UNIVERSITÉ DE NEUCHÂTEL

# LONG TIME BEHAVIOR OF STOCHASTIC POPULATION MODELS

Thèse

présentée à la Faculté des sciences
pour obtenir le grade de docteur ès sciences par

**Bastien Marmet**

soutenue le 06 Juin 2014
en présence des membres du jury

| | | |
|---|---|---|
| Prof. Michel Benaïm | Université de Neuchâtel | Directeur et Rapporteur |
| Dr. Mathieu Faure | Université d'Aix-Marseille | Rapporteur |
| Prof. Florent Malrieu | Université de Tours | Examinateur |
| Prof. Christian Mazza | Université de Fribourg | Rapporteur |
| Prof. Felix Schlenk | Université de Neuchâtel | Examinateur |

Institut de mathématiques, Université de Neuchâtel,
Rue Emile–Argand 11, CH–2000 Neuchâtel

UNIVERSITÉ DE NEUCHÂTEL

**Faculté des sciences**
Secrétariat-décanat de Faculté
Rue Emile-Argand 11
2000 Neuchâtel - Suisse
Tél: + 41 (0)32 718 2100
E-mail: secretariat.sciences@unine.ch

# IMPRIMATUR POUR THESE DE DOCTORAT

**La Faculté des sciences de l'Université de Neuchâtel autorise l'impression de la présente thèse soutenue par**

## Monsieur Bastien MARMET

Titre:

## "Long time behavior of stochastic population models"

**sur le rapport des membres du jury composé comme suit:**

- Prof. Michel Benaïm, Université de Neuchâtel, directeur de thèse
- Prof. Felix Schlenk, Université de Neuchâtel
- Prof. Florent Malrieu, Université de Tours, France
- Prof. Christian Mazza, Université de Fribourg
- Dr Mathieu Faure, Université d'Aix-Marseille, France

Neuchâtel, le 12 juin 2014

Le Doyen, Prof. P. Kropf

# Abstract

The object of this thesis is the study of the asymptotic behavior of two classes of stochastic population models that both can be viewed as small random perturbations of a deterministic population dynamic. The first of these models is a discrete time Markov chain modeling the evolution of a population game with $N$ players and $d$ types. The second model is a constrained diffusion process modeling the evolution of a $d$-species system with noise. Both these processes model population dynamics where all species but one ultimately go extinct. In this work we investigate the transitory behavior of these systems before extinction. More precisely we will give results about the absorption times and the quasi-stationary distributions by linking the behavior of our systems with that of the deterministic dynamics they approximate. In the first model the appropriate tool will be the theory of stochastic approximation algorithms with constant step size. In the second model, our model lack of sufficient regularity will lead us to prove some Freidlin-Wentzell-like results.

Keywords : Markov Chain, Diffusion Process, Stochastic Approximation Algorithm, Quasi-Stationary Distributions, Large Deviations, Chain Recurrence.

# Résumé

L'objectif de cette thèse est l'étude du comportement asymptotique de deux classes de modèles stochastiques de population qui peuvent être vu comme des petites perturbations aléatoires d'une dynamique de population déterministe. Le premier de ces deux modèles est une chaîne de Markov à temps discret qui modélise l'évolution d'un jeu de population à $N$ joueurs et $d$ types. Le second modèle est un processus de diffusion sur un compact qui modélise l'évolution d'un système de population à $d$ espèces avec du bruit. Ces deux processus modélisent des systèmes de population où, à l'infini, toutes les espèces sauf une s'éteignent. Ce travail consiste en l'étude du comportement transitoire de ces processus avant extinction. Plus précisément on donnera des résultats concernant les temps d'absorption et les mesures quasi-stationnaires en reliant le comportement de nos processus avec celui du système dynamique déterministe qu'ils approximent. Dans le premier modèle les outils idoines viennent de la théorie des algorithmes d'approximation stochastiques à pas constant, tandis que, dans le second cas, notre modèle manquant de régularité, nous serons amené à prouver des résultats à la Freidlin-Wentzell.

Mots Clés : Chaîne de Markov, Processus de Diffusion, Algorithme d'Approximation Stochastique, Mesures Quasi-Stationnaires, Grandes Déviations, Récurrence par Chaîne.

# Remerciements

Mes remerciements vont d'abord à mon directeur de thèse Michel Benaïm. Il y a tant de choses pour lesquelles je lui suis redevable, merci d'avoir toujours été disponible quand j'avais besoin d'aide et d'avoir toujours su trouver l'explication juste qui illuminait d'un angle nouveau mes questions. Vous avez su trouver un juste équilibre entre présence et indépendance, me soutenant dans les passages difficiles tout en me laissant libre d'évoluer à mon rythme dans la direction qui m'intéressait. A tout moment impliqué, même de loin, vous m'avez souvent dit "Bastien, je ne vous oublie pas", moi non plus je ne vous oublierai pas Michel.

Je témoigne également ma reconnaissance à mon jury : Merci à Christian Mazza pour avoir accepté de rapporter ma thèse et m'avoir fait découvrir le concept de conférence dans une station thermale. Merci à Felix Schlenk, voisin à la ville comme au travail, auteur de nombreux bons mots devenus une sorte de mythique. Merci à Florent Malrieu qui m'a enseigné une grande partie de ce que je sait des probabilités, de par sa bonne humeur et ses talents pédagogiques il fut et reste un modèle pour moi de l'enseignant que je voudrai devenir. Enfin merci à Mathieu Faure, merci au chercheur Mathieu Faure pour son travail consciencieux de rapporteur, pour m'avoir initié aux monde des mesures quasi-stationnaire mais surtout pour avoir toujours répondu présent pour me guider au début de ma thèse, merci aussi à l'ami Mathieu pour les soirées jeux et les quizz du cerf entre Parachocos

Personne n'en arrive à faire des études pendant 10 ans sans avoir été marqué par des enseignants hors-normes, merci à tous les enseignants sans qui je n'en serais pas là, plus particulièrement merci à Olivier Adamowicz qui, le premier, m'a donné le gout des maths, merci à Nicolas Tosel qui a su me faire voir la beauté des mathématiques et merci à Florent Malrieu pour m'avoir fait découvrir le monde merveilleux de l'aléatoire.

Je me dois également de remercier une personne sans qui je me serai arraché le peu de cheveux qu'il me reste, merci à Patrice Quinton d'avoir, tout au long de cette année, eu le courage et la patience nécessaire pour me libérer d'un labyrinthe administratif ubuesque.

Tous ceux qui ont un jour travaillé à l'institut de Mathématiques de Neuchâtel ont été frappés par la bonne ambiance qui y règne et par les journées rythmées par les traditionnelles pauses cafés. J'ai eu durant mon doctorat le plaisir et l'honneur de contribuer à cet esprit saloon si particulier. Grâce à cette ambiance de studieuse détente se sont crées des amitiés qui, je l'espère, perdureront bien après mon départ. Parmi tous ces collègues devenus amis, chacun a quelque chose pour plaire, chacun à son petit mérite mais certains sortent du lot.

Merci à David, toujours de bonne humeur, même face à une déferlante de blagues douteuse à base de "Frih Deh Bi De Hu". Je garderai le souvenir des matchs de tennis partagés et, qui sait, peut-être un jour on assistera enfin à une victoire Suisse.

Merci à P.N. pour son humour si particulier, si "jurassien", merci pour tous ces footings du soir, parfois à la lampe torche, et ces excursions au ski. Au plaisir de se retrouver de nouveau sur la poudreuse.

Merci à Régis, grand chef du saloon dont je fut le modeste écuyer, j'espère avoir fait honneur à ton héritage, malgré un style surement plus direct diront certains. Merci pour toutes les soirées jeux, tu es un de ces amis discret mais dont l'absence se ferait cruellement ressentir.

Merci à Olivier pour m'avoir initié à la salle de sport de l'uni, si j'ai tant changé physiquement ces 4 dernières années, c'est en grande partie grâce à toi. Merci surtout pour tes mythiques week-end à Ste-Croix, ils resteront, sans aucun doute, le meilleur souvenir de mon séjour ici et je me réjouis par avance d'y retourner.

Merci à Yoann et Basile, les inséparables Rennais, Chaumontais, Dijonnais pour ces longues discussions de politique française, ces soirées poker et ces cours de ski.

Merci Mireille d'avoir eu la patience de supporter ma présence jour après jour en face de toi dans le bureau, merci d'avoir écouté sans jamais me dire de la fermer, résumé des infos du jour, discussions politiques ou le perpétuel feuilleton de mes mésaventures avec l'agreg. Merci aussi pour l'ambiance que tu sais mettre, que ce soit à l'institut ou au Graduate Colloquium. Avec toi à l'institut l'esprit Graduate ne mourra pas.

Je ne vous oublie pas Alex, Amandine, Ana, Antoine, Carl-Erik, Edouard, Erwann, Grégory, Kola, Laure, Miquel, Muriel, Raphaël, Sacha, Stéphane et Thibault, merci à vous et tous les autres amis et collègues, chacun à votre manière vous avez rendu mon séjour ici inoubliable.

Un grand merci à Christine, toujours présente et prête à aider avec le sourire, même face aux tracasseries administratives les plus décourageantes. Tu fut un grand soutien au cours de ces 4 années.

Merci aussi aux étudiants de Neuchâtel qui ont su supporter avec le sourire mes très longues feuilles d'exercices et mon obsession de la grammaire. L'enseignement fut un véritable bouffée d'air frais lors des moments les plus décourageants.

Avant d'arriver à Neuchâtel j'ai eu la chance de visiter d'autres villes et universités et les amis que j'ai rencontrés au cours de ces années sont aussi chers à mon coeur. Merci à Angela, ton amitié a rendu ces années de prépa agréables. Merci à Thibaut et Corinne, maîtres escrimeurs et maîtres du jeu, pour tous ces bons moments passés à Rennes et ailleurs. Merci à tous mes amis rencontrés à l'ENS, en particulier Cécile P., Cécile T., Laurent, Marie K. et Thomas D.

Merci à Christophe pour tout, la concision n'étant pas mon fort je ne rentrai pas dans ce qui deviendrai une pléthore de détails. Tu sais à quel point ton amitié m'est précieuse.

Merci enfin à mes parents, à mon frère, à ma soeur et à toute ma famille. Merci de votre soutien indéfectible quelque soit la situation et merci d'avoir toujours su me rappeler qu'il y a un monde à l'extérieur.

Enfin merci à vous qui lisez cette thèse, si j'ai pu vous apporter ou vous apprendre quelque chose, cela justifie amplement tout le temps et le travail que j' y ai consacré

*But there's no sense crying over every mistake.*
*You just keep on trying till you run out of cake.*
*And the Science gets done.*
Jonathan Coulton

# Contents

# Chapter 1

# Introduction

### Contents



## 1.1 Motivation

How do mathematical models arise ? Why do you make those assumptions ? At the very core of applied mathematics lies the need of a compromise between a model being as close as possible to the real world and us still being able to work on that model. If, for example, I'm trying to model the number of individual in an animal population, ideally I want to be able to account for the influence of every single event that might have had an influence, like the exact weather, the strength of the wind, the nature of the soil, the influence of every other animal in the neighborhood, the temperature etc. Of course it is impossible to gather all these data and even if we had them, our model would become so complicated that we couldn't say anything about it. Besides, knowing that a field mice ate some seeds nearby might give us some info on the dynamics of the rhinoceros population but that insight appears so insignificant. So, for simplicity, we will discard it or include it in a big black box called randomness. Roughly our models are built that way: We have something we want to study, what we think will have a significant influence on our model and will be observable we call it a variable and we will try to model it accurately, the rest, that is things we don't believe will matter, things we believe may have influence but can't observe/control and all we didn't think of, we put it together and call it "randomness", which is a dignified word for saying "We don't know what it is and how it works".

Now we get a model that mixes a number of observable variables and some randomness, we could write it this way: What happens at time $n+1$ for our system $X_.$ is a function of what happened in the past for our system and of some random effects, or random variables, denoted $R_.$

$$X_{n++1} = f(X_0, X_1, \cdots, X_n, R_0, R_1, \cdots, R_n)$$

Then again this will get complicated quickly, besides we don't know how the random effects $R$ depends on each other. We will have to make some assumptions to be able to predict something with our model. First we will suppose that the situation of our system at time $n$ gives us all the insight we need about the system, what happened before doesn't matter. If we now discarded the random effects completely we would get deterministic systems like recurrent sequences or ordinary differential equations. However these systems can be pretty rigid and predictable in their behavior and the real world often exhibit chaotic behavior, so we will keep our randomness. But for now our randomness is "too wild" for us to be able to say anything useful about the model, we will need to make some assumptions, those assumptions are mostly of these three rough types:

- The random effects on $X_n$ are completely independent of the random effects that happened in the past.
- The random effects are not independent from each other but knowing what the random effect was yesterday gives us all the info we could get on what the random effects will be today
- The random effects are not independent but what will happen today will, on average, be what happened yesterday.

