative frequency of species abundances in the ecosystem: the number of individuals for every species are counted and an histogram of the relative frequency of species with a given abundance class is constructed. In the neutral hypothesis the RSA is equivalent to calculate the distribution of the number of individuals per specie, that is the probability for a specie to have  $n$  individuals (at stationarity). The persistence time distribution is instead the probability that a species appears (by immigration or mutation) at  $t = 0$  and migrates or goes extinct at  $t$ .

Neutral theory in fact stems from MacArthur and Wilson's idea of dynamical equilibrium between a

local community communicating with a larger metacommunity. The theory gained a lot of attraction in the physics academia because it treats ecological systems as thermodynamics treats physical systems: its aim is to model population dynamics using only few parameters and treating all species at the same trophic level as identical in order to reproduce the emerging patterns. The fact this theory presents a set of hypothesis allowing to test their validity against the data. Moreover, thanks to its simple formulation, it allows us to treat the results on the species abundance dynamics analytically and computationally.

In the first chapter we first give an overview of just the necessary mathematical concepts related to this thesis (e.g. stochastic differential equations), with a brief explanations of their intuitive meaning. We shortly introduce the generating functions, which will be exploited in the last chapter to compute the persistence times and at last we present two algorithms we used to make the computations, highlighting their strengths and limitations.

In the second chapter, we present the main mathematical framework of the Neutral Theory with emphasis on their ecological meaning: the Birth-Death Master equation, Van Kampen's system size expansion and both Janzen-Connel and Chesson-Warner hypothesis. Then we study explicitly two different models in one dimension, one analogous to the grand-canonical ensemble and one with a more realistic ecological meaning, and the stochastic version of the Lotka-Volterra equations in two dimensions.

Finally in the third chapter we analyze the persistence times for the ecological ensemble. We briefly cover this subject from two points of view: we find an analytic expression for the persistence times through the generating functions formalism and we show the results of our simulation against the theory.

# Chapter 1

## Mathematical and Computational concepts

In this chapter we are going to illustrate how a Markov stochastic processes can be described by the Chapman-Kolmogorov differential equations. We then present the Birth-Death Master equation, the Fokker-Planck equation and the Liouville equation. We will then explain how an homogeneous process becomes stationary and why this allows us to meaningfully compute the mean, the variance and the time correlation function; we will introduce the generating function and its properties, highlighting its value for studying the temporal evolution of a stochastic system. Lastly we will briefly present two simulation algorithms, the Markov and the Gillespie algorithm, and explain their differences.

### 1.1 Stochastic differential equations

We are interested in studying processes whose intrinsic nature is probabilistic, and thus evolve in time following probability laws.

This kind of process is called a stochastic process and can be described by a set of random variables  $\mathbf{x}(t) = (x_1(t), x_2(t), \dots, x_n(t))$  [15, p 77-8]. We fully know this process if, for every possible state, we know either all joint probability densities

$$p(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2, \dots) \tag{1.1}$$

or either the conditional probability densities

$$p(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i | \mathbf{y}_j, \tau_j, \dots) = \frac{p(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i; \mathbf{y}_j, \tau_j, \dots)}{p(\mathbf{y}_j, \tau_j, \dots)} \tag{1.2}$$

We define the *sample space* of the system as the set of all possible values assumed by the random variables and we refer to continuous variables with  $\mathbf{x}$ ,  $\mathbf{y}$ ,  $\mathbf{z}$  and to integer variables as  $\mathbf{n}$ ,  $\mathbf{m}$ . We restrict ourselves to stochastic processes which are also markovian.

A Markov process is a process described by a conditional probability which depends only on its most recent state, and does not explicitly depend on the previously assumed states. This can be stated by requiring the conditional probability assumes the form

$$p(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i | \mathbf{y}_j, \tau_j, \dots) = p(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i; \mathbf{y}_j, \tau_j)$$

and is a way to express that the system is "memoryless", namely its evolution depends only on its own current state and not its history. This condition is powerful because we can cast the conditional probability in the form

$$\begin{aligned} p(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i, \mathbf{y}_j, \tau_j) &= p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i, \mathbf{y}_j, \tau_j) p(\mathbf{x}_2, t_2, \dots, \mathbf{x}_i, t_i, \mathbf{y}_j, \tau_j) = \\ &= p(\mathbf{x}_1, t_1 | \mathbf{x}_2, t_2) p(\mathbf{x}_2, t_2 | \mathbf{x}_3, t_3) \dots \cdot p(\mathbf{x}_i, t_i | \mathbf{y}_j, \tau_j) p(\mathbf{y}_j, \tau_j) \end{aligned} \tag{1.3}$$

by using the Markov property multiple times. By the considerations just made and the fact that,

$$\sum_B P(A \cap B \cap C..) = P(A \cap C..)$$

we can write the integral formulation of the Chapman-Kolmogorov equation as

$$p(\mathbf{z}, t | \mathbf{y}, \tau) = \int dx^n p(\mathbf{z}, t | \mathbf{x}, t') p(\mathbf{x}, t' | \mathbf{y}, \tau).$$

This equation is the core of stochastic calculus. We can also express it in its differential form, and we assume that for all  $\epsilon > 0$ :

- $\lim_{\Delta t \rightarrow 0} \frac{p(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} = W(\mathbf{x} | \mathbf{z}, t)$  uniformly in  $\mathbf{x}$ ,  $\mathbf{z}$ , and  $t$  for  $|\mathbf{x} - \mathbf{z}| \geq \epsilon$ ;
- $\lim_{\Delta t \rightarrow 0} \frac{\int_{|\mathbf{x} - \mathbf{z}| < \epsilon} dx^n (x_i - z_i) p(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} = A_i(\mathbf{z}, t) + O(\epsilon)$  uniformly in  $\mathbf{z}$ ,  $\epsilon$  and  $t$ ;
- $\lim_{\Delta t \rightarrow 0} \frac{\int_{|\mathbf{x} - \mathbf{z}| < \epsilon} dx^n (x_i - z_i)(x_j - z_j) p(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} = B_{ij}(\mathbf{z}, t) + O(\epsilon)$  uniformly in  $\mathbf{z}$ ,  $\epsilon$  and  $t$ ;
- $\lim_{\Delta t \rightarrow 0} \frac{\int_{|\mathbf{x} - \mathbf{z}| < \epsilon} dx^n (x_i - z_i)(x_j - z_j)(x_k - z_k) p(\mathbf{x}, t + \Delta t | \mathbf{z}, t)}{\Delta t} = O(\epsilon)$  are shown to vanish [7, p.48].

Now we can write the Chapman-Kolmogorov differential equation:

$$\begin{aligned} \partial_t p(\mathbf{z}, t | \mathbf{y}, \tau) = & - \sum_i \frac{\partial}{\partial z_i} [A_i(\mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] \\ & + \sum_{i,j} \frac{1}{2} \frac{\partial^2}{\partial z_i \partial z_j} [B_{ij}(\mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] \\ & + \int dx^n [W(\mathbf{z} | \mathbf{x}, t) p(\mathbf{x}, t | \mathbf{y}, \tau) - W(\mathbf{x} | \mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] \end{aligned} \quad (1.4)$$

We note that a solution to the differential Chapman-Kolmogorov equation isn't necessarily a solution of the integral Chapman-Kolmogorov equation. It can be shown that a non-negative solution of the differential equation exists and satisfies the integral one under some non restricting assumptions:

- $\mathbf{A}(\mathbf{x}, t)$  and  $\mathbf{B}(\mathbf{x}, t)$  positive semi-definite;
- $W(\mathbf{x} | \mathbf{y}, t)$  non-negative;
- the initial condition of the system is  $p(\mathbf{z}, t | \mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z})$ .

From now on we assume these conditions to be always satisfied. In particular the initial condition specified above represents the fact that the system can be found in a precise point of the state space.

The Chapman-Kolmogorov differential equation (to which we will refer only as the Chapman-Kolmogorov equation) is made of three distinct terms that, when null in some combination, give rise to three main classes of processes. We now briefly discuss them.

### 1.1.1 The jump process and the master equation (ME)

If we put the drift and the diffusion coefficients equal to zero, we obtain the pure jump process:

$$\begin{aligned} A_i(\mathbf{z}, t) &= 0 \quad \forall i \\ B_{ij}(\mathbf{z}, t) &= 0 \quad \forall i, j \end{aligned}$$

$$\partial_t p(\mathbf{z}, t | \mathbf{y}, \tau) = \int dx^n [W(\mathbf{z} | \mathbf{x}, t) p(\mathbf{x}, t | \mathbf{y}, \tau) - W(\mathbf{x} | \mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] \quad (1.5)$$



We can write equation (1.5) in first order in  $\Delta t$ :

$$p(\mathbf{z}, t + \Delta t | \mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z}) \left[ 1 - \int dx^n W(\mathbf{x} | \mathbf{y}, t) \Delta t \right] + W(\mathbf{z} | \mathbf{y}, t) \Delta t \quad (1.6)$$

and we see that if  $W(\mathbf{z} | \mathbf{y}, t) \neq 0$  there is a finite probability that the process jumps from  $\mathbf{y} \rightarrow \mathbf{z}$  producing a discontinuity in the sample path (independently from the fact that the sample space is continuous or not); instead if  $\mathbf{y} = \mathbf{z}$  the first term expresses the probability that the system will not exit its state.

Equation (1.6) is important from a computational point of view because it can be used to obtain an approximation of the transitional probabilities, by using the finite difference method, that we can implement in an algorithm in order to simulate the evolution of a Markovian system (see subsection 1.3.1).

It's also important to express the discrete version of this equation since we will be extensively using it in the next chapters:

$$\partial_t P(\mathbf{n}, t | \mathbf{n}', t') = \sum_{\mathbf{m}} [W(\mathbf{n} | \mathbf{m}, t) P(\mathbf{m}, t | \mathbf{n}', t') - W(\mathbf{m} | \mathbf{n}, t) P(\mathbf{n}, t | \mathbf{n}', t')] \quad (1.7)$$

Its interpretation may be more clear noting that it represents the probability of a transition from state  $\mathbf{n}' \rightarrow \mathbf{n}$ : we sum over all states  $\mathbf{m}$  and each term of the summation is the difference between the probability of entering the state  $\mathbf{n}$  arriving from  $\mathbf{m}$  minus the probability of the opposite transition.

Equation (1.7) is also called the Master Equation (from now on also called ME for brevity) and is of fundamental importance in ecology because it is the simplest way of modelling the population of a specie and is at the core of Neutral Theory.

### 1.1.2 The diffusion process and the Fokker-Planck equation (FPE)

If we assume the third term of Chapman-Kolmogorov equation (eq. (1.4)) to be zero, we obtain the Fokker-Planck equation (from now on indicated as FPE for brevity)

$$\begin{aligned} W(\mathbf{z} | \mathbf{y}, t) &= 0 \\ \partial_t p(\mathbf{z}, t | \mathbf{y}, \tau) &= - \sum_i \frac{\partial}{\partial z_i} [A_i(\mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] + \sum_{i,j} \frac{1}{2} \frac{\partial^2}{\partial z_i \partial z_j} [B_{ij}(\mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] \end{aligned} \quad (1.8)$$

This equation describes a process which has a drift and a diffusion component. There are no discontinuities in its sample path because there is no jump term, nonetheless the sample path is nowhere differentiable due to the presence of diffusion. This equation is important because many cases of interest are described by a Master Equation that is often not tractable analytically so we resort to a Van Kampen's system size expansion of second order and obtain a Fokker-Planck equation, which maintains in its formulation the key behaviour of the system.

### 1.1.3 Deterministic evolution and Liouville's equation

If we assume that only the first term in the Chapman-Kolmogorov equation (eq. (1.4)) is nonzero we get Liouville's deterministic equation:

$$\begin{aligned} W(\mathbf{z} | \mathbf{y}, t) &= 0 \\ B_{ij}(\mathbf{z}, t) &= 0 \quad \forall i, j \\ \partial_t p(\mathbf{z}, t | \mathbf{y}, \tau) &= - \sum_i \frac{\partial}{\partial z_i} [A_i(\mathbf{z}, t) p(\mathbf{z}, t | \mathbf{y}, \tau)] \end{aligned} \quad (1.9)$$

The importance of this equation lies in the fact that it describes a deterministic motion analogous to what is found in classical mechanics. It does not describe the *average* evolution of the system, it describes the system neglecting all fluctuations as if it were deterministic: in this sense we can think of obtaining the Liouville equation by taking the Fokker-Planck equation and neglecting the diffusion term of the process. By defining  $\mathbf{x}(\mathbf{y}, t)$  as the solution of the ordinary differential equation

$$\frac{d\mathbf{x}(t)}{dt} = \mathbf{A}[\mathbf{x}(t), t]$$

with initial conditions  $\mathbf{x}(\mathbf{y}, \tau) = \mathbf{y}$ , the solution to equation (1.9) with initial condition  $p(\mathbf{z}, \tau | \mathbf{y}, \tau) = \delta(\mathbf{z} - \mathbf{y})$  is

$$p(\mathbf{z}, t | \mathbf{y}, \tau) = \delta[\mathbf{z} - \mathbf{x}(\mathbf{y}, t)]$$

We do not report the proof which can be found in [7, p 54].

## 1.2 Ergodicity and expected values

The fact that a stochastic process is described by a stochastic differential equation is not enough to be sure that the expected values of its random variables are well-defined. Intuitively we need the process to reach a stationary state described by a stationary probability distribution so that, if the process is not at equilibrium, we can be sure that after a certain relaxation time it has become stationary. If this condition is satisfied then the expected values computed from successive realizations of the random variables will converge, as the system evolves, to their stationary value.

### 1.2.1 Stationary process

We define a stationary process as a process which has reached equilibrium and thus successive realizations of the random variable,  $\mathbf{x}_i$  and  $\mathbf{x}_j$ , follow the same probability distribution  $p(\mathbf{x}, t | \mathbf{y}, \tau)$  but in general they may not be independent, even if the process is markovian.

A stationary process thus is only governed by a one time probability distribution or a joint probability distribution which are time-translation invariant (special case of eq. (1.1))[7, p. 57]

$$p(\mathbf{x}_1, t_1 + \epsilon; \mathbf{x}_2, t_2 + \epsilon; \dots) = p(\mathbf{x}_1, t_1; \mathbf{x}_2, t_2; \dots) = p_s(\mathbf{x}_1; \mathbf{x}_2; \dots) \quad \forall \epsilon$$

while conditional probabilities may still depend upon time differences. This allows us to set the initial condition at time  $t = 0$  for the two times conditional probabilities by substituting  $\epsilon = -\tau$

$$p(\mathbf{x}, t + \epsilon | \mathbf{y}, \tau + \epsilon) = p(\mathbf{x}, t - \tau | \mathbf{y}, 0)$$

and this is convenient for markovian processes, which can be expressed as a product of two times conditional probabilities and a one time probability (as in equation (1.3)).

### 1.2.2 Homogeneity

In the special case of a Markov process with initial conditions  $(\mathbf{y}, 0)$ , we have that by defining the one time probability and the conditional probability as realizations of the stationary distribution, we get [7, p. 60]

$$\begin{aligned} p(\mathbf{x}, t) &= p_s(\mathbf{x}, t | \mathbf{y}, 0) \\ p(\mathbf{x}, t | \mathbf{y}, 0) &= p_s(\mathbf{x}, t | \mathbf{y}, 0) \end{aligned}$$

which we can be thought as a process forced in the initial condition by some external agent and then allowed to relax back to its stationary state.

We can therefore define an homogeneous process as a Markov process that reaches stationarity independently of its initial conditions, that is it satisfies

$$\lim_{t \rightarrow \infty} p(\mathbf{z}, t | \mathbf{y}, 0) = p_s(\mathbf{z}) \quad (1.10)$$

and it becomes stationary fast enough. Clearly in the limit of  $t \rightarrow \infty$  every conditional probability of the kind of eq. (1.2) can be expressed as the product of stationary probability distributions.

Condition (1.10) is not enough to have an ergodic process because we need to reach a stationary state fast enough: we discuss this detail in the next paragraph.

### 1.2.3 Ergodicity

A Markov process is said to be ergodic if condition (1.10) holds because it guarantees that every probability distribution (which can be expressed as the product of conditional distributions and a one time distribution) reaches stationarity *and* if this condition is sufficiently rapid. A way to measure the speed of relaxation is to study the autocorrelation function, which is defined as

$$G(\tau, T) = \frac{1}{T} \int_0^T dt x(t)x(t + \tau)$$

It can be shown [7, p.59] that for  $T \rightarrow \infty$  the auto-correlation function converges in mean square to its stationary value

$$\lim_{T \rightarrow \infty} \langle (G(\tau, T) - \langle G(\tau, T) \rangle)^2 \rangle = 0 \quad \longrightarrow \quad \langle G(\tau) \rangle = \langle x(t)x(t + \tau) \rangle_s$$

If the process is stationary we have that the expected value of the mean does not change in time and thus we can write the one time correlation function as

$$R(\tau) = \langle x(t)x(t + \tau) \rangle_s - \langle x \rangle_s^2 \quad (1.11)$$

and if the one time correlation function presents exponential behaviour  $R(\tau) = A \exp(-\tau/\tau_c)$  where  $\tau_c$  is called the correlation time, the expected value of the mean converges in mean square to its stationary value. This allows us to obtain an approximation of the stationary value by averaging the values assumed by the random variable during a single realization of the process at stationarity (and not by averaging different means obtained from different realizations of the process)

$$\lim_{T \rightarrow \infty} \langle (\bar{x}(T) - \langle \bar{x}(T) \rangle)^2 \rangle = 0 \quad \longrightarrow \quad \langle x(T) \rangle = \bar{x}(T) = \sum_i x_i = \langle x \rangle_s$$

### 1.2.4 Expected values

If a process is homogeneous then it reaches stationarity for  $t \rightarrow \infty$ . If equation (1.10) holds then computing the mean values of a stochastic process is valid in the sense that, as we make more and more measurements, we get a progressively better estimate:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i(t_i) = \langle x \rangle_s \quad (1.12)$$

Most stochastic systems of interest satisfy eq. (1.10) reaching stationarity with exponential behaviour. In order to get an estimate of this behaviour we can compute the one time correlation function

$$R(\tau) = \sum_i x(t_i)x(t_i + \tau) - \sum_i x^2(t_i) \sim A \exp(\tau/\tau_c) \quad (1.13)$$

by measuring our process at stationarity. This is just an estimate because the one time correlation function approaches its stationary value in the mean square limit.

In the case of a Markov process at stationarity we can simplify our calculations by a lot. We simply report the *regression theorem*[7, p.65] which states that in the case of a one-dimensional system with variance  $\Delta n_s^2$  whose mean is given by a linear evolution equation of the type  $x(t|x_0) = -Ax(t|x_0)$ , the one time correlation function at stationarity is given by

$$\frac{dR(t)}{dt} = -AR(t) \quad \longrightarrow \quad R(t) = R(0) \exp(-t/\tau_c) \quad (1.14)$$

with initial condition  $R(0) = \Delta n_s^2$ .

### 1.2.5 Generating functions

We define the generating function of a probability distribution as [7, p. 275]

$$G(s, t) = \sum_{n=0}^{\infty} s^n P(n, t|n_0, 0) \quad (1.15)$$

where  $s$  and  $n$  are one-dimensional variables. The generating function has the following properties [7, p. 39, 276]:

- $G(s, t = 0) = s^{n_0}$
- $G(s = 0, t) = P(n = 0, t) = 1 - \mathcal{P}(t)$ , where  $\mathcal{P}(t)$  is called the survival probability.

which are easy to verify.

The main advantage of the generating function is that it allows us to find the survival probability of the system and from that we can study the persistence times distribution which is an emergent pattern common to a lot of ecosystems (see sec. 3.2).

## 1.3 Algorithms and pseudocode

In order to simulate equation (2.1) we adopt two different algorithms, the finite difference method and the Gillespie algorithm (also known as stochastic simulation algorithm, SSA).

### 1.3.1 Finite difference method

Referring to equation (1.6), which is the expansion truncated to first order in  $\Delta t$  of equation (1.7), we can rearrange its terms into a finite difference remembering the initial condition  $p(\mathbf{z}, t|\mathbf{y}, t) = \delta(\mathbf{y} - \mathbf{z})$ :

$$\frac{\Delta p(\mathbf{z}, t + \Delta t|\mathbf{y}, t)}{\Delta t} = \left[ - \int dx^n W(\mathbf{x}|\mathbf{y}, t) \delta(\mathbf{y} - \mathbf{z}) + W(\mathbf{z}|\mathbf{y}, t) \right]$$

And for processes on a discrete state space

$$\Delta P(\mathbf{n}, t + \Delta t|\mathbf{n}', t) = W(\mathbf{n}|\mathbf{n}', t) \Delta t + \sum_{\mathbf{m}} W(\mathbf{m}|\mathbf{n}', t) \Delta t \delta(\mathbf{n}' - \mathbf{n})$$

The finite difference method makes an approximation in first order  $\Delta t$  of the transition rates, causing the presence of a methodic error and creating a dependence upon  $\Delta t$  which must be chosen finding a compromise between the number of wasted draws and accuracy of the simulation. These are its own

disadvantages: reasonable values for  $\Delta t$  are of the order of  $\sim 10^{-1} \div 10^{-3}$ , but as the interval is reduced more draws are rejected; instead the methodic error is a source of uncertainty that is added to the truncation error always present in the algorithms.

We can describe the pseudo-code used as:

1. define the transition probability from  $\mathbf{n}' \rightarrow \mathbf{n}$  as  $P(\mathbf{n}|\mathbf{n}') = W(\mathbf{n}|\mathbf{n}')\Delta t$ ;
2. draw a random number  $r$  from the uniform distribution in the interval  $[0,1]$ ;
3. check if  $r \leq P(\mathbf{n}|\mathbf{n}')$ : if so update the state, if not return to step 1).

The principal advantage of this algorithm is its easiness of implementation: it was mainly used at the beginning to check the correctness of the simulation, in order to transition to the more powerful Gillespie algorithm.

### 1.3.2 Gillespie algorithm

The Gillespie algorithm is used specifically to simulate Markov processes governed by a Master Equation as (1.7). In order to do a simulation of the BDME (2.1) it's necessary to associate to it a corresponding differential equation

$$\begin{aligned} \partial_t P(n, t | n_0, 0) &= [P(n-1, t)b(n-1) + P(n+1, t)d(n+1)] - P(n, t)[b(n) + d(n)] \\ &\Downarrow \\ \dot{n} &= [b(n) - d(n)] \end{aligned} \quad (1.16)$$

instead for simulating the stochasticity for a system of two differential equations it's necessary to associate to it the corresponding BDME. We write the case for the Lotka-Volterra equations which will be studied in sec. 2.5

$$\begin{aligned} \begin{cases} \dot{x} = ak_1x - k_2xy \\ \dot{y} = k_2xy - k_3y \end{cases} \\ \Downarrow \\ \partial_t P(x, y, t | x', y', t') &= \begin{cases} [P(x-1, y, t)(x-1) - P(x, y, t)x] ak_1 + \\ [P(x+1, y, t)(x+1)y - P(x, y, t)xy] k_2 + \\ [P(x, y-1, t)x(y-1) - P(x, y, t)xy] k_2 + \\ [P(x, y+1, t)(y+1) - P(x, y, t)y] k_3 \end{cases} \end{aligned} \quad (1.17)$$

Its key concept is that a master equation can be thought as the result of  $R$  different independent reactions whose probability of happening depends only on the current state of the system. It can be shown that probability of a transition  $r_{nm}$  to state  $n$  happening at time  $t + \tau$  from state  $m$  is given by [5, chap. 4.2]

$$p(n, t + \tau | m, t) = W[n|m]e^{-W_m\tau} \quad (1.18)$$

where  $W[n|m]$  are the transition rate from state  $m$  to state  $n$  and  $W_m = \sum_n W[n|m]$  is the sum of all transition rates exiting state  $m$ .

The main advantages of the Gillespie algorithm are that it doesn't make approximations in  $\Delta t$  for the transition probabilities and the time step is automatically determined by the algorithm itself, depending on the transition rates, resulting in an exact simulation of the stochastic process. We can describe the pseudo-code used as:

1. compute the transition rates of the BDME;
2. draw a random number  $\tau$  from the exponential distribution with mean  $W_m$  and update the time;
3. draw a random number  $r$  from the uniform distribution in the interval  $[0,1]$ ;
4. determine which transition is made by checking the first  $j$  such that  $r \leq \sum_j W[j|m]/W_m$  and update the state.



# Chapter 2

## Ecological concepts

In this chapter we are going to illustrate the main ecological ideas behind Neutral Theory: its assumptions, its advantages and its disadvantages. Then we will introduce the Birth-Death Master equation for the simple case of the grand-canonical ensemble and an ecological meaningful ensemble. For both of them we will briefly explain the validity of the approximation in Taylor series of the birth and death rates, find the expected value of the mean and variance, the theoretical RSA and its continuous approximation obtained from the Fokker-Planck equation associated to the BDME.

### 2.1 Neutral Theory

Historically previous attempts at modelling inter-species competition were made by using systems of ordinary differential equations which led to a deterministic evolution of the system. This kind of approach has some flaws:

- it does not account for the intrinsic stochasticity of ecological systems;
- the number of parameters to estimate grows quadratic with system size;
- the problem becomes analytically intractable or computationally expensive as the system size grows.

Thus the lack of ecological data prevents us to estimate accurately all the parameters needed and the intractability prevents us from clearly understanding if and how universal patterns emerge from the model.

Neutral Theory comes in rescue due to its assumption:

- it treats only species at the same trophic level, thus excluding from its domain prey-predator competition;
- it is neutral, meaning species are competitively equivalent and their abundance is driven by demographic stochasticity;
- it uses birth, death, immigration and speciation as the main drivers of population change;
- it is formulated in a stochastic framework.

It can be viewed as a thermodynamic approach to modelling population dynamics because each species is thought as made of individuals all obeying the same probabilistic laws (all species have same birth and death rates). We will show that using basic ecological processes such as birth, death, immigration and speciation under the neutral hypothesis we are able to capture and model two emergent macro-ecological patterns.

We adopt a Master Equation to model population change. We define as  $\mathbf{n} = (n_1, n_2, \dots, n_S)$  the vector whose components represent the population of the specie  $i$ , where  $S$  is the number of species. In order to model species abundance dynamics we use Neutral Theory and therefore:

- we assume that the transition rates are only determined by birth, speciation and death;
- that the transition rates for each species are dependent only on their own abundance;
- we assume neutrality and therefore instead of treating the ME as a vector equation, since every component of the vector is independent, we can just write the birth-death master equation in one dimension;
- we identify the transition rates as  $b(n)$  for birth (and speciation is accounted in the term  $b(0)$ ) and  $d(n)$  for death.

For clarity we stop expressing the initial condition dependence of the BDME and just write:

$$\partial_t P(n, t | n_0, t') = [P(n-1, t)b(n-1) + P(n+1, t)d(n+1)] - P(n, t)[b(n) + d(n)] \quad (2.1)$$

where the birth rate when the population is zero,  $b(0) > 0$ , means that there is a speciation rate that prevents the extinction of the system.

If we assume that detailed balance holds, that is  $\forall n$  we have:

$$b(n)P(n) - d(n+1)P(n+1) = 0 \quad (2.2)$$

then equation (2.1) is stationary since the right argument is identically zero. The stationary solution is [7, p.237]

$$P_s(n) = P(0) \prod_{z=1}^n \frac{b(z-1)}{d(z)} \quad (2.3)$$

where  $P(0)$  is chosen as to normalize the previous equation.