In the first case we will say that the random variables are independent, in the second case we will say that they form a Markov process and in the last case we will say that they from a martingale (from the French "martingale" which is a piece of horse tack designed to restrained how the horse can move its head, similarly here we want to restrain chance).

These three notions are in increasing order of complexity. In this thesis we will only consider Markov process, independent variables restrain to much the possible behavior of the system and the martingale restrain to little for us to give significant results.

You can argue that, by creating a black box of things we don't know and simplifying the model to be able to work on it we might have strayed too far from the real problems. In short that our model is fundamentally wrong, to that I will respond by quoting George Box "Essentially, all models are wrong, but some are useful", of course our model is wrong but if it allows us to discover new things and/or explain what actually happens then it's all that matters.

For a long time scientists were interested in the evolution of individuals in a population, trying to changes in the size and species composition of populations, and the biological and environmental processes influencing those changes. The precursors of the field of population biology: Thomas Robert Malthus (1766-1834), Benjamin Gompertz (1779-1865) and Pierre François Verhulst (1804-1849) tried to model the growth of population with equations. Each one improved the model of its predecessors and even then their models, though crude, led to real world applications and predictions. Malthus, for example, predicted that the world population would one day exceed the production of food, leading to what is now called the "Malthusian catastrophe". Even though Malthus equation is outdated, the theory of the "Malthusian catastrophe" is still a topical subject. Another example comes from Gompertz works on human mortality, works which were then used by insurance companies to calculate costs of life insurance.

In the 20th century the field of mathematical biology was extended to the study of multi-type population dynamics by the works of Alfred James Lotka (1880-1949), Vito Volterra (1860-1940) and Crawford Stanley Holling (1930-) on predator-prey dynamics. The famous Lotka-Voltera equations were introduced independently by Lotka to model the evolution of the populations of a plant and an herbivorous animal and by Volterra using statistical data on fishes in the Adriatic Sea. Later Holling extended their model which was then used to models lots of two-species interaction, e.g. the moose-wolf interaction in Isle Royale National Park [38]. Nowadays the field of population dynamics is very vast and there is no way I could give even basic insight of all of it in this thesis. Instead I will focus on the part of it I'm most close to and that part is persistence theory.

The question behind persistence is "Under which conditions do a population or multiple interacting populations coexist ?". As we know over time some species disappear, sometimes on their own, sometimes due to human influence, however not all the species who suffer of a negative impact from human influence go extinct. We can wonder why the Dodo went extinct while the rats don't, even though humans have been trying for millennia to exterminate the latter. In the past 20 years this issue of the long-term survival of interacting populations has received an ever increasing attention in the field of populations biology. This lead to the introduction of the concepts of persistence and permanence for both deterministic models and stochastic models. In deterministic models, such as differential equations, persistence is often equated with the existence of an attractor bounded away from the extinction states, permanence also called uniform persistence requires that attractor to be global. For the past 30 years there has been an extensive literature on methods for verifying permanence and or persistence. These models provided great insight in the behavior of population models but remained rigid. In order to refine these models and allow for some "roughness" and/or influence of unpredictable outer events, randomness has been added to these models, leading to models with much more varied behavior and, one might hope, more realistic ones too. However, stochastic models such as stochastic differential equations, introduced new difficulties in the notions of persistence and permanence. The requirement that trajectories stay bounded away from the extinction states is too strong as population trajectories in stochastic models can and often will wander arbitrarily close to the extinction states. These models are then said to be stochastically persistent if there is a positive probability to remain away from extinction, see [45] for a review on the subject.

Again these models where there is a positive probability to remain away from extinction give great insight but do not allow to study the whole variety of possible behaviors. When studying finite population stochastic models, the underlying theory of Markov processes shows that, extinctions being absorbing states and species dying out with positive probabilities, extinction in finite time happens almost surely. Yet, in the real world, with large sized pools of population, we don't observe that inevitable extinction. This finite extinction time may then be very large and the system may remain in some sort of "metastable state" bounded away from extinction for a long time. These mathematical models have been corroborated by biologists who remarked that some interacting populations, while doomed to ultimately settle on an "extinction state" with some of the species going extinct seems to settle in some some kind of population equilibrium. The models we want to study are of that last type: multiple interactions species that are mathematically doomed to extinction (or at least one species will go extinct) but who still seems to maintain some kind of "persistent equilibrium" for a long time.

To convince you that such a problem isn't just a mathematical puzzle we will use a real world example : the side-blotched lizards

In the south-western part in the United States lives a very special species of lizards: Uta Stansburiana or Side-Blotched Lizards. The male individuals of this species exhibit a polymorphism in the colors of their throat. What is very interesting is that, though of the same species, these differently colored lizards have wildly different behaviors, dividing between "usurpers", "sneakers" and "guards".

The orange males are the "usurpers". They are the largest and the most aggressive. They have a large territory with many females. This make them vulnerable to the trickery cuckoldry of the "sneakers". However their size gives them the advantage when facing the "guards" whose females they then steal.

The yellow males are the "sneakers". They are the smallest and don't have a territory, instead they mimic the females and sneak into the territory of other males to mate with their females. The sheer size of the "usurpers" territory makes them vulnerable to this trickery while the "guards" watch more closely over their females and are thus not taken by the "sneakers" willy schemes.

The blue males are the "guards", they only have a small territory with only one or two females over which they watch closely. The smaller size of their territory and their vigilance makes them able to see through the "sneakers" trickery. How-[1] ever, due to them being smaller than the "usurpers" they lose the physical confrontations and their females as a consequence.

1. Pictures taken from Nature 340:240-243

When these lizards mate, the male offsprings are always of the same color as the father.

So far each color has an advantage against one type and is weak against the other. If you are familiar with game theory or playground games you might recognize the rock paper scissor dynamic a.k.a. the shifumi game.

As we said before, to be able to predict things about a real world problem with a mathematical models you have to make simplifying hypotheses. Here we will assume that the total number of lizards stays constant over time and we will "simplify" the time scale in such a way that time will be discrete and at each unit of time one lizard dies and one another is born with a random color distributed according to a law that depends on the distribution of each type in the current population. The balance between the species leads us to a model where the birth of an yellow male is more likely if most of the population is orange. Indeed if there are a lot of orange and a few yellow, each yellow individual will be able to sneak and cuckold a lot of orange lizards and thus mate with a lot of female. In the alternate situation where there are lots of blue the yellow lizards will have difficulties mating but the orange will have an advantage, so the orange population will grow and after a time the yellow will get the advantage and reproduce a lot. Then when there are a lot of yellow there is an intense competition between the yellow for reproduction and few orange lizards to cuckold, the blue lizards will then be the more efficient and their population will grow, and so the cycle continues.

The dynamic we would get is that of a Markov chain defined as such :

Let $(p_{i,j}(x))_{i,j\in\{1\dots3\}}$ be a family of real-valued continuous functions on $\Delta = \{x \in \mathbb{R}^d \mid \sum_{i=1}^{d} x_i = 1\}$ such that, for all $x \in \Delta$ :

$$\forall i \neq j \quad p_{i,j}(x) = 0 \Leftrightarrow x_i x_j = 0, \tag{1.1}$$

$$p_{i,i}(x) = 0, \tag{1.2}$$

$$0 \leqslant p_{i,j}(x) \leqslant 1, \tag{1.3}$$

$$\sum_{i,j=1}^{d} p_{i,j}(x) \leqslant 1. \tag{1.4}$$

$p_{i,j}(x)$ is the probability that, when the population is in state $x$, a $i$ lizard will die and an $j$ will be born.

Let $(X_k^N)$ be the random walk on $\Delta_N$ defined by:

$$\mathbb{P}\left[X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i) | X_k^n = x\right] = p_{i,j}(x)$$

where $(e_i)_{i=1\cdots d}$ is the canonical base of $\mathbb{R}^3$. $X_k^N$ is then the vector of type frequencies at time $k$. The jump $X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)$ means that a $i$ lizard will die and an $j$ will be born at time $k$. The conditions on the family $(p_{i,j})$ mean that :

- At each time $k$, it is always possible that a $i$ lizard dies and a $j$ lizard is born ,as long as there are $j$ lizards on the population.
- Extinct lizards type don't come back to life.

We define

$$p_i(x) = \sum_{j=1}^{d} p_{j,i}(x)$$

$$q_i(x) = \sum_{j=1}^{d} p_{i,j}(x)$$

$p(x)$ the vector of coordinates $p_i(x)$ and $q(x)$ the vector of coordinates $q_i(x)$. We also define $F(x) = p(x) - q(x)$ and we will link the behavior of $X_k^N$ to that of the dynamical system $\dot{x} = F(x)$.

Now that we have a mathematical model we can make predictions and then confront these predictions with what happens in the real world as a mathematical model is only a good model if it can predict things that actually happen. Here I will spoil the conclusion of the first article of this thesis [31] and tell you that our model predicts that, after a long time during which the population maintain some kind of equilibrium the population will ultimately rest in a single species equilibrium.

The question now is "Is this model any good ?". Of course you can't blindly trust the maths to explain the world, you have to test and challenge it. In our case biologists studied these lizards population and their conclusion is that these population currently maintain a three species equilibrium and that in the long time the population will only consist of blue males. Why blue ? Apparently blue males care more about their mates and children and have thus a slightly better fertility and also because, contrary to orange and yellow individuals, blue lizards can live peacefully together without attacking or cuckolding each other. On top of that biologists found groups of blue lizards that developed some cooperative behavior, these observations reinforces their conclusion that blue lizards are ultimately best suited for ultimate survival

What does this example tell us ? First it teaches us that, aside from moral considerations, cheating and stealing other people mates is, from an evolutionary point of view, a bad idea. In the long time all you will achieve is going extinct. Secondly it tells us that our model isn't completely stupid. Knowing that, we can proceed to the maths with peace of mind.

## 1.2 Results summary

This thesis consists of the two research article "Quasi-stationary distributions for stochastic approximation algorithms with constant step size" and "Long time behavior of 1/2 Hölder diffusion population processes".

### 1.2.1 Models introduction

At the root of these works lies a simple yet rich model of population process: $(X_k^N)$ the random walk on $\Delta_N = \Delta \cap \frac{1}{N}\mathbb{Z}^d$ defined by:

$$\mathbb{P}\left[X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)|X_k^N = x\right] = p_{i,j}(x)$$

where $(e_i)_{i=1\cdots d}$ is the canonical base of $\mathbb{R}^d$. This type of model often occurs in population games, see e.g. [42], [53], [3]. In this setting $N$ represents the size of the population. Each individual has a type $i$ and $X^N$ represents then the vector of proportion of each type. The jump $X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)$ means that an individual switches his type from $i$ to $j$ at time $k$. Under assumptions on the family $(p_{i,j})$ that :

- At each time $k$, it is always possible that a player switches from his type $i$ to another type $j$ that is currently present in the population.
- No individual switches to an absent type. This makes sense for models based on strategy switching from imitations or models arising in ecology.

Typically the coefficients $p_{i,j}(x)$ will take the form $p_{i,j}(x) = x_i x_j \lambda_{i,j}(x)$ with $\lambda_{i,j}(x) > 0$. As a consequence of this the chain will ultimately rest in one of the extinction states, that is the vertices of the simplex. If $X_n^N$ is interpreted as a vector of proportion of individual of each type, this event would mean that every type except one has gone extinct and our population doesn't evolve anymore. What we want to study is the behavior of the chain before this ultimate extinction of all except one type. Indeed, we will show that this extinction time can be very long and that, before this extinction, the system will seemingly settle in a "transitory or metastable equilibrium". We will thus study said "equilibirum before extinction".

However, when $N$ gets large, such a process gets difficult to study, its behavior depending on $O(N^{d-1}d^2)$ different coefficients. We need to make some simplifications.

Let $f$ be a function of class $\mathcal{C}^2$ and let us look at the expansion in $N$ of $\mathbb{E}[f(X_{k+1}^N)|X_k^N = x]$.

$$\begin{aligned}\mathbb{E}[f(X_{k+1}^N)|X_k^N = x] &= \mathbb{E}[f(X_{k+1}^N) - f(x)|X_k^N = x] + f(x) \\ &= f(x) + \sum_{i,j}\left(f(x + \frac{e_j - e_i}{N}) - f(x)\right)p_{i,j}(x)\end{aligned}$$

Taking $F_i(x) = \sum_j (p_{j,i}(x) - p_{i,j}(x))$ and $\sigma(x)$ such that $\sigma_{i,j}(x) = -(p_{j,i}(x) + p_{i,j}(x))$ and $\sigma_{i,i}(x) = \sum_j (p_{j,i}(x) + p_{i,j}(x))$ we obtain

$$\mathbb{E}[f(X_{k+1}^N)|X_k = x] = f(x) + \frac{1}{N} < \nabla f(x), F(x) > + \frac{1}{2n^2}Tr(D^2 f(x)\sigma) + o(\frac{1}{N^2})$$

If we only take into account the first term in the expansion we obtain an Euler scheme for approximating the ODE $\dot{x} = F(x_t)$. In the first part of this thesis we study the behavior of the process $X_{\cdot}^N$ by comparing its behavior with that of the mean-field ordinary differential equation $\dot{x} = F(x)$. In [20], Faure and Schreiber studied a similar problem for randomly perturbed discrete time dynamical systems. Their approach was a motivation and an inspiration for the results proved in [31].