We can also write the BDME in its FP form by making a Van Kampen's system size expansion [7, p.250]. Assuming we have a large total population  $N$ , able to fluctuate in the interval  $[0, \infty)$  we change variables in eq (2.1) by putting  $x = n/N$  and write

$$\frac{\partial P(x, t)}{\partial t} = \left\{ b\left(x - \frac{1}{N}\right)P\left(x - \frac{1}{N}, t\right) - b(x)P(x, t) \right\} + \left\{ d\left(x + \frac{1}{N}\right)P\left(x + \frac{1}{N}, t\right) - d(x)P(x, t) \right\} \quad (2.4)$$

lastly we can expand in Taylor series around  $x$  and truncating in second order to get a FPE assuming both  $b(x)$  and  $d(x)$  as smooth functions of  $x$ :

$$\frac{\partial p(x, t)}{\partial t} = \frac{\partial}{\partial x} [(d(x) - b(x)) p(x, t)] + \frac{1}{2} \frac{\partial^2}{\partial x^2} [(d(x) + b(x)) p(x, t)] \quad (2.5)$$

Lastly we consider how we can approximate the birth and death rates. In general we do not know the analytic expression for them and thus approximate them by expanding in a Taylor series assuming that the population  $n$  is free to live in the interval  $[0, \infty)$ :

$$\begin{aligned} \frac{b(n)}{n} &= b_0 + \frac{b_1}{n} + o\left(\frac{1}{n^2}\right) \\ \frac{d(n)}{n} &= d_0 + \frac{d_1}{n} + o\left(\frac{1}{n^2}\right) \end{aligned} \quad (2.6)$$

The previous approximation is heuristically justified because we do not know the functional form of the birth and death rates and thus resolve to making an approximation that tries to get not only the gist of the process, but also models subtler behavior without being too unrealistic by requiring the knowledge of coefficients of greater order. Mathematically not imposing any kind of hard constraint neither on the total population of the system and neither on the population of a single specie, leaving it able to fluctuate in the range  $n \in [0, \infty)$ , means that we are modelling the behaviour of a metacommunity composed of a large number of individuals. Finally there is an ecological meaning in this approximation:



we can interpret the Taylor series as the spatial density of the birth rate, where the dependence from space is implicit and it is expressed through the population density. The fact that species have a birth rate dependent on their own spatial density is also postulated in two famous hypothesis: the Janzen-Connell hypothesis and the Chesson-Warner hypothesis. The Janzen-Connell hypothesis assumes that a specie, which has an advantage at the expense of another one, stays in the proximity of it; for example predators stay in the proximity of a prey[9]. The Chesson-Warner hypothesis instead is experimentally validated and assumes that species react to fluctuations in the environment and thus their inter-species and intra-specie competition is affected. For example, in the case a specie is met with favorable environmental conditions, then it will grow faster than other species that may be not advantaged by these external conditions [4]. To model the Chesson-Warner one should introduce and environmental noise: an aspect not considered in this work.

## 2.2 The 1D Grand-Canonical Ensemble

We study the BDME (eq. (2.1)) assuming a special case of the birth and death transition probabilities. Considering the Taylor expansion of (2.6), we assume  $d_0 = d > 0$ ,  $d_1 = 0$  and  $b_0 = b_1 = b > 0$  obtaining the birth and death rates

$$b(n) = b(n+1) \quad n \geq 0 \qquad d(n) = dn \quad n \geq 0 \qquad (2.7)$$

which we can substitute in eq. (2.1) obtaining

$$\partial_t P(n, t | n_0, 0) = [P(n-1, t)bn + P(n+1, t)d(n+1)] - P(n, t) [(b+d)n + b] \qquad (2.8)$$

Detailed balance is easily seen to hold. In fact every possible path that returns to its initial condition  $n$  in our 1D system can be thought of a succession of 1 unit closed loops, and for each of these loops time-reversibility holds in fact:

$$T[n+1|n]T[n|n+1] = T[n|n+1]T[n+1|n]$$

Thus we can use eq. (2.3) in order to obtain our stationary solution

$$P_s(n) = \left(1 - \frac{b}{d}\right) \sum_{z=1}^n \frac{bz}{dz} = (1-r)r^n \qquad (2.9)$$

where we put  $r = b/d$ . The previous solution is valid only when  $b < d$ , otherwise if  $r \geq 1$  the geometric series will not converge and the probability distribution will not be normalizable. The functional form of the RSA is identical to the Bose-Einstein distribution for non-degenerate energy levels, from which it is named after.

We first proceed by calculating the expectation values and fluctuations around the mean in the stationary case. We have

$$\mathbb{E}(n) = \langle n \rangle_s = \sum_{n=0}^{\infty} nP(n) = \sum_{n=0}^{\infty} nr^n(1-r) = \frac{r}{1-r} \qquad (2.10)$$

$$\mathbb{E}[(n - \mathbb{E}(n))^2] = \Delta n_s^2 = \sum_{n=0}^{\infty} [n - \langle n \rangle_s]^2 P(n) = \sum_{n=0}^{\infty} n^2 r^n (1-r) - \langle n \rangle_s^2 = \frac{r}{(1-r)^2} \qquad (2.11)$$

We immediately note that as  $r \rightarrow 1^-$  the mean and the fluctuation diverge to infinity. In fact as we already noted  $r = 1$  is a special case where the probability distribution (2.9) is not normalizable. We also look at the ratio of the variance to the mean in order to see if there are points where the fluctuations around the mean get much larger than the mean itself, thus indicating a critical point where the system is particularly erratic. We obtain

$$\frac{\Delta n_s^2}{\langle x \rangle_s} = \frac{1}{1-r}$$

from which we see that for  $r \rightarrow 1^-$  even their ratio diverges.

In order to study the time dependence of our system the easiest approach is to make a Van Kampen's system size expansion and then obtain the corresponding Fokker-Planck equation, which we can solve in the stationary case to obtain a continuous probability distribution for the population to compare to the discrete case and ultimately, neglecting the fluctuations in the FPE, obtain the Liouville equation in order to find the time evolution of the mean. There is no hard constraint on the population of our system thus the Van Kampen's system size expansion is justified and from equation (2.5) we can write:

$$\frac{\partial p(x, t)}{\partial t} = \frac{\partial}{\partial x} \{[x(d-b) - b]p(x, t)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{[x(d+b) + b]p(x, t)\} \quad (2.12)$$

It's interesting to solve the previous equation in the stationary case by imposing the time derivative to be zero and obtaining

$$\frac{\partial_x p}{p} = -\frac{(d-b)(2x+1)}{x(d+b)+b}$$

which has solution

$$p(x) = \frac{e^{-2x\left(\frac{d-b}{d+b}\right)} [x(b+d) + b]^{-\left(\frac{d-b}{d+b}\right)^2}}{\int_0^\infty e^{-2x\left(\frac{d-b}{d+b}\right)} [x(b+d) + b]^{-\left(\frac{d-b}{d+b}\right)^2} dx} = \mathcal{N} e^{-2x\left(\frac{d-b}{d+b}\right)} [x(b+d) + b]^{-\left(\frac{d-b}{d+b}\right)^2}$$

$$\text{with } \mathcal{N} = (d+b) \left[ \frac{2(d-b)}{(d+b)^2} \right]^\alpha \frac{e^{-2b\left(\frac{d-b}{d+b}\right)^2}}{\Gamma(\alpha)}$$

It is useful to illustrate how to find the above solution starting from the standard form of the Fokker-Planck equation, so that we can underline the ecological meaning of its coefficients. By substituting the next change of variables in equation (2.12)

$$y = x(b+d) + b \qquad \tau = t(d+b)$$

we get

$$\begin{aligned} \frac{\partial p(y, \tau)}{\partial \tau} &= \frac{\partial}{\partial y} \left\{ \left[ y \frac{(d-b)}{(d+b)} - \frac{2bd}{(d+b)} \right] p(y, \tau) \right\} + \frac{(d+b)}{2} \frac{\partial^2}{\partial y^2} \{yp(y, \tau)\} \\ &= \frac{\partial}{\partial y} \left\{ [\mu y - \tilde{b}] p(y, \tau) \right\} + D \frac{\partial^2}{\partial y^2} \{yp(y, \tau)\} \end{aligned} \quad (2.13)$$

Where we have

- $\mu = 1/\tau_c = (d-b)/(d+b)$  as a drift coefficient;
- $\tilde{b} = 2bd/(d+b)$  as a speciation rate;
- $D = (d+b)/2$  as a individual death rate.

We know that equation (2.13) has a stationary solution proportional to a Gamma distribution[1, p. 15]

$$P_s(y) \propto y^{\tilde{b}/D-1} e^{-y/(D\tau_c)} = y^{(4bd/(d+b)^2-1)} e^{-2y(d-b)/(d+b)^2} \quad (2.14)$$

We can simplify this equation by making another change of variables

$$\alpha = \frac{4bd}{(d+b)^2} \qquad w = 2y \frac{(d-b)}{(d+b)^2}$$

and find that our non-normalized probability density function is  $p_s(w) = w^{(\alpha-1)} e^{-w}$ . We can normalize it and find that our final, normalized, probability density function is a Gamma distribution

$$p_s(w) = \frac{w^{(\alpha-1)} e^{-w}}{\Gamma(\alpha)}$$

By substituting back the inverse change of variables, we can get the stationary probability distribution solution of eq. (2.13) is

$$P_s(y) = \frac{\left[\frac{2(d-b)}{(d+b)^2}\right]^\alpha y^{\alpha-1} e^{-y \left[\frac{2(d-b)}{(d+b)^2}\right]}}{\Gamma(\alpha)}$$

and ultimately find the normalized probability distribution solution of eq. (2.12)

$$p(x) = (d+b) \left[\frac{2(d-b)}{(d+b)^2}\right]^\alpha \frac{e^{-2b \frac{(d-b)}{(d+b)^2}}}{\Gamma(\alpha)} [x(b+d) + b]^{-\left(\frac{d-b}{d+b}\right)^2} e^{-2x \frac{(d-b)}{d+b}} \quad (2.15)$$

Liouville's equation (eq. (1.9)) is found by dropping the term of second order in the Fokker-Planck equation and from it we can study the deterministic equation describing the temporal evolution of the population. We can write

$$\frac{\partial p(x, t)}{\partial t} = -\frac{\partial}{\partial x} \{ [x(b-d) + b] p(x, t) \}$$

from which we obtain [7, p. 54]

$$\dot{x} = [x(b-d) + b] = b(x+1) - dx \quad (2.16)$$

We can easily integrate the previous equation with initial condition  $n(0) = n_0$  and find

$$x(t|n_0) = n_0 e^{-(d-b)t} + \frac{r}{1-r} \left(1 - e^{-(d-b)t}\right) = \langle x \rangle_s + (n_0 - 1) e^{-(d-b)t} \quad (2.17)$$

and immediately note that for  $t \rightarrow \infty$  the time evolution of the population approaches the stationary value  $\forall n_0$ .

Lastly we study the correlation time of the system  $R(t) = \langle x(t_0 + t)x(t_0) \rangle - \Delta n_s^2$  with  $\Delta n_s^2$  as in definition 2.11. We can use the regression theorem because our system is described by a Markov process which has a homogeneous evolution for the mean that can be made linear by a change of variables. Thus at stationarity the one time correlation function is

$$\frac{dR(t)}{dt} = -(d-b)R(t) \quad \longrightarrow \quad R(t) = \Delta n_s^2 e^{-(d-b)t} \quad (2.18)$$

This function describes a system whose state in the limit  $t \rightarrow \infty$  becomes uncorrelated with the state assumed in  $t = 0$  and which reaches this condition with exponential behaviour. The ergodic conditions are satisfied and we can say that this is a homogeneous process with correlation time  $\tau_c = 1/(d-b)$ .

## 2.3 The Ecological Ensemble

In this section we study an ensemble which has an analytic solution and is also more suited to model ecological systems.

This ensemble uses three different parameters (one more than the grand-canonical ensemble) and with a different functional form. The birth and death rates are now

$$\begin{cases} b(0) = \nu & n = 0 \\ b(n) = bn & n \geq 1 \end{cases} \quad d(n) = dn \quad n \geq 0 \quad (2.19)$$

with  $d > 0$ ,  $b > 0$  and  $\nu > 0$ . The main difference with the Bose-Einstein ensemble is that when  $n = 0$ , corresponding to the extinction of all individuals of the specie, in order to not let the system die we introduce a speciation rate  $\nu$  which assures that a new individual of a *different* specie will restart the birth-death process. We can now write eq. (2.1) in the ecological ensemble and find:

$$\partial_t P(n, t|n_0, 0) = P(n-1, t)b(n-1) + P(n+1, t)d(n+1) - P(n, t)[(b+d)n] \quad (2.20)$$

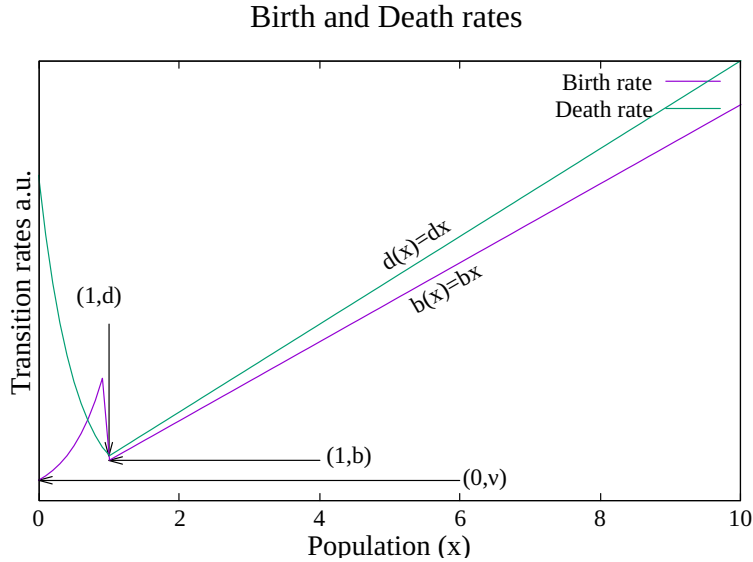


Figure 2.1: Visual representation of the transition rates.

Detailed balance is still easily seen to hold, thus we can use eq. (2.3) in order to obtain our stationary solution

$$\begin{cases} P_{s0}(0) = \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} & n = 0 \\ P_{s1}(n) = \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} \frac{r^n}{n} & n \geq 1 \end{cases} \quad (2.21)$$

where we put  $r = b/d$ . We note that if  $r \geq 1$  the logarithm series will not converge and the probability distribution will not be normalizable.

This ensemble is important because it describes ecologic dynamics independently of the underlying density distribution of the individuals; it explains well the dynamics of species living inside a meta-community composed of a large number of individuals, whose main driver of population change are the birth and death rate while the speciation rate assumes the important role of guaranteeing the appearance of new species in the metacommunity[1]. In particular it is found that, studying areas of inter-continental size, the birth and death rates obtained experimentally have the functional form of equation 2.19 with a speciation rate  $\nu \ll 1$  [12]. The distribution 2.21 was in fact proposed by Fisher[6] in order to explain the differences in the population between rare and common species and it suited well the data. Nonetheless it was challenged by Preston in 1948[14] who found that the RSA shows a triphasic mode dependent on the dimension of the sampled area.

As we did in the case of the Bose-Einstein distribution, we proceed by calculating the mean and the variance at stationarity.

$$\mathbb{E}(n) = \langle n \rangle_s = \sum_{n=1}^{\infty} n P_{s1}(n) = \sum_{n=1}^{\infty} \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} \frac{r^n}{n} = \frac{f(r, \nu)r}{1-r} \quad (2.22)$$

$$\mathbb{E}[(n - \mathbb{E}(n))^2] = \Delta n_s^2 = \sum_{n=0}^{\infty} [n - \langle n \rangle_s]^2 P_{s1}(n) = \frac{f(r, \nu)r}{(1-r)^2} [1 - f(r, \nu)r] \quad (2.23)$$

$$\frac{\Delta n_s^2}{\langle n \rangle_s} = \frac{1}{1-r} (1 - f(b, \tilde{\nu})) \quad (2.24)$$

where  $f(r, \tilde{\nu}) = \tilde{\nu} \frac{1}{1 - \tilde{\nu} \ln(1-r)}$  and  $\tilde{\nu} = \frac{\nu}{b}$ . Even in this case there is a critical point in  $r \rightarrow 1^-$  where the mean and the variance diverge and also the ratio of the variance to the mean diverges.

In order to study the time dependence of the system we do a Van Kampen's system size expansion and write the associated Fokker-Planck equation (see eq. 2.4). In this case the birth rate has a Dirac delta in zero and this is problematic because, in order to do a Van Kampen's system size expansion, we

need the transition rates to be smooth functions of  $x$ . In order to do so we divide the Fokker-Planck equation in two intervals,  $(0, 1)$  and  $[1, \infty)$ . In the first interval we will make an arbitrary choice for the transition rates while in the second interval we do not have problems in extending the domain of the transition rates from the discrete to the continuum, analogously to what we did in the Bose-Einstein ensemble. So we write

$$\begin{aligned}\frac{\partial p_0(x, t)}{\partial t} &= \frac{\partial}{\partial x} \{[d(x) - b(x)]p_0(x, t)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{[d(x) + b(x)]p_0(x, t)\} & 0 < x < 1 \\ \frac{\partial p_1(x, t)}{\partial t} &= \frac{\partial}{\partial x} \{x(d - b)p_1(x, t)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{x(d + b)p_1(x, t)\} & x \geq 1\end{aligned}$$

where we have left the transition rates unspecified in the first case of the FPE. We will choose an exponential behaviour in the interval  $(0, 1)$  for them because it will led to a uniform probability distribution as is in the discrete case, while the FPE in  $[1, \infty)$  we will use as transition rate  $b(x) = bx$  and  $d(x) = dx$ .

We proceed to solve the previous equation in the stationary case requiring the time derivative to be zero

$$0 = \frac{\partial}{\partial x} \{[d(x) - b(x)]p_0(x, t)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{[d(x) + b(x)]p_0(x, t)\} \quad 0 \leq x < 1 \quad (2.25)$$

$$0 = \frac{\partial}{\partial x} \{x(d - b)p_1(x, t)\} + \frac{1}{2} \frac{\partial^2}{\partial x^2} \{x(d + b)p_1(x, t)\} \quad x \geq 1 \quad (2.26)$$

and we focus on the first case. In order to have a uniform probability in the first interval we need the coefficient multiplying the probability distribution to be identically zero. This happens if we choose an exponential behaviour for the transition rates

$$b(x) = \nu e^{2x} \qquad d(x) = de^{-2(x-1)}$$

where we choose the coefficients as to have  $b(0) = \nu$  and  $d(1) = d$ . From this choice we obtain

$$0 = \frac{1}{2} [d(x) + b(x)] \frac{dp_0(x)}{dx}$$

and since a sum of exponential is never identically zero, we have to put  $p'_0 \equiv 0$ , thus  $p_0 = \mathcal{N}_0$ . The second case can be easily solved by integration and we get

$$0 = \mu x p_1 + D \frac{\partial}{\partial x} \{x p_1\} \quad \longrightarrow \quad p_1(x) = \mathcal{N}_1 \frac{e^{-\frac{\mu}{D}x}}{x}$$

The indefinite integral of the probability distribution on the right is the exponential integral, which diverges in  $x = 0$ . We change the names to the coefficients putting  $\mu = d - b$  and  $D = (d + b)/2$  in order to return to the standard naming in a Fokker-Planck equation; the coefficients have the same meaning as in the Bose-Einstein distribution, but we notice how there in no coefficient  $\tilde{b}$  corresponding to a speciation rate. Because of the choice made of transition rates, int the approximation to the continuum the speciation rate does not appear and the system goes extinct for  $t \rightarrow \infty$ .

Instead there is no problem in the interval  $[1, \infty)$  and we only have to normalize the piece-wise probability distribution obtained. In order to do that it's meaningful to compare the discrete RSA (eq. (2.21)) and the continuous RSA in the case of large population  $n \gg 1$ , that is  $r \rightarrow 1^-$ . We write  $r^n = \exp(n \ln r)$  and compare the two expression

$$\begin{aligned}P_{s0}(0) &= \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} & \iff & p_0 = \mathcal{N}_0 \\ P_{s1}(n) &= \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} \frac{e^{n \ln(r)}}{n} & \iff & p_1(x) = \mathcal{N}_1 \frac{e^{-\frac{\mu}{D}x}}{x}\end{aligned}$$

By putting  $d = b + \epsilon$  and  $\epsilon \rightarrow 0^+$  and expanding in zero

$$\ln \left( 1 - \frac{\epsilon}{b + \epsilon} \right) \sim -\frac{\epsilon}{b} \qquad -\frac{\mu}{D} \sim -\frac{\epsilon}{b}$$

and we recover the fact that as the population grows larger the continuous approximation is identical to the discrete case in first order of  $\epsilon$ ; we already saw that for  $r \rightarrow 1^-$  the mean diverges. From these observations we choose that in order to normalize our probability distribution we put  $\mathcal{N}_1 = \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)}$  and find  $\mathcal{N}_0$  by imposing the normalizing condition

$$\mathcal{N}_0 = 1 + \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} \text{Ei} \left( -\frac{\mu}{D} \right)$$

$$\text{Ei}(z) = \int_1^{\infty} \frac{\exp(-zx)}{x} dx \quad \text{is the Exponential integral}$$

After all these considerations we get that our final probability distribution solution of (2.25) is

$$\begin{cases} p_0(x) = 1 + \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} \text{Ei} \left( -\frac{\mu}{D} \right) & 0 < x < 1 \\ p_1(x) = \frac{\nu}{b} \frac{1}{1 - \frac{\nu}{b} \ln(1-r)} \frac{e^{-\frac{\mu}{D}x}}{x} & x \geq 1 \end{cases} \quad (2.27)$$

Lastly we study the deterministic behaviour of the system. We recover Liouville's equation from the Fokker-Planck equation and write

$$\begin{cases} \frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x} [[de^{-2(x-1)} - \nu e^{2x}]p(x,t)] & 0 < x < 1 \\ \frac{\partial p(x,t)}{\partial t} = -\frac{\partial}{\partial x} [[x(d-b)]p(x,t)] & x > 1 \end{cases} \quad (2.28)$$

and here we begin to see some problems caused by the piecewise division of the FPE and the arbitrary choice of the transition rates. In the interval  $(0, 1)$  we would have a nonsensical solution that would led our system to have a *negative* number of individuals. Therefore we choose to discard the first case and extend the second case from  $[1, \infty) \rightarrow (0, \infty)$  thus obtaining

$$\dot{x} = [x(b-d)] = bx - dx \quad (2.29)$$

We note that when  $n = 0$  there is no speciation rate prevent the complete extinction of the species. However we proceed by integrating this equation with the initial condition  $n(0) = n_0$  and find

$$x(t|n_0) = n_0 e^{-(d-b)t} \neq \langle x \rangle_s \quad (2.30)$$

and immediately note that, differently from the grand-canonical distribution, for  $t \rightarrow \infty$  the mean does not approach the stationary value obtained from the discrete probability distribution but instead is driven to extinction.

Lastly we study the correlation time of the system. We can use the regression theorem because our system is described by a Markov process whose mean is described by a linear evolution (although not quite correct in the interval  $(0, 1)$ ). Thus at stationarity the one time correlation function is

$$\frac{dR(t)}{dt} = -(d-b)R(t) \quad \longrightarrow \quad R(t) = \Delta n_s^2 e^{-(d-b)t} \quad (2.31)$$

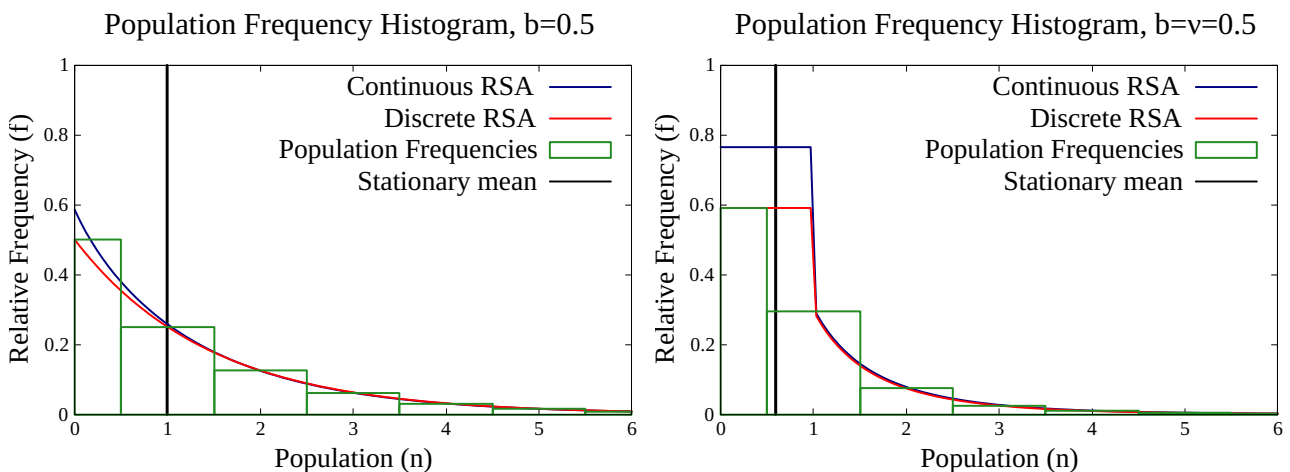
As in the grand-canonical ensemble this function describes a system whose state at time  $t \rightarrow \infty$  becomes uncorrelated with a state at  $t_0$  and which reaches this condition with exponential behaviour. The ergodic conditions are satisfied and this is in fact an homogeneous process with correlation time  $\tau_{c} = 1/(d-b)$ , identical to the Bose-Einstein distribution.

## 2.4 Results

In this section we will present the results of the simulation of both ensembles.

First we show some plots for the well behaved cases of  $r = 0.5$  (figure 2.3a for the Bose-Einstein ensemble, figure 2.3b for the ecological ensemble) and  $r = 0.9$  (figures 2.5a and 2.5c for the Bose-Einstein ensemble, figures 2.5b and 2.5d for the ecological ensemble), which are away from the critical point  $r = 1$ . For both cases and both ensembles we choose  $d = 1$  and used  $500k$  steps; we used the Gillespie algorithm to do the simulation. In order to compute the mean and the variance we discarded all simulated values which had a time  $t \leq 5 \cdot \tau_c$  (found theoretically in eq. (2.18) for the Bose-Einstein distribution and in eq. (2.31) for the Fisher distribution); moreover we used a weighted mean to compute the expectation values, whose weight for each value  $x(t_i)$  was given by the time spent by the system in that state  $\Delta t = t_{i+1} - t_i$ .

Figure 2.2: Histograms for  $b = 0.5$ ,  $\nu = 0.5$  and  $d = 1$ .



(a) Graph showing the histogram of a simulation in the Bose-Einstein ensemble.