If we now take into account the second order term we recognize the infinitesimal generator of a stochastic differential equation of the following form.

$$dX_t^{(N)} = F(X_t^{(N)})dt + \frac{1}{\sqrt{N}}\Sigma(X_t^{(N)})dB_t$$

where $\sigma = \Sigma\Sigma^*$. We write this process $X_t^{(N)}$ instead of $X_t^N$ to avoid confusion between the discrete time Markov chain and the diffusion process.

The study of this SDE is the subject of the second part of this thesis. In [46], Schreiber, Benaïm and Atchadé gave criteria for the persistence of a class of SDE on the $d$-dimensional simplex of the following form

$$dX_t = X_t \circ F(X_t)dt + X_t \circ \sigma(X_t)dB_t$$

The main difference of our model is the loss of the Lipschitz property of the diffusion term. This seemingly small difference will lead to a whole different behavior, we will prove that our model will be absorbed in finite time by the boundary, whereas Schreiber, Benaïm and Atchadé model remains in the relative interior of the simplex for all times.

### 1.2.2 Quasi-stationary distributions for stochastic approximation algorithms with constant step size

Our first article ([31]) setting is as such

We denote by $\Delta$ the simplex of $\mathbb{R}^d$.

$$\Delta = \{x \in \mathbb{R}^d \ ; \ \forall i = 1 \cdots d \ \ x_i \geqslant 0 \ \ \& \ \ \sum_{i=1}^{d} x_i = 1\}$$

We let $\mathring{\Delta}$ denote the relative interior of $\Delta$ and

$$\Delta_N = \Delta \cap \frac{1}{N}\mathbb{Z}^d$$

$$\mathring{\Delta}_N = \mathring{\Delta} \cap \frac{1}{N}\mathbb{Z}^d.$$

Let $F : \Delta \to \mathbb{R}^d$ be a locally Lipschitz vector field such that :

$$\forall x \in \Delta \quad \sum_{i=1}^{d} F_i(x) = 0$$

The topology considered will be the topology induced by the classical $\mathbb{R}^d$ metric topology on $\Delta$. Throughout the paper, if $A$ is a subset of a metric space $(E, d)$, we will denote by $N^\varepsilon(A)$ its $\varepsilon$-neighborhood

$$N^\varepsilon(A) = \{x \in E \ ; \ d(x, A) < \varepsilon\}.$$

We consider a family of Markov chains $(X_n^N)_{n\in\mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in the $d$-dimensional discrete simplex $\Delta_N$.

We denote by $\mathcal{F}_n^N$ the $\sigma$-algebra generated by $\{X_i^N, i = 1, ..., n\}$. For $A \in \mathcal{F}$ we let $\mathbb{P}_x[A] = \mathbb{P}[A|X_0 = x]$.

**Standing Hypothesis 1.2.1 :**
*The Markov process $X^N$ has the following properties :*

*(i)* $X_{n+1}^N - X_n^N = \frac{1}{N}(F(X_n^N) + U_{n+1}^N)$

*(ii)* $\mathbb{E}[U_{n+1}^N|\mathcal{F}_n^N] = 0$

*(iii) There exists* $\Gamma \geqslant 0$ *such that* $\|U_n^N\| \leqslant \sqrt{\Gamma}$

*(iv) The boundary of the simplex is an absorbing set:*

*(a) for all* $x \in \partial\Delta$ $\mathbb{P}_x[X_1^N \in \partial\Delta] = 1$

*(b) for all* $x \in \Delta$ $\mathbb{P}_x[\exists n : X_n^N \in \partial\Delta] = 1$

*(v)* $X^N$ *restricted to* $\mathring{\Delta}_N$ *is irreducible*

$$\forall x, y \in \mathring{\Delta}_N \quad \mathbb{P}_x[\exists n : X_n = y] > 0$$

*and aperiodic*

$$\forall x \in \mathring{\Delta}_N \quad gcd(\{n \,;\, \mathbb{P}_x[X_n = x] > 0\}) = 1$$

We then study the quasi-stationary measures of these process and link those measure to the deterministic mean-field dynamic $\dot{x} = F(x)$ and prove similar results as Faure and Schreiber in [20] regarding the absorption time of our process, the limit set of these quasi-stationary measures when $N$ goes to infinity and the support of the limiting measures.

**Definition 1.2.2:**

Let $\alpha$ be a probability measure on $\mathring{\Delta}$. We say that $\alpha$ is a **quasi-stationary distribution** (QSD) for the process $X_\cdot^N$ if, for all $n \geq 0$ and any measurable set $A \subset \mathring{\Delta}$,

$$\alpha(A) = \mathbb{P}_\alpha\left(X_n^N \in A|\tau_N > n\right).$$

where $\tau_N = \inf\{t > 0 \mid X_t^N \in \partial\Delta\}$

**Theorem 1.2.3** (3.3.2) **:**
*For all $N \in \mathbb{N}$ there exists an unique quasi-stationary distribution $\mu_N$ for the process $X_\cdot^N$.*

We also prove that the absorption time goes "very quickly" to infinity as $N$ grows large.

**Definition 1.2.4:**
Let $(\varphi_t)_{t\geqslant 0}$ be a flow generated by an ODE $\dot{x} = b(x)$. A set $A$ is called an attractor for the flow $\varphi$ if

– $A$ is invariant, i.e

$$\forall t \geqslant 0 \quad \varphi_t(x) \in A \Leftrightarrow x \in A$$

– There is a neighborhood $U$ of $A$ such that $A$ attracts $U$ in the sense that

$$\lim_{t \to \infty} d(\varphi_t(U), A) = 0$$

Without loss of generality we can suppose that $U$ is positively invariant, i.e that $\forall t \geqslant 0 \quad \varphi_t(x) \in U \Rightarrow x \in U$. Indeed, should $U$ not be positively invariant we can consider $V = \gamma_+(U) = \{\varphi_t(x) \, | t \geqslant 0, x \in U\}$ which is still a neighborhood of $A$, is still attracted by $A$ and is positively invariant.

**Theorem 1.2.5** (3.3.3) **:**
*Starting from the quasi-stationary distribution $\mu^N$, the law of the absorption time is a geometric law of parameter $1 - \rho_N$. If we further assume that the deterministic mean-field dynamic $\dot{x} = F(x)$ admits an attractor $A \subset \mathring{\Delta}$, then, there exists $\gamma > 0$ such that the following estimate holds :*

$$0 \leqslant 1 - \rho_N \leqslant O\left(\frac{e^{-\gamma N}}{N}\right)$$

*Thus, there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mu^N}[\tau_N] \geqslant C N e^{\gamma N}$$

From a pratical point of view, such a extinction time may be so large that we might never observe it. It thus makes sense to look at what happens before this absorption. We will then look more closely at the quasi-stationary measures of our process which have the role of an "equilibrium before extinction" and give more information about these measures when the population is very large.

**Theorem 1.2.6** (3.3.4) **:**
*We suppose that the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$. Then the set of limit points of $\{\mu^N\}$ for the weak* topology is a subset of the set of invariant measures for the flow $\{\varphi_t\}$.*

Knowing that our QSD are "close" to an invariant measure for the deterministic dynamic is enlightening but not precise enough as these limit measure could well have their support of the boundary and we want to look at our problem through the prism of persistence, that is we want conditions to ensure some kind of multi-species equilibrium. To obtain such results we need some additional assumptions.

**Theorem 1.2.7** (3.4.9) **:**
*We suppose that the flow $\{\varphi_t\}$ associated with the mean dynamic $\dot{x} = F(x)$ has an attractor $A \subset \mathring{\Delta}$. We also suppose that the process satisfies some additional hypotheses, in particular we suppose that $X^N_{\cdot}$ satisfies a large deviation principle. Then the limiting measure $\mu$ has its support in the union of the $\mathcal{L}$-quasi-attractors, in particular $Supp(\mu) \subset \mathring{\Delta}$.*

The $\mathcal{L}$-quasi-attractors are sets defined by the large deviations rate functions but they have the nice properties of also be attractors for the deterministic dynamic.

### 1.2.3 Long time behavior of 1/2 Hölder diffusion population processes

In the second article we study the behavior of a family of processes on the $d$-dimensional simplex $\Delta$ that arise from the same multi-type population process as before. The lack of Lipschitz regularity in the diffusion term lead to a behavior very different from the Lipschitz case. Our setting is as such

We will denote by $\circ$ the component by component product in $\mathbb{R}^d$.

$$(x_1, x_2, \cdots x_d) \circ (y_1, y_2, \cdots y_d) = (x_1 y_1, x_2 y_2, \cdots, x_d y_d)$$

We consider a family of Markov processes $(X_t^N)_{t \in \mathbb{R}_+}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\Delta$ defined by

$$dX_t^{(N)} = X_t^{(N)} F(X_t^{(N)}) dt + \frac{1}{\sqrt{N}} \sqrt{X_t^{(N)}} \circ \sigma(X_t^{(N)}) dB_t \tag{1.5}$$

Throughout the chapter, these hypotheses will always be assumed to hold

**Hypothesis 1.2.8 :**

*(i)* $F : \Delta \to \mathbb{R}^d$ *is a L-Lipschitz vector field*

*(ii)* $\forall x \in \Delta \quad \sum_{i=1}^{d} x_i F_i(x) = 0$

*(iii)* $\sigma$ *is a continuously derivable application from* $\Delta$ *to* $\mathcal{M}_{d,l}(\mathbb{R})$

*(iv)* $\forall x \in \Delta$ *and* $\forall j \in \{1, \cdots, l\} \quad \sum_{i=1}^{d} \sqrt{x_i} \sigma_{i,j}(x) = 0$

*(v)* *For all* $i \in \{1 \cdots d\}$ *and all* $x \in \Delta$, *we have* $(\sigma\sigma^*)_{ii}(x) > \varepsilon$

We denote by $\mathcal{F}_n^N$ the $\sigma$-algebra generated by $\{X_i^N, i = 1, ..., n\}$. For $A \in \mathcal{F}$ we let $\mathbb{P}_x[A] = \mathbb{P}[A | X_0 = x]$.

When studying SDE of the form

$$dX_t = X_t \circ F(X_t) dt + X_t \circ \sigma(X_t) dB_t$$

a simple exponential martingale argument or the use of the strong uniqueness property show that $X_t \in \mathring{\Delta}$ almost surely for all $t$. Such a behavior is no more true when the diffusion term is no more Lipschitz, in fact we get that

**Theorem 1.2.9** (4.3.1) **:**
*Let* $\varepsilon > 0$. *Let* $\tau_N = Inf\{t > 0\,;\ X_t^{(N)} \in \partial\Delta\}$.
*Then* $\mathbb{P}_x[\tau_N < \infty] = 1$

We now know that our process goes extinct almost surely, but what happens before this extinction ? Similarly to what happened in the random walk model, is there a quasi-stationary distribution ? The answer is yes.

**Theorem 1.2.10** (4.4.3) **:**
*For all $N$ there exists a QSD for the process $X^N_\cdot$*

Again we are interested in what happens when the population size $N$ is very large. We will thus look more into the behavior of the extinction times and the QSD when $N$ goes to infinity.

**Theorem 1.2.11** (4.4.7) **:**
*Starting from $\mu^N$, the law of the absorption time has an exponential law with parameter $\theta(\mu_N)$. The following estimate holds :*

$$0 \leqslant 1 - e^{\theta_N} \leqslant O\left(\frac{1}{N}\right)$$

*where $\theta_N = \theta(\mu^N)$.*

*Thus, there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mu^N}[\tau_N] \geqslant CN$$

**Theorem 1.2.12** (4.4.8) **:**
*We suppose that the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$. Then the set of limit points of $\{\mu^N\}$ for the weak* topology is a subset of the set of invariant measures for the flow $\{\varphi_t\}$.*

Due to the non-Lipschtiz character of our diffusion term the traditional Freidlin-Wentzell theory don't apply. However, as we saw in [31], large deviations principle can give us insights on the behavior of our process and are thus always interesting to have. As a consequence we proved large number and large deviations results analogous to the classical results, however the lack of regularity on the diffusion term forces our results to be much weaker than the classical Lipschitz results. These results apply to more general diffusion processes, namely we study $X^\varepsilon_t$ such that

$$dX^\varepsilon_t = b(X^\varepsilon_t)dt + \sqrt{\varepsilon}\Sigma(X^\varepsilon_t)dB_t$$

where $X^\varepsilon_0 = x \in \Delta$, $t \in [0, 1]$, $b$ is Lipschitz continuous with constant $L$ and $\Sigma$ is non negative, $1/2$ Hölder continuous with the same constant $L$. Both $\Sigma$ and $b$ vanish on $\partial\Delta$ such that $X^\varepsilon_t \in \Delta$ a.s. We want to compare the trajectories of the SDE with those of the ODE.

$$\dot{x}_t = b(x_t)$$

**Theorem 1.2.13** (4.5.1) **:**
*Let $m_\varepsilon(T) = \sup_{0 \leqslant s \leqslant T} \|X^\varepsilon_s - x_s\|$, where $x_s$ is the solution of the ODE $\dot{x}_t = b(x_t)$ with*

*the initial condition* $x_0 = X_0^\varepsilon$

$$\forall \delta > 0 \quad \mathbb{P}[m_\varepsilon(T) \geqslant \delta] \leqslant \frac{T\varepsilon\|\Sigma\|_\infty}{\delta}$$

*In particular, we get*

$$m_\varepsilon(T) \xrightarrow[\varepsilon\to 0]{\mathbb{P}} 0$$

Sometimes Large Numbers estimates are not enough, we may also want to estimate the probabilities of very rare events, events which though rare may be of a great significance, think for example of winning the lottery or of a meteor crashing on Earth. The vents are "large deviations" from the norm (as in "expected outcome") and estimates on the probabilities of said events are called large deviations principles. For a thorough theory of large deviations we refer to [16]. In our case we have such an estimate on how the trajectories of the SDE deviates from the trajectories of the mean ODE.

**Theorem 1.2.14** (4.5.2) **:**
$X_t^\varepsilon$ *satisfies a large deviation principle on* $\mathcal{C}([0,1])$

## 1.3 Contents outline

This thesis is organized in three main parts. After this introduction comes an extensive chapter detailing various tools used in the research articles and some partial reviews of subjects that were inspiration and motivation for this work.

First comes a brief introduction to the theory of Markov processes, and more precisely, the two types of Markov processes we will study: discrete time and space Markov chains and diffusion processes. For the more probabilistically inclined reader this is very classical but they are a necessary stepping stone for the rest of the readership. We then give some results about the long time behavior of these type of processes, again old news for the probabilists. We also give some historical results abut the behavior of very general 1-dimensional constrained diffusion, these results are of interest because in Chapter 3 we will work on $d$-dimensional constrained diffusion with more particular coefficients who nonetheless exhibits a similar behavior. After that we give a basic theory of quasi-stationary measure, one of the most important object in latter chapter and a partial review on persistence which is the motivation of this work and knowing the problematics and results of persistence theory gives a new dimension to the problematics tackled herein. Finally we talk about random perturbations of dynamical systems, in the aspects of both stochastic approximation algorithms and Freidlin-Wentzell theory as both these theory are instrumental in the results of the latter chapters. Then Chapter 4 consists of the article "Quasi-stationary distributions for stochastic approximation algorithms with constant step size" in review at the Annals of Applied Probability and Chapter 5 consists of the article "Long time behavior of 1/2 Hölder diffusion population processes" submitted nowhere yet.

# Chapter 2

# Fundamental Tools

When comes the time of writing his/her thesis every PhD student is presented with a dilemma: Try to make it easily accessible to a large readership like master students and take the risk of boring the more knowledgeable reader for who this is very classical or keep it concise and thus obtuse for anyone who isn't familiar with the field. I tried to find the best compromise I could, trying to interest and not to lose the curious non-specialist while attempting not to bore too much someone for who all of this is old news. So, if you are already familiar with the theory of Markov processes I suggest you jump straight to Section 2.3.

The aim of this Chapter is, not only to give results about the tools used in my research, but also to give more context on the field of study of population process via the prisms of both persistence and quasi-stationary distributions. It is not by far exhaustive but merely an appetizer.

In this Chapter I will first give some basic insight about the Markov processes I consider, that is Markov chains and diffusion processes, and their asymptotic and "quasi-asymptotic" behavior. Then I will give results about the two important tools in used in my research: First quasi-stationary distributions, the very objects I studied, and persistence, persistence results were a motivation of my work and, while my results are not per-se persistence results, they go in the same general direction and have a common flavor. Second random perturbations of dynamical systems which are precisely the kind of Markov processes I worked with.

**Contents**

## 2.1 Markov processes

The theory of Markov processes is very rich and already fills a huge number of well-written books, we will only give basic theory of the two kind of processes we will study: Markov chains with discrete time and space and diffusion processes. The curious reader can sate his thirst of knowledge (and proofs) in, e.g., [36], [18], [37] and, for the French speaking reader [22].

### 2.1.1 Markov chains

In this section $E$ will be a countable space and $\mathcal{E} = \mathcal{P}(E)$ will be the $\sigma$-algebra of all the subsets of $E$.

**Definition 2.1.1** (Markov chain)**:**
Let $X = (X_n)_{n\in\mathbb{N}}$ be a process taking values in $E$. We will say that $X$ is a *Markov chain* of initial law $\mu$ if, for all $n \in \mathbb{N}$ and all $(x_0, x_1, \cdots, x_{n+1}) \in E^{n+2}$ the following holds

$$\mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n, \cdots X_1 = x_1, X_0 = x_0] = \mathbb{P}[X_{n+1} = x_{n+1} | X_n = x_n]$$

$$\mathcal{L}(X_0) = \mu$$

Furthermore if, for all $(x, y) \in E^2$, the transition probability from $x$ to $y$ $\mathbb{P}[X_{n+1} = y | X_n = x]$ doesn't depend on $n$ we will say that the Markov chain is homogeneous. An homogeneous Markov is fully characterized by its initial law (that is the law of $X_0$) and the transition matrix $P$ defined by $P(x, y) = \mathbb{P}[X_{n+1} = y | X_n = x]$ (we will still call $P$ a matrix when $E$ is infinite even if it's technically an operator on the functions on $E$).

From now on every Markov chain will be homogeneous.

**Proposition 2.1.2 :**

- *Let $X$ be a Markov chain with transition matrix $P$ and initial law $\mu$. Then, for all $n \in \mathbb{N}$ and all $(x_1, \cdots, x_n) \in E^n$*

$$\mathbb{P}[X_n = x_n, \cdots X_1 = x_1, X_0 = x_0] = \mu(x_0) P(x_0, x_1) P(x_1, x_2) \cdots P(x_{n-1}, x_n)$$

  *Conversely, if that equality holds for for all $n \in \mathbb{N}$ and all $(x_1, \cdots, x_n) \in E^n$, then $X$ is a Markov chain with transition matrix $P$ and initial law $\mu$.*
- *We recursively define the power of $P$ by $P^0 = Id$, $P^1 = P$, and*

$$P^m(x, y) = \sum_{z\in E} P^{m-1}(x, z) P(z, y) = \sum_{z\in E} P(x, z) P^{m-1}(z, y)$$

  *This definition coincides, when $E$ is finite, with the classical matrix product and, when $E$ is infinite, with the composition of operators.*

*Then, for all integers $n, m$ and all $(x, y) \in E^2$ we get*

$$\mathbb{P}[X_{n+m} = y | X_n = x] = P^m(x, y)$$

**Theorem 2.1.3** (Weak Markov property) **:**
*Let $\Omega = E^{\mathbb{N}}$ and $\mathcal{A} = \mathcal{E}^{\otimes \mathbb{N}}$, then, for all $\mathcal{A}$ measurable, positive or bounded function $f$ on $E^{\mathbb{N}}$ and all $n \in \mathbb{N}$,*

$$\mathbb{E}[f(\theta_n(X)) | X_0, \cdots X_n] = \mathbb{E}_{X_n}[f(X)]$$

*where $\theta_n$ is the left $n$-shift operator on $E^{\mathbb{N}}$*

**Definition 2.1.4** (Stopping time)**:**
Let $X_n$ be a homogeneous Markov chain and let $\mathcal{F}_n = \sigma(X_0, \cdots X_n)$ be the $\sigma$-algebra generated by the family $(X_0, \cdots X_n)$. Let $T$ be a random variable taking integer values. We will say that $T$ is a *stopping time* adapted to the chain $X$ if, for all $n \in \mathbb{N}$

$$\{T = n\} \in \mathcal{F}_n$$

We also define the $\sigma$-algebra of the events anterior to $T$ by

$$\mathcal{F}_T = \left\{ A \in \mathcal{A} = \mathcal{E}^{\otimes \mathbb{N}} \middle| \forall n \in N \ A \cap \{T = n\} \in \mathcal{F}_n \right\}$$

Let $A \in \mathcal{E}$, we define $T_A = inf\{n \in \mathbb{N} \, ; \, X_n \in A\}$. Then $T_A$ is a stopping time.

**Theorem 2.1.5** (Strong Markov property) **:**
*Let $\Omega = E^{\mathbb{N}}$ and $\mathcal{A} = \mathcal{E}^{\otimes \mathbb{N}}$, then, for all $\mathcal{A}$ measurable, positive or bounded function $f$ on $E^{\mathbb{N}}$, all $n \in \mathbb{N}$, and all adapted stopping time $T$.*

$$\mathbb{E}[f(\theta_T(X)) \mathbb{1}_{T < +\infty} | \mathcal{F}_T] = \mathbb{1}_{T < +\infty} \mathbb{E}_{X_T}[f(X)]$$

**Proposition 2.1.6** (Recurrence classes) **:**
*A Markov chain induces an oriented graph structure on its state space called the transition graph, in the following way: there is an oriented edge from $x$ to $y$ if and only if $P(x, y) > 0$, that is, if and only $\mathbb{P}[X_1 = y | X_0 = x] > 0$. We will say that $x$ leads to $y$, and write $x \rightsquigarrow y$ if there is a path on the graph from $x$ to $y$ or, equivalently, if there exists $n$ such that $\mathbb{P}[X_n = y | X_0 = x] > 0$. We say that $x$ communicate with $y$ and write $x \leftrightsquigarrow y$ if $x \rightsquigarrow y$ and $y \rightsquigarrow x$.*

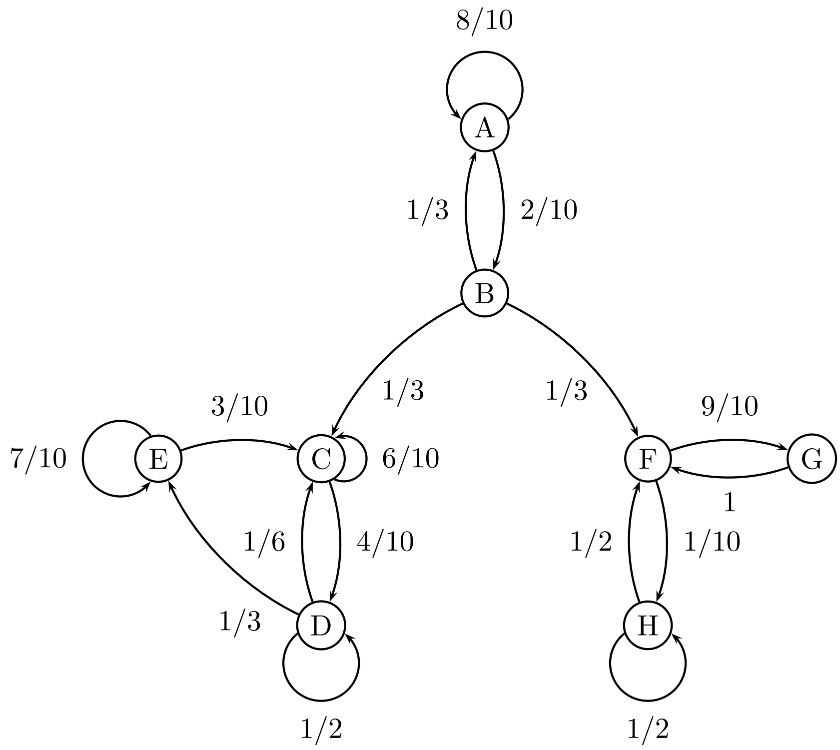
*The relation $x \leftrightsquigarrow y$ is an equivalence relation and allows us to partition $E$ into classes. We define the following sets*

$$R = \{x \in E \mid \mathbb{P}_x[X_n = x \text{ for infinitely many } n] = 1\} \text{ the set of recurrent points}$$

$$T = E \setminus R = \{x \in E \mid \mathbb{P}_x[X_n = x \text{ for infinitely many } n] = 1\} \text{ the set of transient points}$$

*We can then partition $R$ into equivalence classes for the relation $\leftrightsquigarrow$, we call these classes recurrence classes. A chain where $E = R$ is called irreducible.*

An example of transition graph



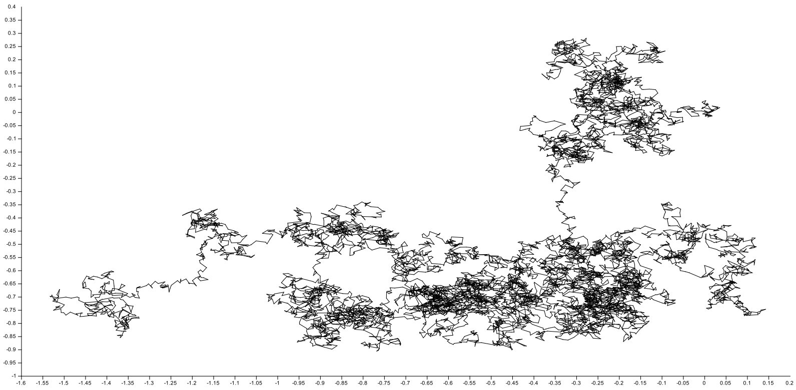

The recurrent states are $C, D, E, F, G$ and $H$, the transient states are $A$ and $B$, the recurrence classes are $\{C, D, E\}$ and $\{F, G, H\}$.