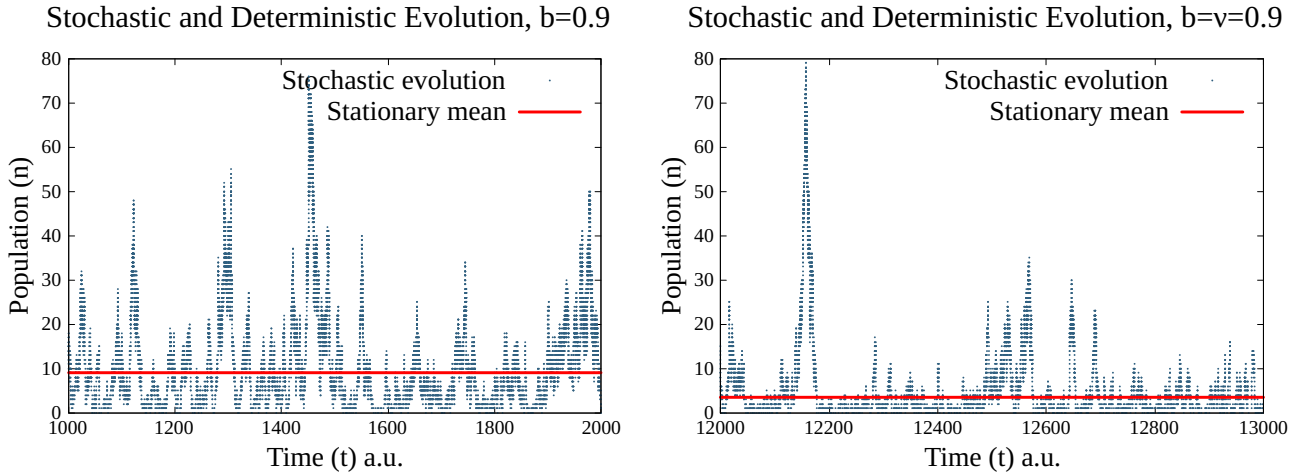
(b) Graph showing the histogram of a simulation in the ecological ensemble.

From the histograms we can clearly see the difference between the Bose-Einstein ensemble (which is displayed on the left) and the ecological ensemble (which is displayed on the right). Both simulations are perfectly fitted by the discrete probability distributions given by eq. (2.9) in the grand-canonical ensemble and by eq. (2.21) in the ecological ensemble; instead the continuous probability distributions given by eq. (2.12) in the grand-canonical ensemble and by eq. (2.27) in the ecological ensemble are seen to overestimate the probability of low values of the population, however they approximate well the tail which better satisfies the limit of large population  $n \gg 1$ . We also found that for values of  $r$  up to  $\sim 0.95$  the simulated mean and variance approximate very well the theoretical ones obtained using eq. (2.10) (grand-canonical ensemble) and eq. (2.22) (ecological ensemble).

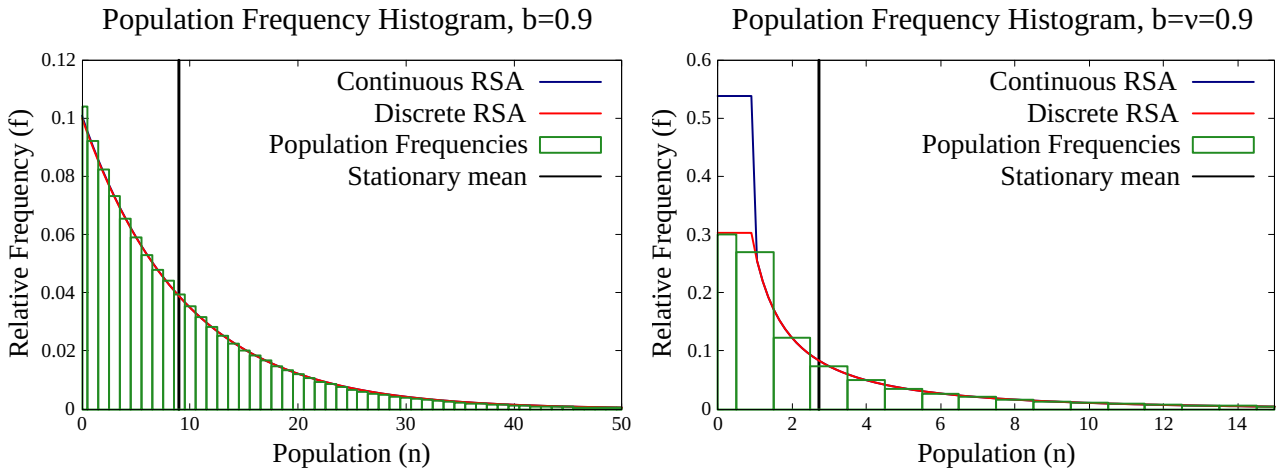
We also tried to do simulations near the critical point and found out that for  $r \sim 0.99$  we don't get correct results for the mean and variance, even if we increase the number of steps tenfold. When looking at the histograms (not reported in this work), we see that the empirical RSA lies above the discrete RSA predicted by Neutral Theory. This happens because as  $r \rightarrow 1^-$  the process becomes a random walk and in order to compute the expected values we need to average the results of multiple simulations and cannot rely upon averaging the values of just one simulation even if it is made arbitrary long: instead we have to average different simulations.

We also checked the transient time of our process for both ensembles. Comparing our theoretical predictions (eq. (2.18) for the Bose-Einstein distribution and eq. (2.31) for the Fisher distribution) with the simulations we found that if we are not too close to the critical point, our continuous deterministic evolution obtained from the Liouville equation describes very well the time evolution of the system away from equilibrium. In figure 2.7a and 2.7b we see both types of evolutions.

Figure 2.4: Time series and histograms for  $b = 0.9, \nu = 0.9$  and  $d = 1$ .

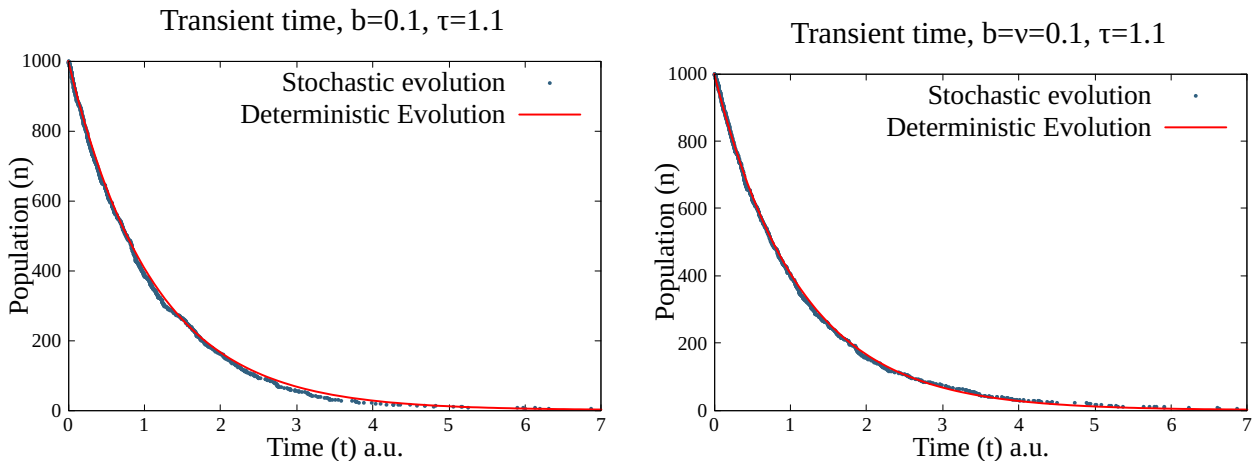


(a) Graph showing the evolution of the population in time at stationarity in the case of the Bose-Einstein ensemble. Theoretical mean and standard deviation:  $9.0 \pm 9.5$ . Simulated mean and standard deviation:  $8.7 \pm 9.2$ .  
 (b) Graph showing the evolution of the population in time at stationarity in the case of the ecological ensemble. Theoretical mean and standard deviation:  $2.7 \pm 4.5$ . Simulated mean and standard deviation:  $2.8 \pm 4.6$ .



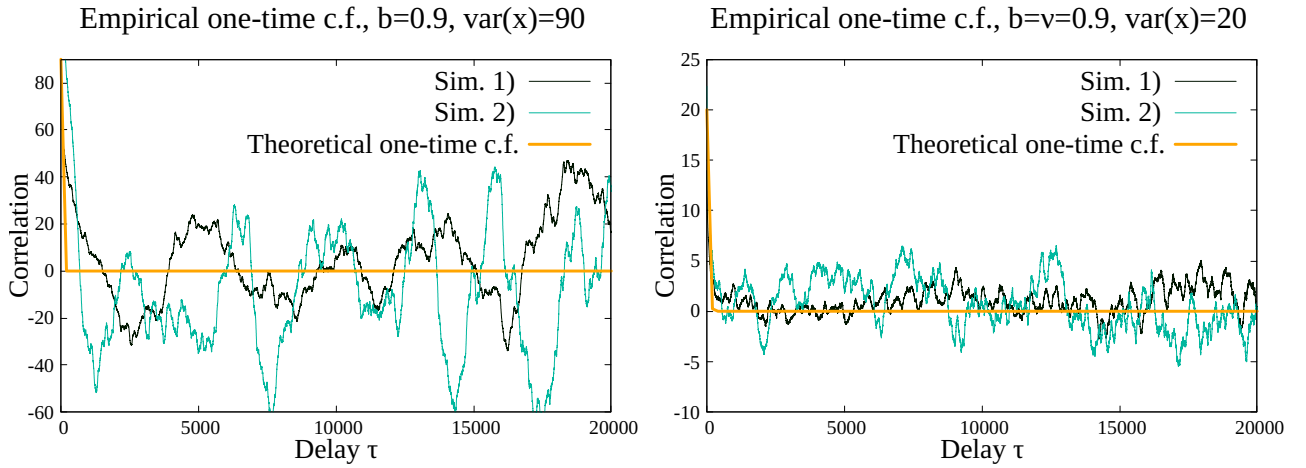
(c) Graph showing the histogram of the above time series in the Bose-Einstein ensemble.  
 (d) Graph showing the histogram of the above time series in the ecological ensemble.

Figure 2.6: Transient time for  $b = 0.1, \nu = 0.1$  and  $d = 1$ .



(a) Graph showing the transient time in the Bose-Einstein ensemble, with  $\tau_c = 1$ .  
 (b) Graph showing the transient time in the ecological ensemble, with  $\tau_c = 1$ .





(a) Graph showing two realization of the one time correlation function in the Bose-Einstein ensemble.

(b) Graph showing two realization of the one time correlation function in the ecological ensemble.

We also checked that the system has reached stationarity by computing the one-time correlation function for both ensembles. In the title of figures 2.8a and 2.8b the theoretical values of the variance are reported. The delay  $\tau$  corresponds not to the time in the simulation, but indicates the delay as a distance between the  $i$ -th value of the simulation and the  $(i + \tau)$ -th value; in the ordinate there is the value of the one-time correlation function. We can clearly see that by averaging the one-time correlation function over multiple simulations we obtain values much closer to zero; this corresponds to the fact that the theoretical one-time correlation function is given by a decaying exponential with correlation time  $\tau_c = 10$  for both ensembles.

At last we checked that our system is effectively at stationarity and that the average population of multiple simulations (and not only the average in time of one simulation) converges to the stationary value. We did it in the most interesting case for the Bose-Einstein ensemble near the critical point putting  $b = r = 0.99$  and  $d = 1$ ; in figure 2.9 the mean of 20 and 40 different simulation averaged together are presented and we can see how there are oscillations of the order of tens of individuals, who become smaller as more simulations are averaged together. For comparison, a single simulation usually can have deviations from the stationary mean of the order of hundreds of individuals.

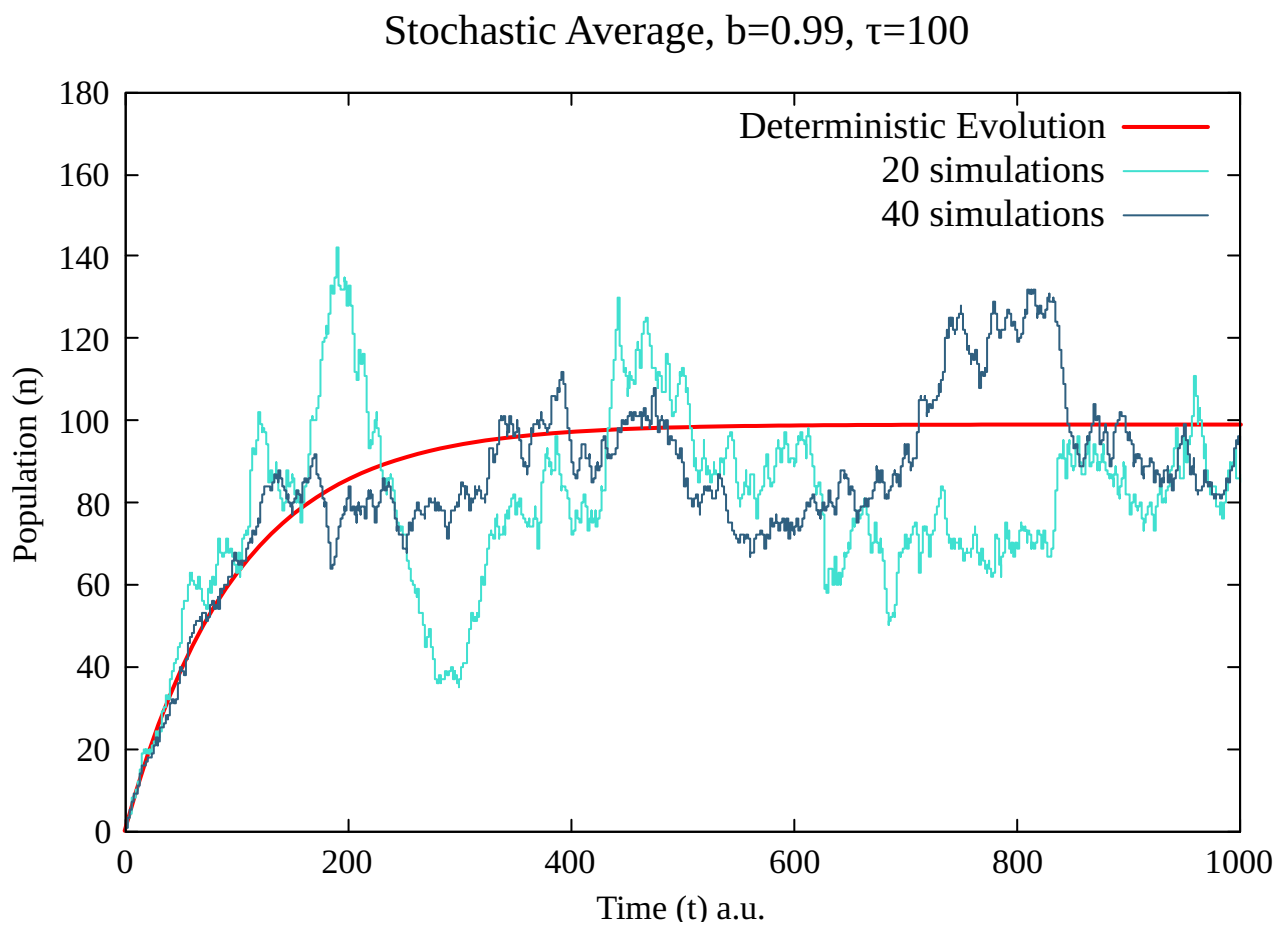


Figure 2.9: Graph showing two different averages, one over 20 and one over 40 different simulations, of the time-evolution of the population of the system.