**Proposition 2.1.7 :**
*We have the following properties about recurrence classes, let $C$ be a recurrence class*

- *For all $x \in C$, all $y \notin C$ and all $n \geqslant 0$ we have $P^n(x, y) = 0$*
- *If $x$ is a recurrent point, then $\sum_{n \geqslant 0} P^n(x, x) = \infty$*
- *$x$ is recurrent if and only if $\mathbb{P}_x[T_x < \infty] = 1$*
- *Conversely, if $x$ is transient, then $\sum_{n \geqslant 0} P^n(x, x) < \infty$*

**Definition 2.1.8:**
Let $x \in E$ and let $k = \gcd\{n \in \mathbb{N} \mid P^n(x, x) > 0\}$. If $k = 1$ then $x$ is said to be *aperiodic*, else $x$ will be said said to have period $k$. Every element in a recurrence class has the same period.

Figure 2.1: 2-dimensionnal Brownian motion



### 2.1.2 Diffusion processes

The theory of diffusion process, though very deep and interesting, is long and intricate, what we will give here is merely a rough and partial introduction to it. What we will do won't be very thorough or even very rigorous, for a more complete theory of this topic we refer to [37], [18] and, for the French speaking reader [22].

**Theorem 2.1.9** (Existence of the Brownian motion) **:**
*There exists a probability space on which it is possible to define a process* $(B_t)_{t\geqslant 0}$ *taking values in* $\mathbb{R}^m$ *with the following properties*

*(i)* $B_0(\omega) = 0$ *for all* $\omega$

*(ii) the map* $t \mapsto B_t(\omega)$ *is a continuous function of* $t$ *for all* $\omega$

*(iii) for every* $0 \leqslant s \leqslant t \leqslant 1$, $B_t - B_s$ *is independent of* $\{B_u \ |u \leqslant s\}$ *and has a* $\mathcal{N}(0, (t-s)Id)$ *distribution.*

*Such a process is called a Brownian motion*

The following figure is a representation of a path of 2-dimensional Brownian motion

We won't give a rigorous construction of the Brownian motion but rather a intuition in one dimension.

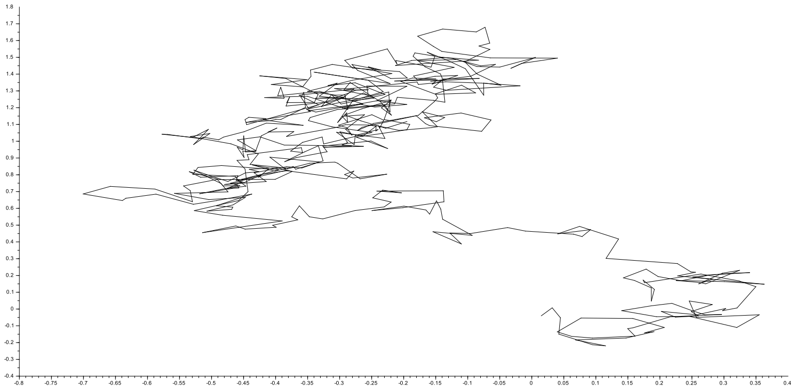
Define the symmetric random walk $X_n$ on $\mathbb{Z}$, that is, for all $x \in \mathbb{Z}$ and all $k \in \mathbb{N}$

$$\mathbb{P}[X^n_{k+1} = x+1 | X^n_k = x] = \frac{1}{2} \qquad \mathbb{P}[X^n_{k+1} = x-1 | X^n_k = x] = \frac{1}{2}$$

Let us now define the process $Z^n_t$ by $Z^n_k = \frac{1}{\sqrt{n}} X^n_{nk}$ if $k \in \frac{1}{n}\mathbb{Z}$ and extending to the real indexes by linear interpolation between points in $\frac{1}{n}\mathbb{Z}$. More precisely

$$Z^n_t = \frac{1}{\sqrt{n}} \left( X^n_{[nt]} + (nt - [nt])(X^n_{[nt]+1} - X^n_{[nt]} \right)$$

Figure 2.2: Trajectory of $X^n$, $n = 100$



Then we got the following theorem

**Theorem 2.1.10** (Donsker) **:**
*The sequence $Z_n$ converges in law to a Brownian motion as n goes to infinity.*

**Theorem 2.1.11** (Properties of the Brownian motion) **:**
*The Brownian motion has the following properties*

(i) *Almost surely, B paths are not Lipschitz continuous anywhere.*

(ii) *Almost surely, B paths have no point of increase ( a local point of increase x for a function $f : [0, ?[\mapsto \mathbb{R}$ is a point where, for some open interval $]a, b[$ containing x we have $f(s) \leqslant f(x)$ for all $s \in ]a, t[$ and $f(x) \leqslant f(s)$ for all $s \in ]x, b[$)*

(iii) *Almost surely, B paths are nowhere differentiable.*

(iv) *Almost surely, B paths are $\gamma$ Hölder continuous for all $\gamma < 1/2$*

(v) *For any $C > 0$, $\frac{1}{\sqrt{C}} B_{Ct}$ is also a Brownian motion.*

(vi) *$tB_{1/t}$ is also a Brownian motion*

(vii) *$-B$ is also a Brownian motion*

**Definition 2.1.12:**
Let $(E, \mathcal{F})$ be a measurable space. A function $P : \mathbb{R}_+ \times E \times \mathbb{R}_+ \times \mathcal{F} \to [0, 1]$ is called a transition probability function if, for all $0 \leqslant s \leqslant t$, all $x \in E$ and all $U \in \mathcal{F}$ it satisfies

(i) $P(s, x, t, .)$ is a probability measure on $\mathcal{F}$.

(ii) $P(s, ., t, U)$ is $\mathcal{F}$-measurable

(iii) For all $r \in [s, t]$

$$P(s, x, t, U) = \int P(r, y, t, U) P(s, x, r, dy)$$

**Definition 2.1.13** (Markov process)**:**
Let $(\Omega, \mathcal{G}, \mathcal{G}_t, \pi)$ be a filtered probability space and let $(E, \mathcal{F})$ be a measurable space. Let $P : \mathbb{R}_+ \times E \times \mathbb{R}_+ \times \mathcal{F} \to [0,1]$ be a transition probability function and let $\mu$ be a probability measure on $\mathcal{F}$. A stochastic process $(X_t)_{t\in\mathbb{R}_+}$ is a Markov process with respect to $(\mathcal{G}_t)$, with transition $P$ and initial distribution $\mu$ if, for all $0 \leqslant s \leqslant t$ and all positive, $\mathcal{F}$-measurable function $f$

$$\mathbb{E}[f(X_0)] = \mu f$$

where $\mu f = \int f(x)\mu(dx)$

$$\mathbb{E}[f(X_t)|\mathcal{G}_s] = \int f(y)P(s, X_s, t, dy) \quad \pi\text{-a.s.}$$

$X$ is said to be an homogeneous Markov process if $P(s, \cdot, t, \cdot)$ only depend on the difference $t-s$, in that case we will write $P(t-s, x, A)$.

**Proposition 2.1.14 :**
*The Brownian motion is an homogeneous Markov process with transition $P$ such that $P(t-s, x, \cdot)$ has a Gaussian law with mean $x$ and standard deviation $\sqrt{t-s}$*

We want to extend the theory of integration to stochastic process, namely we want to give some sense to the expression

$$\int_0^t X_s dB_s$$

where $X_s$ is a stochastic process.

**Definition 2.1.15:**
Let $X_s$ be a left-continuous, adapted and locally bounded process. Let $t_0^n = 0, t_1^n, \cdots, t_n^n = t$ be a family of partitions of $[0,t]$ with mesh going to 0. Then we define

$$\int X_s dB_s = \lim_{n\to\infty} \sum_{i=0}^{n-1} X_{t_{i-1}^n}(B_{t_i^n} - B_{t_{i-1}^n})$$

where the limit is taken for the convergence in probability.

**Proposition 2.1.16 :**
*Let $X_s$ be a left-continuous, adapted and locally bounded process. We have the following properties*

$$\mathbb{E}\left[\int X_s dB_s\right] = 0$$

$$\mathbb{E}\left[\left(\int X_s dB_s\right)^2\right] = \mathbb{E}\left[\int X_s^2 ds\right]$$

We could extend this theory of integration to integral with respect to a semi-martingale, for that we refer to [40] and [41]. Now that we have given a sense to integral of stochastic process with respect to the Brownian motion we want to give a sense to an extension of ordinary differential equations, that is stochastic differential equations. One can see stochastic differential equations as a dynamical systems where the variation of $X_t$ doesn't only depend on a function of $X_t$ but also on random effect. Morally we want to solve things that look like $\dot{x}_t = F(x_t)$ + some sort of small increment of a white noise random process, something like "$\frac{\sigma(x_t)dB_t}{dt}$" which of course we know don't exist.

We thus want to give sense to equations of the form

$$\begin{cases} dX_t = F(t, X_t)dt + \sigma(t, X_t)dB_t \\ X_0 = \xi \end{cases}$$

**Definition 2.1.17:**
We say that $X_t$ is a solution of the stochastic differential equation

$$dX_t = F(t, X_t)dt + \sigma(t, X_t)dB_t$$

if $X_t$ satisfies the stochastic integral equation

$$X_t = X_0 + \int_0^t F(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$$

Such a process is also called a *diffusion process*

When studying the theory of ordinary differential equations there is a very strong tool for proving existence and uniqueness of solutions: the Cauchy-Lipschitz theorem. In the field of stochastic differential equations, the situation gets a little more complicated ; there are different ways to define the existence of a solution and also different ways to define the uniqueness of a solution.

**Definition 2.1.18:**
Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a filtered probability space and let $B$ be a $\mathcal{F}_t$ Brownian motion. We say that $X_t$ is a strong solution of the SDE

$$\begin{cases} dX_t = F(t, X_t)dt + \sigma(t, X_t)dB_t \\ X_0 = \xi \end{cases}$$

if

- $X_0 = \xi$ $\mathbb{P}$-a.s.
- for all $t \geqslant 0$ , $\mathbb{P}$-a.s.

$$\int_0^t \|F(s, X_s)\|ds < \infty \qquad \int_0^t \|\sigma(s, X_s)\|^2 ds < \infty$$

$$X_t = X_0 + \int_0^t F(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s$$

**Definition 2.1.19:**
Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a filtered probability space. We say that the SDE

$$\begin{cases} dX_t = F(t, X_t)dt + \sigma(t, X_t)dB_t \\ X_0 = \xi \end{cases}$$

has the strong uniqueness property if, for all $\mathcal{F}_t$ standard Brownian motion $B$ and all strong solutions $X^1, X^2$ of the SDE with $X_0^1 = X_0^2$, we have

$$\forall t \geqslant 0 \quad X_t^1 = X_t^2 \ \mathbb{P}\text{-a.s.}$$

**Definition 2.1.20:**
We say that the SDE

$$\begin{cases} dX_t = F(t, X_t)dt + \sigma(t, X_t)dB_t \\ X_0 \overset{\mathcal{L}}{=} \mu \end{cases}$$

admits a weak solution $X$ if there exists a filtered probability space$(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ with a $\mathcal{F}_t$ Brownian motion $B$ and there exists a continuous $\mathcal{F}_t$ adapted process $X$ such that

$$\begin{cases} X_t = X_0 + \int_0^t F(s, X_s)ds + \int_0^t \sigma(s, X_s)dB_s \\ X_0 \overset{\mathcal{L}}{=} \mu \end{cases}$$

**Definition 2.1.21:**
We say that the SDE

$$\begin{cases} dX_t = F(t, X_t)dt + \sigma(t, X_t)dB_t \\ X_0 \overset{\mathcal{L}}{=} \mu \end{cases}$$

has the weak uniqueness property if, given two weak solutions $(\Omega^i, \mathcal{F}^i, (\mathcal{F}_t^i)_{t\geqslant 0}, \mathbb{P}^i, B^i, X^i)$ with $X_0^1 \overset{\mathcal{L}}{=} X_0^2$, the two processes $X^1$ and $X^2$ are equal in law

$$\mathcal{L}(X^1) = \mathcal{L}(X^2)$$

**Theorem 2.1.22** (Yamada-Watanabe) **:**
*Strong existence imply weak existence. Strong uniqueness imply weak uniqueness. The converses are false.*

*Weak existence plus strong uniqueness imply strong existence.*

There are lots of results on strong and weak existence and uniqueness, see e.g. [37], [25] and [41]. Mostly the Lipschtiz continuity of the coefficients will be enough to ensure strong existence and uniqueness. When the coefficients fail to be Lipschitz continuous the situation gets more complicated.

**Theorem 2.1.23 :**
*Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a filtered probability space. We suppose that $F$ and $\sigma$ are locally*

*Lipschitz continuous in the second variable, in the sense, for all $n \geqslant 0$, $\exists C_n > 0$ s.t. $\forall t \in [0,n]$, $\forall x, x' \in B(0,n) = \{x \in \mathbb{R}^) \mid \|x\| \leqslant n\}$*

$$\|\sigma(t,x) - \sigma(t,x')\| + \|F(t,x) - F(t,x')\| \leqslant C_n \|x - x'\|$$

*Then the SDE has the strong uniqueness property.*

**Theorem 2.1.24 :**
*Let $(\Omega, \mathcal{F}, (\mathcal{F}_t)_{t\geqslant 0}, \mathbb{P})$ be a filtered probability space and let $B$ be a $\mathcal{F}_t$ Brownian motion. We suppose that $F$ and $\sigma$ are Lipschitz continuous in the second variable, in the sense, for all $n \geqslant 0$, $\exists C_n > 0$ s.t. $\forall t \in [0,n]$, $\forall x, x' \in \mathbb{R}^d$*

$$\|\sigma(t,x) - \sigma(t,x')\| + \|F(t,x) - F(t,xi')\| \leqslant C_n \|x - x'\|$$

$$\mathbb{P}\left[\int_0^n \|\sigma(t,0)\|^2 dt + \int_0^n \|F(t,0)\|^2 dt < +\infty\right] = 1$$

*Then there exist an unique strong solution to the SDE*

When the coefficients $\sigma$ and $F$ don't depend on time we can give weak existence and uniqueness results.

**Theorem 2.1.25** (see [25]) **:**
*We consider the SDE*

$$\begin{cases} dX_t = F(X_t)dt + \sigma(X_t)dB_t \\ X_0 \overset{\mathcal{L}}{=} \mu \end{cases}$$

*If $F$ and $\sigma$ are continuous and bounded and there exists $l > 1$ such that $\int \|x\|^{2l}\mu(dx) < \infty$. Then the SDE admits a weak solution*

**Theorem 2.1.26** (see [41]) **:**
*If one of the following assumptions holds then the SDE*

$$\begin{cases} dX_t = F(X_t)dt + \sigma(X_t)dB_t \\ X_0 \overset{\mathcal{L}}{=} \delta_x \end{cases}$$

*admits a weakly unique weak solution*

- *$F$ and $\sigma$ are $\alpha$-Hölder continuous ($1 > \alpha > 0$) and $\exists \varepsilon$ s.t. $\sigma\sigma^T(x) \geqslant \varepsilon Id_{\mathbb{R}^d}$.*
- *$F$ and $\sigma$ are Lipschitz continuous.*
- *$F$ and $\sigma$ are bounded and locally Lipschitz continuous.*

In the case of ODE we have a very classic result known as the chain rule, that is, for $x_t \in \mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$, if $\dot{x}_t = F(t, x_t)$ and $y = g(t, x_t)$ then

$$\dot{y}_t = \frac{\partial g}{\partial t}(t, x_t) + \langle \nabla g(t, x_t); F(t, x_t)\rangle$$

where $\nabla g(t,x)$ is the vector $\left(\frac{\partial g}{\partial x_1}, \cdots, \frac{\partial g}{\partial x_d}\right)$ In the case of SDE this chain rule doesn't hold anymore, some additional terms appear.

We only give the Itō rule for functions taking values in $\mathbb{R}$. The general rule for functions taking values in $\mathbb{R}^k$ can easily be deduced by applying it to every coordinate function.

**Theorem 2.1.27** (Itō rule) **:**
*Let $X_t$ be a strong solution of the SDE*

$$\begin{cases} dX_t = F(t,X_t)dt + \sigma(t,X_t)dB_t \\ X_0 = \xi \end{cases}$$

*and let $g(t,x)$ be a $\mathcal{C}^2$ map from $\mathbb{R}_+ \times \mathbb{R}^d$ to $\mathbb{R}$. Let $Y_t = g(t,X_t)$. Then $Y_t$ verifies*

$$\begin{cases} Y_t = \int_0^t \frac{\partial g}{\partial t}(s,X_s)ds + \int_0^t \langle \nabla g(s,X_s); F(s,X_s)\rangle ds + \int_0^t \langle \nabla g(s,X_s); \sigma(s,X_s)\rangle dB_s \\ \qquad + \frac{1}{2}\int_0^t Tr(\sigma\sigma^T(s,X_s)\nabla^2 g(s,X_s))ds \\ Y_0 = g(0,\xi) \end{cases}$$

Of course we wouldn't have introduced these diffusion processes if they weren't Markov process.

**Theorem 2.1.28** (see [22]) **:**
*Let $\sigma$ and $F$ be functions from $\mathbb{R}^d$ to $\mathbb{R}^d$ such that, for all $x, x' \in \mathbb{R}^d$*

$$\|\sigma(x) - \sigma(x')\| + \|F(x) - F(x')\| \leqslant C\|x - x'\|$$

*Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and let $X_t$ be a solution of*

$$dX_t = F(X_t)dt + \sigma(X_t)dB_t$$

*Then $X_t$ is a Markov process with respect to $\mathcal{F}_t$ with transition semi-group $P_t$ such that, for all bounded measurable function $f$*

$$P_t f(x) = \mathbb{E}_x[f(X_t)]$$

**Theorem 2.1.29** (see [22]) **:**
*Let $(\Omega, \mathcal{F}, \mathcal{F}_t, \mathbb{P})$ be a filtered probability space and let $X_t$ be a solution of*

$$dX_t = F(X_t)dt + \sigma(X_t)dB_t$$

*Then $X_t$ has the strong Markov property, that is, if $T$ is a stopping time and $\phi$ be a Borel function for $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ to $\mathbb{R}_+$, then*

$$\mathbb{E}\left[\mathbb{1}_{T<\infty}\phi(\{X_{T+t}\ ,\ t \geqslant 0\})|\mathcal{F}_T\right] = \mathbb{1}_{T<\infty}\mathbb{E}_{X_T}[\phi]$$

*where, for all $x \in \mathbb{R}^d$, $\mathbb{P}_x$ is the law on $\mathcal{C}(\mathbb{R}_+, \mathbb{R}^d)$ of a solution of*

$$\begin{cases} dX_t = F(X_t)dt + \sigma(X_t)dB_t \\ X_0 = 0 \end{cases}$$

We have now the most important definitions and tools for the study of Markov processes. What will be of interest to us is what happens to these process as $t$ goes to infinity.

## 2.2 Long time behavior of Markov processes

Let $(X_t)_{t\in\mathbb{T}}$ be a time homogeneous strong Markov process with state space $E$, $\mathbb{T}$ will be either $\mathbb{N}$ or $\mathbb{R}_+$.

**Definition 2.2.1:**
We define the family of operators $(P_t)_{t\geqslant 0}$ on $\mathcal{C}(E,R)$ the set of continuous functions from $E$ to $\mathbb{R}$ by

$$P_t f(x) = \mathbb{E}_x[f(X_t)]$$

The Markov property implies that these operators form a semi-group, that is $P_t P_s = P_{t+s}$. In discrete time we will write $P = P_1$ and $P^n = P_n$.

**Definition 2.2.2** (Invariant measure)**:**
Let $\mu$ be a probability measure on $E$, we say that $\mu$ is *invariant* if, for all $t \in \mathbb{T}$, $\mu P_t = \mu$ (if $E$ is finite $\mu P$ is taken as the product of a line vector and a matrix, if $E$ is infinite $\mu P$ is defined by $\mu P = P^*\mu$ where $P$ is the dual operator of $P$).

An equivalent formulation would be that, for all $f \in \mathcal{C}(E,R)$, $\mathbb{E}_\mu[f(X_t)] = \int f d\mu$.

$\mu$ will be said to be symmetric or reversible if, for all $f$ and $g$ in $\mathcal{C}(E,R)$ and all $t > 0$,

$$\mathbb{E}_\mu[f(X_0)g(X_t)] = \int f P_t g d\mu = \int g P_t f d\mu = \mathbb{E}_\mu[g(X_0)f(X_t)]$$

A symmetric probability measure is trivially invariant. The converse is false: not every invariant measure is reversible.

**Definition 2.2.3:**
An invariant probability measure $\mu$ is said to be ergodic if, for all Borel set

$$P_t \mathbb{1}_B = \mathbb{1}_B \ \mu - a.s. \quad \Longrightarrow \quad \mu(B) \in \{0,1\}$$

That is, the only $\mu$-almost surely invariant sets are $\mu$-trivial.

**Proposition 2.2.4** (see [51]) **:**
*Let $Inv(X)$ be the set of all invariant probability, then $Inv(X)$ is a convex set and its extremal points are ergodic measure. In particular, there always exists an ergodic measure, provided there exists an invariant measure.*

**Proposition 2.2.5 :**
*Let $\mu$ be an invariant measure. $P_t$ is a contraction semi-group. Thus it can be extended to $\mathbb{L}^p(\mu)$ for $1 \leqslant p \leqslant \infty$.*

**Definition 2.2.6:**
Let $(P_t)_{t\geqslant 0}$ be a transition semi-group. Let $\mathcal{C}_0(E)$ be the space of all continuous functions from $E$ to $\mathbb{R}$ that goes to 0 at infinity. The semi-group $(P_t)_{t\geqslant 0}$ is said to have the Feller property if

- For all $t > 0$, $P_t(\mathcal{C}_0(E)) \subset \mathcal{C}_0(E)$
- For all $f \in \mathcal{C}_0(E)$ and all $x \in E$

$$\lim_{t\to 0} P_t f(x) = f(x)$$

If, moreover, for all function $g$ that is measurable and bounded, we have

$$P_t g \in \mathcal{C}_0(E)$$

Then the semi-group will be said to have the strong Feller property.

A Markov process whose semi-group has the Feller property (respectively the strong Feller property) will be called a Feller process (respectively a strongly Feller process).