## 2.5 The 2D Lotka-Volterra Equation

In this section we will cover shortly and qualitatively the stochastic Lotka-Volterra equation describing a prey-predator system, highlighting the main differences with the one dimensional systems studied in the previous chapter and its unique properties but without dwelling too much on the underlying math since an analytic solution to the associated BDME doesn't exist.

In order to model competition between species we can resort to differential equations. In the simple case of a prey-predator system, we immediately realize that the rate of change of preys cannot depend only on its number as was the case for mono-trophics systems governed by births and deaths, but must be influenced in some way by the fact that predators are actively hunting down the preys. The most straightforward way to model this interaction is to introduce a new term that is the product of the population of preys and predators, multiplied by a coefficient. The standard prey-predator differential equation associated to a bi-dimensional system is

$$\begin{cases} \dot{x} = ak_1x - k_2xy \\ \dot{y} = k_2xy - k_3y \end{cases}$$

and its interpretation is simple. We have 3 different terms:

- $ak_1x$  is the product of the resources generation rate  $a > 0$  that for example represents the abundance of food, of  $k_1 > 0$  which represents the probability of a prey finding the food and  $x$  which represents that if more preys are presents, more encounters will happen and preys will reproduce faster;
- $k_2xy$  is the product of the prey-predator encounter rate  $k_2 > 0$  multiplied by the numbers of both populations; it is subtracted from the population of preys and added to the population of predators because it represents the fact that a predator has successfully killed a prey and had an offspring;
- $k_3y$  is the product of the death rate  $k_3 > 0$  analogous to the one seen in the systems previously studied.

The Lotka-Volterra deterministic system of equations presents two points of equilibrium in  $\mathbf{O} = (0, 0)$  and  $\mathbf{Q} = (k_3/k_2, ak_1/k_2)$ . The presence of the stationary solution in  $\mathbf{O}$  corresponds to the extinction of both species, and the presence of the trivial solutions obtained by putting either  $x = 0$  ( $y = 0$ ) corresponding to the exponential growth(extinction) of preys(predators) limits the existence of the solution to the first quadrant, where  $x, y > 0$ . Its first integral is  $f(x, y) = k_2(x + y) - k_3 \ln(x) - ak_1 \ln(y)$  and since it is a convex function which goes to  $+\infty$  as  $x \rightarrow 0^+$  and  $x \rightarrow +\infty$  uniformly in  $y$  (and vice versa) we show that all solutions are closed curves in the first quadrant that are cyclically repeated.

As we already saw, we can find the associated Birth-Death Master equation associated to a deterministic equation. From eq. (1.3.2) we have that the 2D BDME associated to the Lotka-Volterra equations is

$$\partial_t P(x, y, t | x', y', t') = \begin{cases} [P(x-1, y, t)(x-1) - P(x, y, t)x] ak_1 + \\ [P(x+1, y, t)(x+1)y - P(x, y, t)xy] k_2 + \\ [P(x, y-1, t)x(y-1) - P(x, y, t)xy] k_2 + \\ [P(x, y+1, t)(y+1) - P(x, y, t)y] k_3 \end{cases}$$

An analytic solution doesn't exist and thus we don't have an explicit expression for the RSA. We can, however, qualitatively study this equation by doing some simulations.

Using the SSA algorithm we have to carefully choose the coefficients  $k_i$  and the starting prey/predator populations to highlight the untamed behaviour of the system. We choose  $a = 1$ ,  $k_1 = 3$ ,  $k_2 = 0.01$  and  $k_3 = 3$ . The point of equilibrium inside the first quadrant is  $\mathbf{Q} = (300, 300)$  and thus we initialize the populations to be  $(x_0, y_0) = (200, 200)$ . We choose coefficients of the order of unity because this allows us to do a more accurate simulation: in the case of the SSA they just modify the rate  $T$  and thus slow or accelerate the evolution in time of the system. We choose populations near the point of equilibrium

## Stochastic and Deterministic Evolution

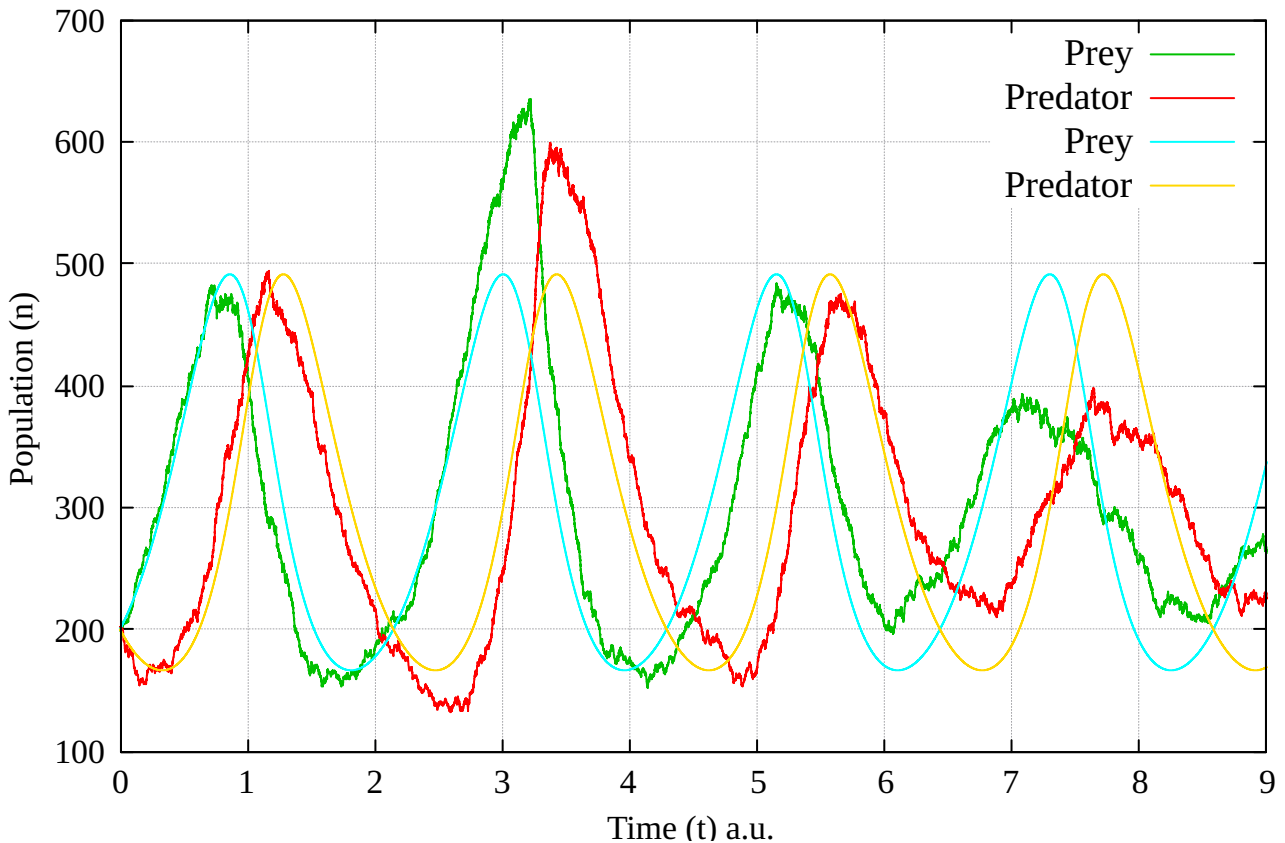


Figure 2.10: Lotka-Volterra deterministic and stochastic time series.

because, as we will see, the stochastic evolution of the system drifts away from equilibrium. The fact that  $k_1 = k_3$  simplifies even more the solution. In order to make the simulation we use  $10^5$  steps.

From the inspection of figure 2.10 we can clearly see that although the deterministic solutions of the Lotka-Volterra equations are periodic and with the same amplitude, the corresponding evolution of the stochastic process isn't so regular. Preys and predators do not have the same amplitude (although similar) neither between them neither for successive maxima or minima, and this worsens with time.

In figure 2.11 we can see the starting evolution of the system in the phase space and corresponds to the first  $10^4$  steps. We can see the start of the stochastic process in  $(200, 200)$  and how it almost follows the deterministic evolution for the first cycle (corresponding to the almost overlap of the deterministic and the stochastic evolution of the first oscillation in figure 2.10). We can recognize that if we put preys on the x-axis and predators on the y-axis the system evolves counter-clockwise because preys can be born and thus drift to the right of the x-axis, while predators can die off and thus drift to zero on the y-axis.

In figure 2.12 we can see the last  $2 \cdot 10^4$  steps. We notice how the system sweeps much larger ranges for the population of the preys and the predators, almost double of the first cycle. The stochastic evolution has definitely departed from the level curve corresponding to the initial condition. Since the system is markovian there is no way, for the system, to remember where it started and greater and greater oscillations are induced until one of the two species goes extinct. If the preys go extinct, the predator will also go extinct; if the predators die off, the preys will exponentially grow (these are the trivial solution of the deterministic Lotka-Volterra equations). This is a great problem in ecology where stochastic model predict that two species can't survive together but one, or both, eventually die off (this is also known as the *atto-fox problem*).

We lastly look at the histograms. We have the histogram of the prey population, of the predator population, and the histogram of both on the phase space. Since we choose  $k_1 = k_3$ , which ultimately

## Stochastic and Deterministic Evolution, End

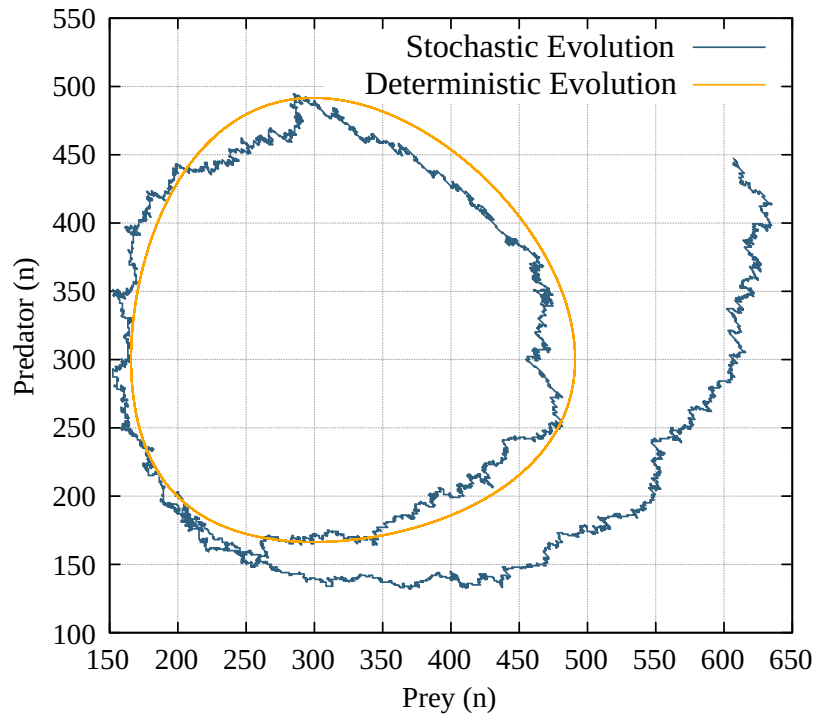


Figure 2.11: Time evolution in the phase space at the beginning.

## Stochastic and Deterministic Evolution, End

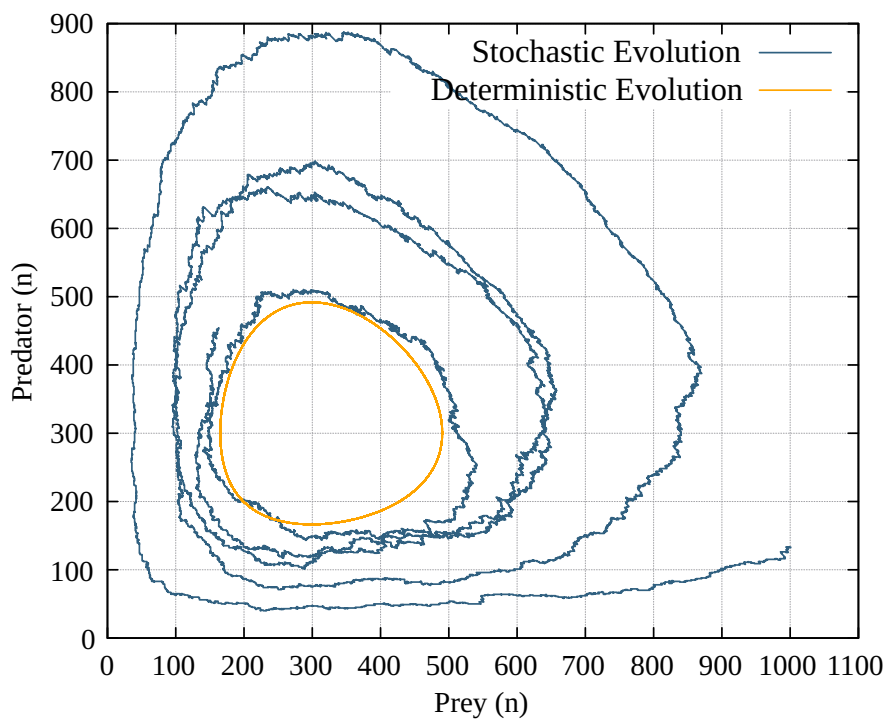


Figure 2.12: Time evolution in the phase space at the end.