**Definition 2.2.7:**
Suppose that $\mathbb{T} = \mathbb{R}_+$ Let $\mathcal{C}_0(E)$ be the space of all continuous functions from $E$ to $\mathbb{R}$ that goes to 0 at infinity. Let $\| \,\textbf{.}\, \|$ be the norm on $\mathcal{C}_0(E)$ defined by $\|f\| = \sup\limits_{x\in E} |f(x)|$
Let $\mathcal{D}(L) = \{f \in \mathcal{C}_0(E)\ s.t.\ \lim_{t\to 0} \frac{P_t f - f}{t}$ exists in $\mathcal{C}_0(E)$ for the $\| \,\textbf{.}\, \|$ convergence$\}$ We define $L$, by

$$Lf = \lim_{t\to 0} \frac{P_t f - f}{t} \qquad \text{if } f \in \mathcal{D}(L)$$

We call $\mathcal{D}(L)$ the domain of $L$. Then $\mathcal{D}(L)$ is a sub-vector space of $\mathcal{C}_0(E)$ and $L$ is a linear operator on $\mathcal{D}(L)$ called the infinitesimal generator $L$ also acts on the probability measure on the adjoint sense, that is

$$\begin{aligned} \mu L : \mathcal{D}(L) &\to \mathbb{R} \\ f &\mapsto \int L f d\mu \end{aligned}$$

**Definition 2.2.8:**
Suppose that $\mathbb{T} = \mathbb{N}$. Then we define the infinitesimal generator of the process by $L = P - Id$

**Proposition 2.2.9 :**
*$\mu$ is an invariant measure if and only if $\mu L = 0$*

**Theorem 2.2.10** (Ergodic Theorem) **:**
*Assume that there exists an unique invariant probability measure $\mu$, which is then also ergodic. Then, for $\mu$-almost every $x$ in $E$ and $f \in \mathbb{L}^1(\mu)$*

$$\lim_{t\to\infty} \frac{1}{t} \int_0^t f(X_s) \quad \left( resp.\ \lim_{n\to\infty} \frac{1}{n} \sum_{j=0}^{n-1} f(X_j) \right) \;=\; \int f d\mu$$

$\mathbb{P}_x$ *almost surely and in* $\mathbb{L}^1(\mathbb{P}_x)$*. Also*

$$\lim_{t\to\infty} \frac{1}{t}\int_0^t P_s f \quad \left( \textit{resp.} \ \lim_{n\to\infty} \frac{1}{n}\sum_{j=0}^{n-1} P^j f \right) = \int f d\mu$$

$\mu$ *almost surely and in* $\mathbb{L}^1(\mu)$*.*

In some cases we can even do without the Cesàro mean, that is prove that $\nu P_t \to \mu$ as $t$ goes to infinity, in the weak* topology sense, that is $X_t \xrightarrow{\mathcal{L}} \mu$. The case where $\mathbb{T} = \mathbb{N}$ and $E$ is countable is one of these. It has been studied for a long time and gave birth to numerous results. It also provides great insights on more general cases. Should you be curious about other cases see e.g. Cattiaux lecture notes [11].

#### 2.2.0.1 Markov chains on countable spaces

**Definition 2.2.11:**
Let $\mu$ be a measure such that, for all $(x, y) \in E^2$

$$\mu(x)P(x, y) = \mu(y)P(y, x)$$

Then $\mu$ is said to be reversible.

**Proposition 2.2.12 :**
*Suppose that* $E$ *is finite and that, for some* $x \in E$ *and every* $y \in E$*, the sequences* $(p^n(x, y))_{n\in\mathbb{N}}$ *converge to a limit* $\mu_y$*. Then* $\mu$ *is an invariant distribution.*

**Definition 2.2.13:**
We know that a state $x$ is recurrent if $\mathbb{P}_x[X_n = x \text{ for infinitely many } n] = 1$. If, in addition, the expected return time $\mathbb{E}_x[T_x]$ is finite, then, we say that $x$ is *positive recurrent*. If $x$ is recurrent but not positive recurrent it is called *null recurrent*.

If $E$ is finite then every recurrent state is positive recurrent.

**Theorem 2.2.14 :**
*Let* $C$ *be a recurrence class, then the following assertions are equivalent:*

*(i) Every state in* $C$ *is positive recurrent.*

*(ii) There exists a state in* $C$ *that is positive recurrent*

*(iii)* $P$ *has an invariant distribution* $\mu$ *which positively charges* $C$

**Theorem 2.2.15 :**

1. *If* $X$ *admits only one recurrence class (for example when it is irreducible) then* $X$ *admits at most one invariant probability measure.*
2. *If* $X$ *is irreducible and recurrent, then there exists an unique invariant probability measure* $\mu$*.*

3. *If $X$ has an invariant recurrence class, then there exists at least one invariant probability measure.*
4. *$X$ admits an unique invariant probability measure $\pi$ if and only if there exists exactly one positive recurrent class. In that case $\pi_x = 1/\mathbb{E}_x[T_x]$.*

**Theorem 2.2.16 :**
*Suppose that the chain is irreducible and aperiodic and suppose that it admits an invariant distribution $\pi$. Then, for any starting distribution $\mu_0$ and all $y \in E$,*

$$\lim_{n\to\infty} \mathbb{P}_{\mu_0}[X_n = y] = \pi_y$$

*In particular, for all $(x,y) \in E^2$,*

$$\lim_{n\to\infty} P^n(x,y) = \pi_y$$

*Moreover, if $E$ is finite, then there exists $A > 0$ and $\rho \in ]0,1[$ such that*

$$\|P^n(x,y) - \pi(y)\| \leqslant A\rho^n$$

**Theorem 2.2.17** (Ergodic Theorem / Chacon-Ornstein Theorem ) **:**
*Suppose that the chain $X$ is irreducible, then, for all $x \in E$ and all starting distribution $\mu_0$*

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=0}^{n-1} \mathbb{1}_{X_k=x} \to \frac{1}{\mathbb{E}_x[T_x]}\right] = 1$$

*If the chain is recurrent, then let $\mu$ be an invariant measure and $N_x^y = \sum_{k=0}^{T_y} \mathbb{1}_{X_k=x}$ be the number of passage in $x$ before reaching $y$. Then, for all $(x,y) \in E^2$,*

$$\mathbb{E}_y[N_x^y] = \frac{\mu(x)}{\mu(y)}$$

*For any positive or $\mu$-integrable functions $f, g : I \to \mathbb{R}$, we have*

$$\mathbb{P}\left[\frac{\sum_{k=0}^{n-1} f(X_k)}{\sum_{k=0}^{n-1} g(X_k)} \to \frac{\int f d\mu}{\int g d\mu}\right] = 1$$

*In particular, if $\mu(E) < \infty$ or equivalently if $X$ is positive recurrent, we have, for any bounded function $f$*

$$\mathbb{P}\left[\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \int f d\mu\right] = 1$$

*where $\mu$ is the unique invariant distribution.*

In the finite case, when there exist absorbing points and the chain is irreducible we can easily prove that the chain will ultimately rest in one of those absorbing points. Does this still hold for more general process with absorbing subsets of the state space ? There is no simple answer in a general setting, however when the process evolves in a 1-dimensional state space some results are known.

#### 2.2.0.2 1-dimensional constrained diffusion process

These results were taken from [25] where you can also find the proofs.

Let $I = ]l, r[$ be an open interval in $\mathbb{R}$ ($-\infty \leqslant l < r \leqslant +\infty$ ) Let $\sigma(x)$ and $b(x)$ be real valued functions of class $\mathcal{C}^1$ such that $\sigma^2(x) > 0$ for all $x \in I$. For $x \in I$, the stochastic differential equation

$$\begin{cases} dX_t = b(X_t)dt + \sigma(X_t)dB_t \\ X_0 = x \end{cases}$$

has a unique solution $X_t^x$ up to the explosion time $\tau = \lim_{n\to\infty} \tau_n$ where

$$\tau_n = \inf\{t\ ;\ X_t^x \notin [l + 1/n, r - 1/n]\}$$

It can be shown that $\lim_{t\to\tau} X_t$ exists and is equal to $l$ or $r$ a.s. on the set $\{\tau < \infty\}$. We define $X_t^x$ to be this limit for $t \geqslant \tau$. Let $W_I$ be the set of continuous paths $w : [0, \infty[ \to [l, r]$ such that $w(0) \in I$ and $w(t) = w(\tau(w))$ for $t \geqslant \tau(w)$ where $\tau(w)$ is the explosion time of the path $\tau(w) = \inf\{t\ ;\ w(t) \in \{l, r\}\}$. $X^x$ is then a $W_I$ valued random variable with $\tau = \tau(X^x)$. Let $\mathbb{P}_x$ be the probability law of $X^x$ on $W_I$. $\{P_x\}$ defines the minimal $L$-diffusion process on $I$ where

$$L = \frac{1}{2}\sigma^2(x)\frac{d}{dx^2} + b(x)\frac{d}{dx}$$

Let $c \in I$ be fixed and define

$$s(x) = \int_c^x exp\left(-\int_c^y \frac{2b(z)}{\sigma^2(z)}dz\right) dy$$

$s(x)$ is then a strictly increasing smooth function on $I$ satisfying

$$Ls(x) = 0 \text{ on } I$$

As a consequence $s(x_t)$ is a local martingale, from that fact stems the following Theorem.

**Theorem 2.2.18 :**
*For a function $f$ from $I$ to $\mathbb{R}$ we will denote*

$$f(l^+) = \lim_{a\to l\ a>l} f(a) \text{ and } f(r^-) = \lim_{b\to r\ r>b} f(b)$$

*1/ If $s(l^+) = -\infty$ and $s(r^-) = \infty$ then, for every $x \in I$*

$$\mathbb{P}_x[\tau = \infty] = \mathbb{P}_x[\limsup_{t\to\infty} X_t = r] = \mathbb{P}_x[\liminf_{t\to\infty} X_t = l] = 1$$

*In particular the process is recurrent, i.e. $\mathbb{P}_x[\exists t \text{ s.t. } X_t = y] = 1$ for all $x$ and $y$ in $I$.*

*2/ If* $s(l^+) > -\infty$ *and* $s(r^-) = \infty$ *then* $\lim_{t\to\tau} X_t$ *exists* $\mathbb{P}_x$ *a.s. and*

$$\mathbb{P}_x[\lim_{t\to\tau} X_t = l] = \mathbb{P}_x[\sup_{t<\tau} X_t < r] = 1$$

*We get a similar result when reversing the roles of* $l$ *and* $r$

*3/ If* $s(l^+) > -\infty$ *and* $s(r^-) < \infty$ *then* $\lim_{t\to\tau} X_t$ *exists* $\mathbb{P}_x$ *a.s. and*

$$\mathbb{P}_x[\lim_{t\to\tau} X_t = l] = 1 - \mathbb{P}_x[\lim_{t\to\tau} X_t = r] = \frac{s(r^-) - s(x)}{s(r^-) - s(l^+)}$$

*In the last two cases the process is not recurrent.*

To get more information on the process we would like to use the function $u$ solution of $Lu = u$. Indeed, from the Itō rule, we would then have $e^{-t}u(X_t)$ as a local martingale which would again give us great insights. However there is no simple expression of the solution of $Lu = u$. Instead, for $x \in I$, we define

$$\kappa(x) = 2\int_c^x exp\left(-\int_c^\xi \frac{2b(z)}{\sigma^2(z)}dz\right)\left(\int_c^\xi exp\left(\int_c^\eta \frac{2b(z)}{\sigma^2(z)dz}\right)\frac{d\eta}{\sigma^2(\eta)}\right)d\xi$$

**Lemma 2.2.19 :**
*Let* $u$ *be the unique solution of*

$$\begin{cases} Lu(x) = u(x) \\ u(c) = 1 \\ u'(c) = 0 \end{cases}$$

*Then*

$$1 + \kappa(x) \leqslant u(x) \leqslant exp\,(\kappa(x))$$

We can remark that

$$\kappa(r^-) < \infty \Rightarrow s(r^-) < \infty$$

and

$$\kappa(l^+) > -\infty \Rightarrow s(l^+) > -\infty$$

**Theorem 2.2.20 :**

*1/ If* $\kappa(r^-) = \kappa(l^+) = \infty$*, then, for all* $x \in I$

$$\mathbb{P}_x[\tau = \infty] = 1$$

*2/ If* $\kappa(r^-) < \infty$ *or* $\kappa(l^+) < \infty$*, then, for all* $x \in I$

$$\mathbb{P}_x[\tau < \infty] > 0$$

*3/*

$$\mathbb{P}_x[\tau < \infty] = 1 \textit{ for all } x \in I$$

*if and only if one of the following cases occurs*

*(i)* $\kappa(r^-) < \infty$ *and* $\kappa(l^+) < \infty$

*(ii)* $\kappa(r^-) < \infty$ *and* $s(l^+) = -\infty$

*(iii)* $\kappa(l^+) < \infty$ *and* $s(r^-) = \infty$

## 2.3 Extinction and survival of population processes

### 2.3.1 Persistence

The goal of the theory of persistence is to answer a simple question (though the answer isn't simple) about mixed population dynamics: "Under which conditions do the populations coexist ?" In this section, we will first talk about the treatment of this problematic in the context of deterministic population process and then move on to the case of stochastic population processes. For additional references you can see [45], [29].