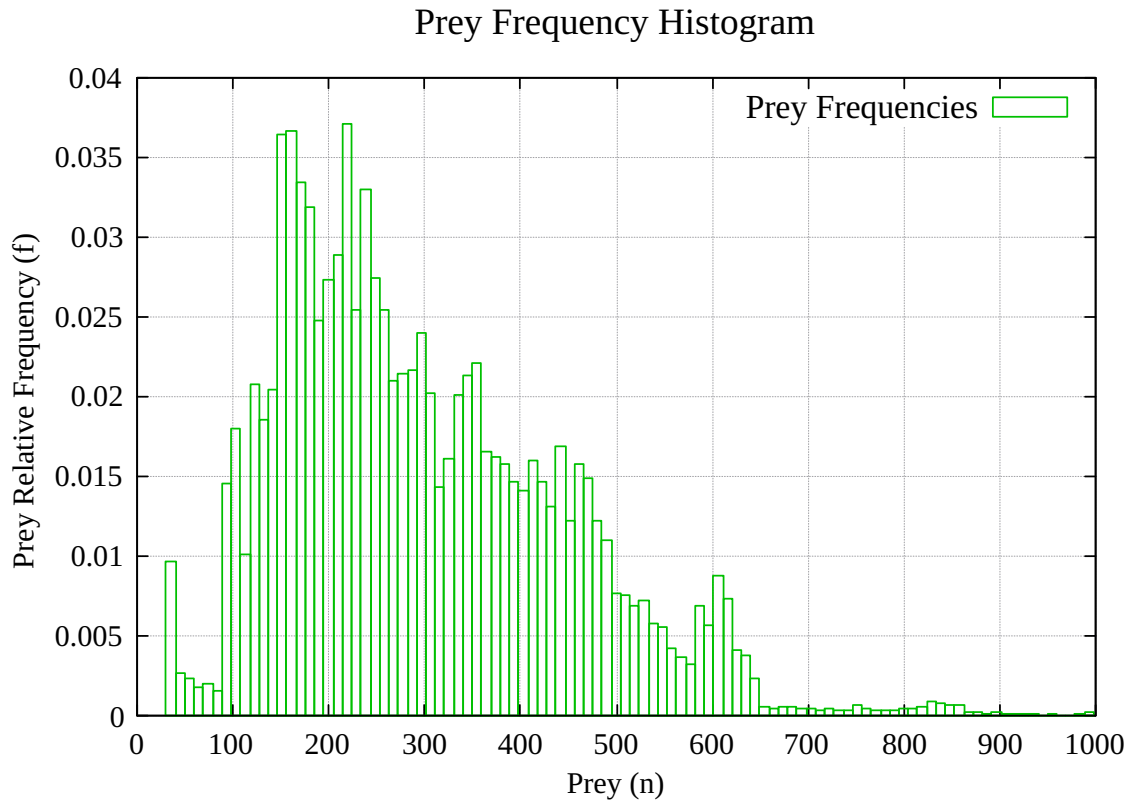


Figure 2.13: Histogram of the prey's population of the whole stochastic time series, of which figures 2.11 and 2.12 show a slice. 100 bins were used for the histogram.

means that preys are equal to predators, we get that both histograms are approximately the same. The majority of the time is spent on the low part of the point of equilibrium because it's where the rate of change  $R_{tot}$  that dictates the time increment is small but without being far enough from the point of equilibrium so that the system isn't inclined to violently rebound from the deviation.

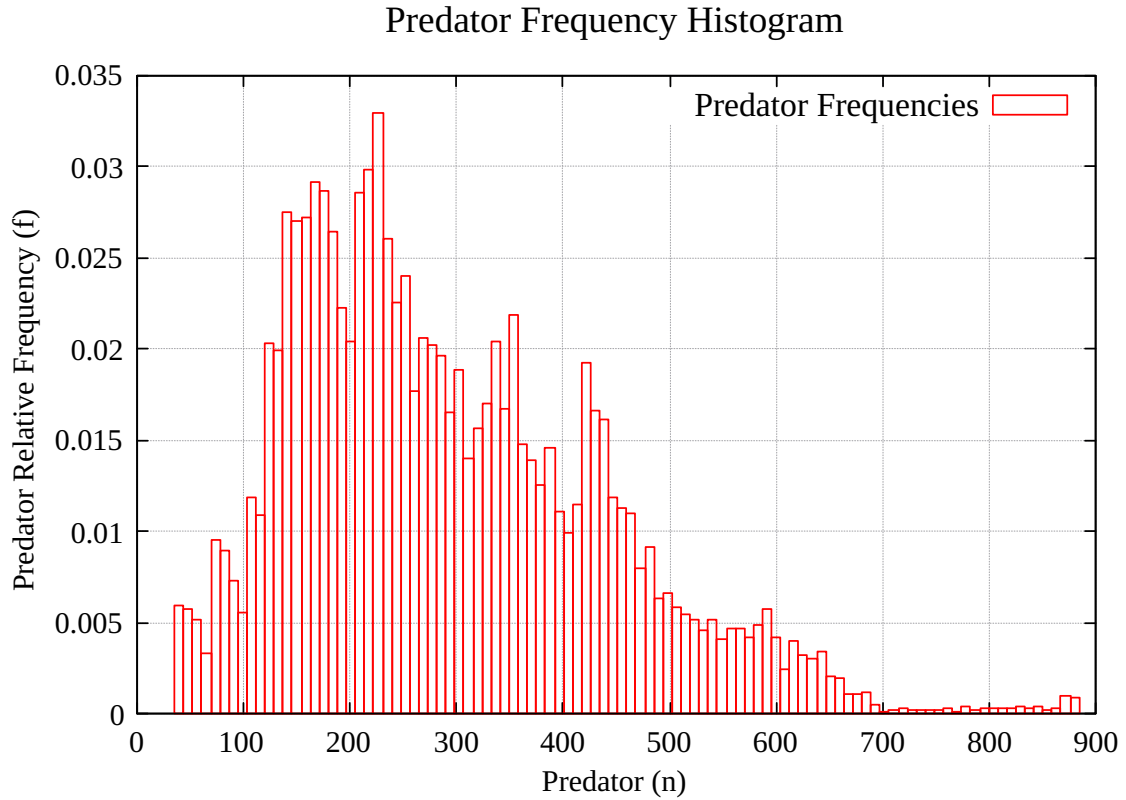


Figure 2.14: Histogram of the predator's population of the whole stochastic time series, of which figures 2.11 and 2.12 show a slice. 100 bins were used for the histogram.

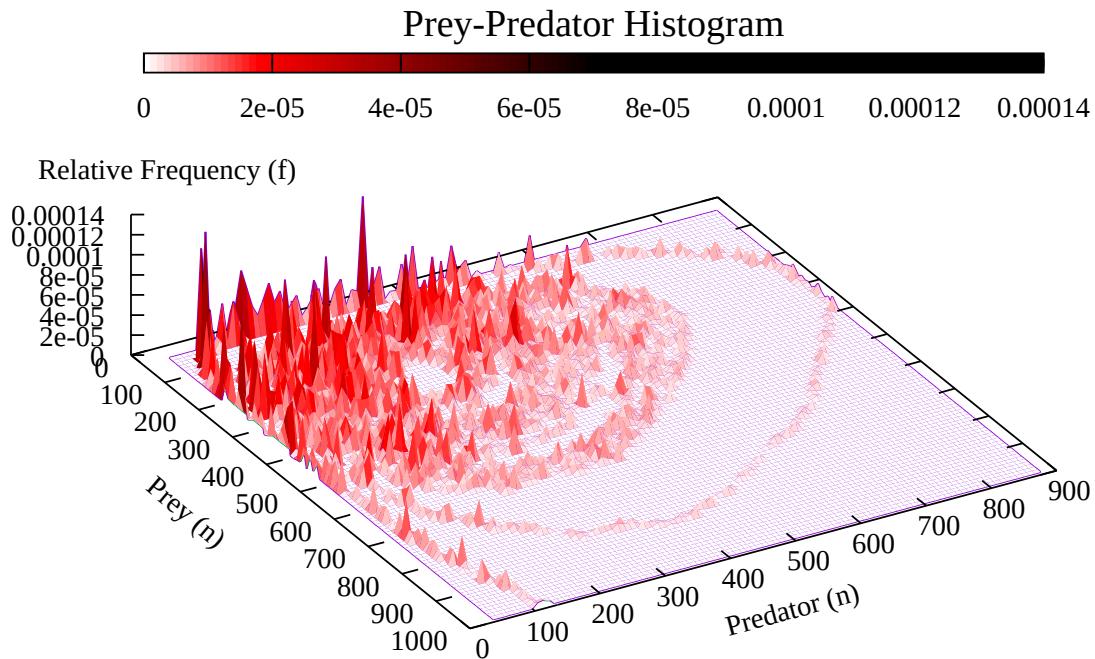


Figure 2.15: Histograms of the prey's and predator's population of the whole stochastic time series. Figures 2.11 and 2.12 show the same path in phase space as points while in this image the third dimension represents the conditional frequencies. A grid of  $100 \times 100$  bins was used.





## Chapter 3

# Persistence times Analysis

As we already presented in section 1.2.5, from the analysis of the generating function, we can define the survival probability of a system and ultimately the distribution of the persistence times of a stochastic process. We briefly recall the definition of the survival probability  $\mathcal{P}(t)$ :

$$\mathcal{P}(t) = 1 - G(s = 0, t) = 1 - P(n = 0, t),$$

i.e. the probability that a species has at least one individual at time  $t$ .

The quantity  $P(n = 0, t)$  is commonly known as the first-passage time through zero of a stochastic process and is of crucial importance when studying ecological processes because it represents the probability that, *locally*, a specie goes extinct.

### 3.1 Empirical Persistence-time distribution

Neutral Theory is based upon the assumption that the evolution of an ecological system is governed by four major processes: birth, death, migration and speciation. Migration becomes particularly important when we consider a local community which communicates with a larger metacommunity and local persistence times represent the time span for which a species can be located in the local community. Since the persistence time distribution can be described from the birth and death master equation that also governs the RSA (as seen in section 2.1), then neutral theory predict that also  $P(n = 0, t)$  will display and emergent universal shape. Indeed, it has been found[16] that empirical persistence-time distributions computed from real macro-ecological data for breeding birds, grasslands, forests and fish follow the same type of functional form, namely a power law with a cutoff.

### 3.2 Theoretical Persistence-time distribution

From the theoretical point of view, it has been shown analytically[13] that in the general for birth and death rates given by Eq. (2.6), the persistence-time distribution follows a power law with exponential cutoff

$$\rho(t|\nu) \propto t^{-\alpha} e^{-\nu t}. \quad (3.1)$$

$\rho(t|\nu)$  is described by two parameters:

- $\nu$ , that is a parameter determined by migration or speciation processes;
- $\alpha$  that is a parameter determined by the spatial structure of environment where the ecosystem lives [2]. Its value ranges from 3/2 for a one-dimensional system to 2 for systems in 3 or greater dimensions and for the mean field case, while the persistence of the system in two dimensions assumes the form  $t^{-2} \ln(t)$ .

We now find the analytic solution for the ecological ensemble.

### Ecological ensemble

We write the generating function associated to the BDME (2.20) in the ecological ensemble (eq. (2.19)) but we rewrite the birth rates as  $b(n) = bn$ , thus ignoring the Dirac delta.

$$\begin{aligned}
 \frac{\partial G(s, t)}{\partial t} &= \sum_{n=0}^{\infty} s^n \frac{\partial P(n, t)}{\partial t} \\
 &= \sum_{n=0}^{\infty} s^n \{ [P(n-1, t)b(n-1) + P(n+1, t)(n+1)d] - P(n, t)[(b+d)n] \} \\
 &= \sum_{n=0}^{\infty} s^{n-1} P(n) [n(1-s)(d-bs)] \\
 &= (1-s)(d-bs) \frac{\partial G(s, t)}{\partial s}
 \end{aligned} \tag{3.2}$$

We can solve this linear PDE by using the method of the characteristic[20]. First we divide by  $d$  and define  $r = b/d$ ; then we redefine the time scale as  $t \rightarrow t/d$ . We obtain

$$\frac{\partial G(s, t)}{\partial t} - (1-s)(1-rs) \frac{\partial G(s, t)}{\partial s} = 0 \tag{3.3}$$

and proceed to define the characteristics

$$\frac{dt}{d\sigma} = 1 \qquad \frac{ds}{d\sigma} = (1-s)(1-rs)$$

with initial condition  $t(0) = 0$  and  $s(0) = s_0$ , from which we find

$$t = \sigma \qquad s_0 = \left( \frac{1-s}{1-rs} \right)$$

By imposing the initial condition

$$G(s, 0) = s_0^{n_0} = F \left( \frac{1-s}{1-rs} \right)$$

we obtain

$$F(z) = \left( \frac{1-z}{1-rz} \right)^{n_0}$$

we find the generating function for the ecological ensemble:

$$\begin{aligned}
 G(s, t) &= \left( \frac{1-A(s, t)}{1-rA(s, t)} \right)^{n_0} \\
 A(s, t) &= \left( \frac{1-s}{1-rs} \right) e^{-(1-r)t}
 \end{aligned} \tag{3.4}$$

From the generating function we can find the persistence time distribution. First we find the survival probability

$$\mathcal{P}(t|r) = 1 - \left( \frac{1 - e^{-(1-r)t}}{1 - r e^{-(1-r)t}} \right)^{n_0}$$

and then in the scaling limit, assuming  $r \rightarrow 1^-$  and  $(1-r)t \simeq \text{const}$ , we can get an approximation

$$\mathcal{P}(t|r) \underset{r \rightarrow 1}{\sim} \frac{n_0(1-r)}{e^{(1-r)t} - 1}$$

and then by taking the time derivative we get the lifetime distribution

$$\rho(t|r) = -\frac{\partial \mathcal{P}(t|r)}{\partial t} = n_0(1-r)^2 \frac{e^{(1-r)t}}{(e^{(1-r)t} - 1)^2} \tag{3.5}$$

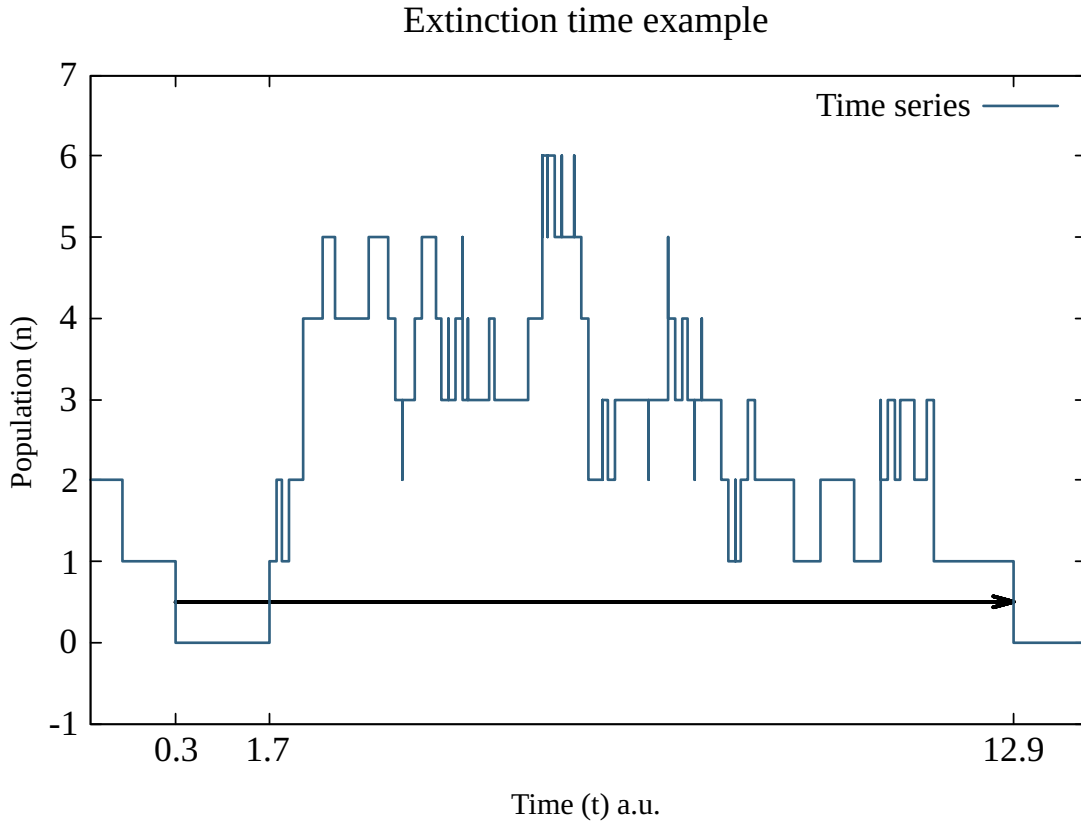


Figure 3.1: Example of a measurement of an extinction time, as indicated by the arrow. Notice how also the time where there are no individuals alive, namely from  $\sim 0.3$  to  $\sim 1.7$  is included.

### 3.3 Simulated Persistence-time distribution

In order to check the validity of eq. 3.5 we do a simulation in the ecological ensemble by:

1. set initial conditions and after that let the system evolve in time for about 100M steps;
2. discard the first part of the time series consisting of 5 times the transient time, in order to make sure the system is stationary;
3. measure the time differences of the time series corresponding to successive population extinctions;
4. after finding the time differences we simply count the frequencies without any weight and using 100 bins.

We compute the extinction times comprising also of the time spent by the system without any individuals alive. This is shown in figure 3.1, where the black arrows starts at the beginning of the extinction of the previous specie and ends as soon as the current species goes extinct again.

In figure 3.2 we can see a simulation near the critical case of  $b = r = 0.99$  and  $d = 1$ . We choose  $\nu = 0.99$  so that the system would not spend too much time extinct and can be thought as a system subject to extreme mutations; the values of speciation and of birth rate are close to the critical point in order to satisfy the scaling assumption. We see that eq (3.5) perfectly fits our data even near the critical point but the tail of the curve is not perfectly approximated because it is difficult to sample.

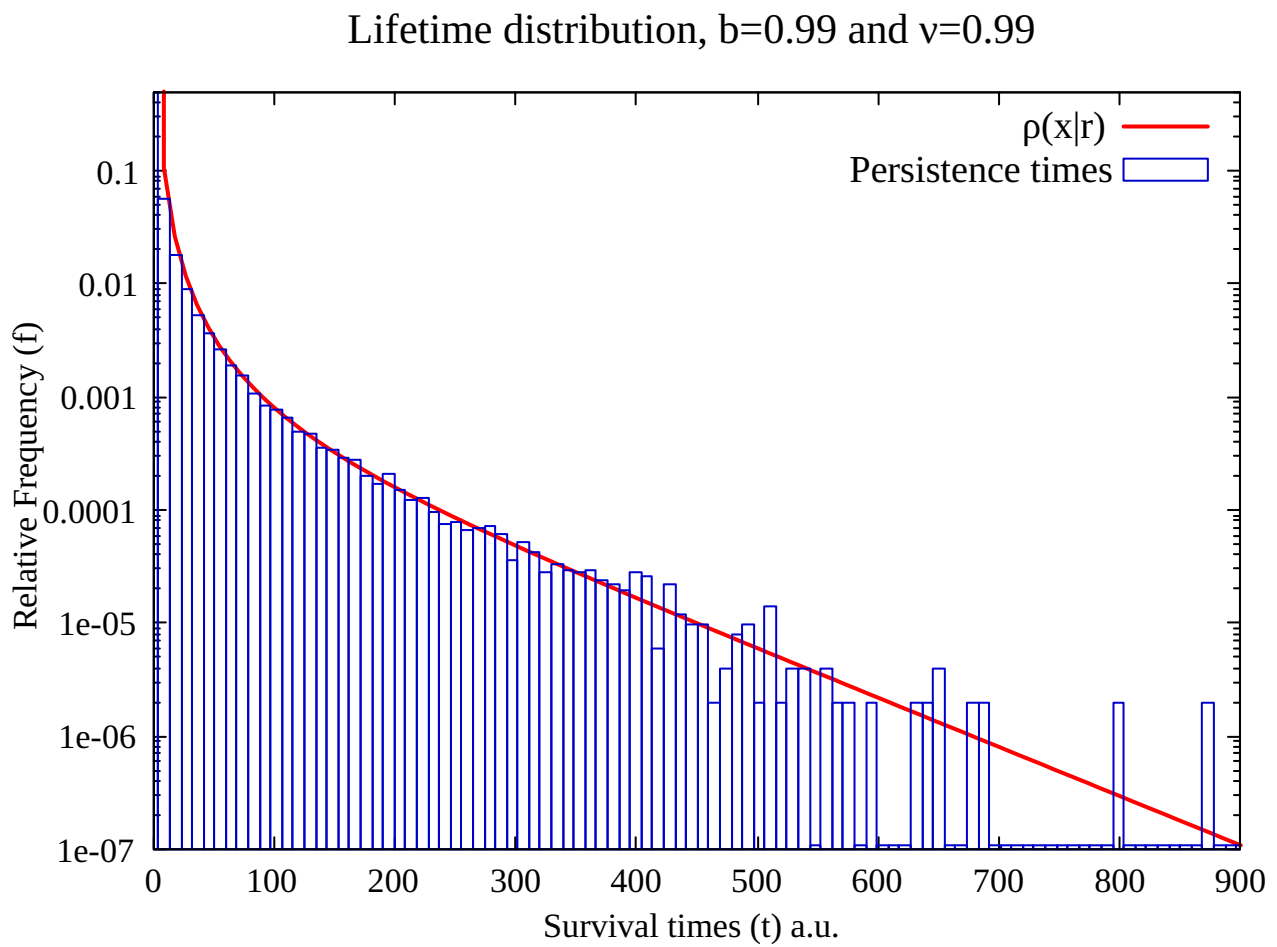


Figure 3.2: Persistence times in the ecological ensemble.  $\rho(t|r)$  is equation (3.5)

# Bibliography

- [1] Sandro Azaele et al. “Statistical Mechanics of Ecological Systems: Neutral Theory and Beyond”. In: *Reviews of Modern Physics* 88 (June 2015). DOI: 10.1103/RevModPhys.88.035003.
- [2] Enrico Bertuzzo et al. “Spatial effects on species persistence and implications for biodiversity”. In: *Proceedings of the National Academy of Sciences* 108.11 (2011), pp. 4346–4351. ISSN: 0027-8424. DOI: 10.1073/pnas.1017274108. eprint: <https://www.pnas.org/content/108/11/4346.full.pdf>. URL: <https://www.pnas.org/content/108/11/4346>.
- [3] Eric Biber. “The Challenge of Collecting and Using Environmental Monitoring Data”. In: *Ecology and Society* 18.4 (2013). DOI: 10.5751/ES-06117-180468. URL: <https://www.ecologyandsociety.org/vol18/iss4/art68/>.
- [4] Peter L. Chesson and Robert R. Warner. “Environmental Variability Promotes Coexistence in Lottery Competitive Systems”. In: *The American Naturalist* 117.6 (1981), pp. 923–943. DOI: 10.1086/283778. eprint: <https://doi.org/10.1086/283778>. URL: <https://doi.org/10.1086/283778>.
- [5] Domitilla Del Vecchio and Richard M. Murray. *Biomolecular Feedback Systems*. Usa: Princeton University Press, 2014.
- [6] R. A. Fisher, A. Steven Corbet, and C. B. Williams. “The Relation Between the Number of Species and the Number of Individuals in a Random Sample of an Animal Population”. In: *Journal of Animal Ecology* 12.1 (1943), pp. 42–58. ISSN: 00218790, 13652656. URL: <http://www.jstor.org/stable/1411>.
- [7] Crispin W. Gardiner. *Handbook of Stochastic Methods, 2nd Edition*. 1997.
- [8] STEPHEN P. HUBBELL. *The Unified Neutral Theory of Biodiversity and Biogeography (MPB-32)*. Princeton University Press, 2001. ISBN: 9780691021287. URL: <http://www.jstor.org/stable/j.ctt7rj8w>.
- [9] Daniel H. Janzen. “Herbivores and the Number of Tree Species in Tropical Forests”. In: *The American Naturalist* 104.940 (1970), pp. 501–528. DOI: 10.1086/282687. eprint: <https://doi.org/10.1086/282687>. URL: <https://doi.org/10.1086/282687>.
- [10] Alfred J. Lotka. “ELEMENTS OF PHYSICAL BIOLOGY”. In: *Science Progress in the Twentieth Century (1919-1933)* 21.82 (1926), pp. 341–343. ISSN: 20594941. URL: <http://www.jstor.org/stable/43430362>.
- [11] Robert H. MacArthur and Edward O. Wilson. *The Theory of Island Biogeography*. Princeton, NJ: Princeton University Press, 1967.
- [12] Simone Pigolotti et al. “Species lifetime distribution for simple models of ecologies”. In: *Proceedings of the National Academy of Sciences* 102.44 (2005), pp. 15747–15751. ISSN: 0027-8424. DOI: 10.1073/pnas.0502648102. eprint: <https://www.pnas.org/content/102/44/15747.full.pdf>. URL: <https://www.pnas.org/content/102/44/15747>.
- [13] Simone Pigolotti et al. “Species lifetime distribution for simple models of ecologies”. In: *Proceedings of the National Academy of Sciences* 102.44 (2005), pp. 15747–15751. ISSN: 0027-8424. DOI: 10.1073/pnas.0502648102. eprint: <https://www.pnas.org/content/102/44/15747.full.pdf>. URL: <https://www.pnas.org/content/102/44/15747>.

- [14] F. W. Preston. “The Commonness, And Rarity, of Species”. In: *Ecology* 29.3 (1948), pp. 254–283. DOI: 10.2307/1930989. eprint: <https://esajournals.onlinelibrary.wiley.com/doi/pdf/10.2307/1930989>. URL: <https://esajournals.onlinelibrary.wiley.com/doi/abs/10.2307/1930989>.
- [15] M. Ross Sheldon. *Introduction to probability models*. Usa: Elsevier, 2013.
- [16] S. Suweis et al. “On species persistence-time distributions”. In: *Journal of Theoretical Biology* 303 (2012), pp. 15–24. ISSN: 0022-5193. DOI: <https://doi.org/10.1016/j.jtbi.2012.02.022>. URL: <http://www.sciencedirect.com/science/article/pii/S0022519312000951>.
- [17] Vito Volterra. “Variazioni E Fluttuazioni Del Numero D’individui In Specie Animali Conviventi”. In: *Memoria della Regia Accademia Nazionale dei Lincei, Series 6 2* (Jan. 1926), pp. 31–113.
- [18] *History of ecology*. 2019. URL: [https://en.wikipedia.org/wiki/History\\_of\\_ecology](https://en.wikipedia.org/wiki/History_of_ecology).
- [19] *Barro Colorado Island*. 2019. URL: [https://en.wikipedia.org/wiki/Barro\\_Colorado\\_Island](https://en.wikipedia.org/wiki/Barro_Colorado_Island).
- [20] *The Method of Characteristics*. 2002. URL: <http://www.scottsarra.org/shock/shock.html>.