Our first setting will be that of an ODE $\dot{x}(t) = F(t, x(t))$ on either $(\mathbb{R}_+)^d$ or $\Delta$ the simplex of $\mathbb{R}^d$. We will denote by $\varphi_t(x)$ the solution at time $t$ of the ODE with initial condition $x(0) = 0$. We should first ask ourselves what do we mean by "coexistence of populations" or "survival" ? In most of the cases the uniqueness of solutions for our ODE will imply that our dynamical system never hits the extinction states, e.g. in the classical replicator dynamic

$$\dot{X} = X \circ \Big( A(X)X - X^T A(X)X \Big)$$

where $X$ is a vector in $(\mathbb{R}_+)^d$ and $A(X)$ a $d \times d$ matrix, the Cauchy-Lipschitz Theorem implies that $X(t) \in (\mathbb{R}_+^*)^d$ for all $t > 0$, provided that $X(0) \in (\mathbb{R}_+^*)^d$. Even more simple is the one-dimensional dynamic $\dot{x} = -x$. It is obvious that this kind of population should fall within the category of "extincting" or "non-persistent" populations, however we still have $x(t) > 0$ for all $t$ if $x(0) > 0$. Thus we want to define deterministic persistence as "doesn't go too close to extinction", more precisely we have the following definitions

**Definition 2.3.1** (see [29])**:**
Let $X$ be our state space and let $\rho$ be a function from $X$ to $\mathbb{R}_+$ such that $\{x \mid \rho(x) = 0\}$ is the extinction set we want to "avoid". Most of the time we will have $X = (\mathbb{R}_+)^d$ or $X = \Delta$ and $\rho(x) = d(x, \partial X)$. A semi-flow $\varphi$ is

- weakly $\rho$-persistent if, for all $x \in \{x \in X \mid \rho(x) > 0\}$

$$\limsup_{t\to\infty} \rho(\varphi_t(x)) > 0$$

- strongly $\rho$-persistent if, for all $x \in \{x \in X \mid \rho(x) > 0\}$

$$\liminf_{t\to\infty} \rho(\varphi_t(x)) > 0$$

– uniformly weakly $\rho$-persistent if there exists $\varepsilon > 0$ such that, for all $x \in \{x \in X \mid \rho(x) > 0\}$

$$\limsup_{t\to\infty} \rho(\varphi_t(x)) > \varepsilon$$

– uniformly strongly $\rho$-persistent also called uniformly $\rho$-persistent if there exists $\varepsilon > 0$ such that, for all $x \in \{x \in X \mid \rho(x) > 0\}$

$$\liminf_{t\to\infty} \rho(\varphi_t(x)) > \varepsilon$$

These definitions are very general in order to cover all cases, even the more pathological ones. Most of the time we will use this definition:

**Definition 2.3.2:**
Let $X = (\mathbb{R}_+)^d$ or $X = \Delta$. A semi-flow $\varphi$ is persistent if there exists $\varepsilon$ such that, for all $x \in \mathring{X}$

$$\forall i \in \{1, \cdots d\} \quad \liminf_{t\to\infty} (\varphi_t(x))_i > \varepsilon$$

There is a plethora of results about these different kind of persistence in various models, we refer to [29], we will however give a general one which will help enlighten the results on stochastic persistence.

**Theorem 2.3.3 :**
*Suppose that there exists a global compact attractor $A \subset \mathring{X}$ for the flow $\varphi_t$, then $\varphi_t$ is persistent.*

If we now want to extend our study to that of dynamical systems with noise these definitions may not work, For example, with a stochastic process $X_t$ on $(\mathbb{R}_+)^d$ with Brownian noise there is no chance that we could have something like

$$\forall i \in \{1, \cdots d\} \quad \liminf_{t\to\infty} (X_t)_i > \varepsilon \text{ almost surely}$$

We thus need new definitions for persistence for stochastic processes. Our setting will first be that of discrete time Markov chain in a compact state space, then we will move on to diffusion process, again on a compact state space.

Firstly, our discrete time setting is that of a Markov population process defined as such

$$X_{t+1} = F(X_t, \xi_{t+1})$$

where $X_t \in K$ is the state of the population at time $t$ (e.g. a vector of densities or frequencies), $K$ is the compact state space and $\xi_t$ is a random variable that determines the environmental conditions at time $t$. We will make the following hypotheses

**Hypothesis 2.3.4 :**

*(i)* $\{\xi_t\}_{t=0}^{\infty}$ *are of i.i.d random variables taking values in a separable metric space* $\Omega$ *(such as* $\mathbb{R}^d$*).*

*(ii)* $F : \mathbf{S} \times \Omega \to K$ *is a continuous function.*

*(iii) There is a closed absorbing subset* $K_0 \subset K$

Most, if not all, of the times, $K$ will be either the $d$-dimensional simplex $\Delta$ in which case $K_0 = \partial\Delta$ or a compact subset of $(\mathbb{R}_+)^d$ in which case $K_0 = \{x \in K \mid \exists i \ x_i = 0\}$. In either case our assumption mean that there is no immigration or spontaneous generation, when a species goes extinct it stays that way. We could relax the boundedness assumption by just asking $X_t$ to be stochastically bounded, see e.g. [45], nevertheless our goal here will be to study stochastic process on the simplex $\Delta$, in which case will have almost sure boundedness, we thus focus on almost surely bounded processes.

As we saw earlier,the aim of persistence theory is to study survival of the population. Following the same direction as the deterministic case we might want to study the probability that, in the long term, our process stays away from the extinction states, that is the asymptotic behavior of $\mathbb{P}[X_t \in U]$ where $U$ is a subset of $K \setminus K_0$. As always with stochastic process, our process can and will make excursions close to the boundary, in that case we might also wonder "how long do these excursions take ?" and "how often do these excursions happen ?". To answer these questions we will study the asymptotic behavior of the occupation probability

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

where $\delta_x$ denotes a Dirac measure at the point $x$ i.e. $\delta_x(A) = 1$ if $x \in A$ and 0 otherwise. For any Borel set $A \subset \mathbf{S}$, $\Pi_t(A)$ is the fraction of time that $X_s$ spends in $A$ for $1 \leqslant s \leqslant t$. The long-term fraction of time that $X$ spends in $A$ is given by $\lim_{t\to\infty} \Pi_t(A)$, provided the limit exists. Understanding this asymptotic behavior with probability one corresponds to the "typical trajectory" perspective.

In deterministic models we defined persistence by "staying bounded away from the border". Similarly we could define stochastic persistence by "staying almost surely bounded away from the extinction state". However we will often have model with noise that are absolutely continuous with respect to the Lebesgue measure, e.g perturbed replicator dynamic $X_{t+1} = X_t(A(X_t) + \xi_t)$ where $\xi_t$ has a normal distribution. In these models the process will never almost surely stay away from the absorption states. We need to relax our definitions. For our new definitions, we introduce the set of the population states within $\eta > 0$ of extinction

$$K_\eta = \{x \in K : d(x, K_0) \leqslant \eta\}.$$

We then have the following definitions

**Definition 2.3.5** (see [45])**:**
The Markov chain $X_t$ is persistent in probability if, for all $\varepsilon > 0$, there exists $\eta > 0$ such

that

$$\limsup_{t\to\infty} \mathbb{P}[X_t \in K_\eta | X_0 = x] \leqslant \varepsilon$$

for all $x \in K \setminus K_0$.

This definition means that reaching low densities or frequencies is very unlikely in the long term. The next definition provides the "typical trajectory" perspective on persistence.

**Definition 2.3.6** (see [45])**:**
The Markov chain $X_t$ is almost surely persistent if, for all $\varepsilon > 0$, there exists $\eta > 0$ such that

$$\limsup_{t\to\infty} \Pi_t[K_\eta] \leqslant \varepsilon \text{ almost surely}$$

whenever $X_0 = x \in K \setminus K_0$.

This definition means that the fraction of time a typical population trajectory spends near extinction states is very small.

We have the following proposition

**Proposition 2.3.7** (see [45]) **:**
*If $(X_t)_{t\geqslant 0}$ is persistent in probability, then the set of weak* limit points of*

$$\frac{1}{t}\sum_{s=0}^{t-1} \mathbb{P}[X_s \in \cdot | X_0 = x]$$

*with $x \in K \setminus K_0$ is non-empty and each of these limit points is an invariant measure.*

*Alternatively, if $X_t$ is almost surely persistent, then the set of weak* limit points of $\Pi_t$ with $X_0 = x \in K \setminus K_0$ is almost-surely non-empty and each of these limit points is almost-surely a positive invariant measure.*

When a unique positive invariant probability measure exists and the system is persistent, one can often show that, if $X_0 = x \in K \setminus K_0$, then the distribution of $X_t$ converges to $\mu$ and $\Pi_t$ converges almost surely to $\mu$. For that we can make use of the following theorem of [33]. This theorem relies on the concept of $\varphi$-irreducibility with respect to a Borel set $B \subset K$. We say that a Markov chain $X_t$ is $\varphi$-irreducibility on $B$ if there exists a Borel measure $\varphi$ on $B$ such that $\varphi(A) > 0$ implies that $\mathbb{P}[X_t \in A \text{ for some } t | X_0 = x] > 0$ for all $x \in B$.

**Theorem 2.3.8** (see [33]) **:**
*Assume that $X_t$ is $\varphi$-irreducible on $K \setminus K_0$ and that there exists a positive function $V : K \setminus K_0 \to \mathbb{R}_+$, a compact set $C \subset K \setminus K_0$, and a constant $\beta > 0$ such that*

$$\mathbb{E}[V(X_1)|X_0 = x] \leqslant (1-\beta)V(x) + \mathbb{1}_C(x) \textit{ for all } x \in K \setminus K_0$$

*Then there exists a unique invariant probability measure $\mu$ and the distribution of $X_t$ converges in the weak* topology to $\mu$. Moreover, $\Pi_t$ almost surely converges in the weak* topology to $\mu$ whenever $X_0 = x \in K \setminus K_0$.*

A drawback of requiring $\varphi$-irreducibility is that it can be difficult to verify or even false for important models. For instance, models where there are a finite number of environmental states (i.e. $\Omega$ is a finite set) rarely satisfy the irreducibility condition and, consequently, may not have a unique positive invariant measure. However, finiteness of the state space allows us to easily make us of Perron-Frobenius-like theorems which might give us the unicity of an invariant probability.

We now give more specific results obtained by Benaïm, Hofbauer and Sandholm in [5], Schreiber, Benaïm and Atchadé in [46] and Benaïm in a work in progress. They studied both discrete time and continuous time models of multiple species interactions (replicator dynamics in [5], discrete time Markov process and diffusion process [46] and partially deterministic Markov process in Benaïm work in progress) and gave criteria for persistence and robust persistence, that is ensuring that persistence also holds for small perturbations of the model.

Their discrete time model is as follows: Let $X_t^i$ denote the density of the $i$-th population at time $t$ and $X_t = (X_t^1, \dots, X_t^k)$ the vector of population densities at time $t$. The fitness $f_i(X_t, \xi_{t+1})$ of population $i$ at time $t$ depends on the population state and on a random variable $\xi_{t+1}$ that represents the state of the environment at time $t+1$. The process $X_t$ follows then the equation

$$X_{t+1} = f(X_t, \xi_{t+1}) \circ X_t$$

where $f(x, \xi) = (f_1(x, \xi), \dots, f_k(x, \xi))$ is the vector of fitnesses and $\circ$ denotes the Hadamard product i.e. component-wise multiplication.

The same assumptions as before are made on the model

**Hypothesis 2.3.9 :**

*(i)* *There exists a compact set $K$ of $\mathbb{R}_+^k = \{x \in \mathbb{R}^k \mid x_i \geqslant 0\}$ such that $X_t \in K$ for all $t \geqslant 0$.*

*(ii)* *$\{\xi_t\}_{t=0}^{\infty}$ is a sequence of i.i.d random variables independent of $X_0$ of law $m$ taking values in a Polish space $E$*

*(iii)* *$f_i(x, \xi)$ are positive functions, continuous in $x$ and measurable in $\xi$.*

*(iv)* *For all $i$, $\sup_{x \in K} \int (\log f_i(x, \xi))^2 \, m(d\xi) < \infty$*

Assumption $(i)$ ensures that the populations remain bounded for all time. Assumptions $(ii)$ and $(iii)$ imply that $\{X_t\}_{t=0}^{\infty}$ is a Markov chain on $K$ and that $\{X_t\}_{t=0}^{\infty}$ is a Feller process. Assumption $(iv)$ is a technical assumption met by many models.

The persistence results are expressed in term of the invasion rates of species $i$ with respect to a measure $\mu$ defined by

**Definition 2.3.10:**
The expected per-capita growth rate at state $x$ of population $i$ is

$$\lambda_i(x) = \int \log f_i(x,\xi) m(d\xi).$$

When $\lambda_i(x) > 0$, the $i$-th population tends to increase when the current population state is $x$. When $\lambda_i(x) < 0$, the $i$-th population tends to decrease when the current population state is $x$. For an invariant probability measure $\mu$, we define the invasion rate of species $i$ with respect to $\mu$ to be

$$\lambda_i(\mu) = \int_K \lambda_i(x)\, \mu(dx)$$

The following proposition clarifies the significance of the invasion rates $\lambda_i(\mu)$

**Proposition 2.3.11 :**
*Let $\mu$ be an invariant probability measure and let $i \in \{1, \cdots, k\}$.. Then, there exists a bounded map $\hat{\lambda}_i : K \to \mathbb{R}$ such that:*

- *For $\mu$-almost every $x \in K$*

$$\mathbb{P}\left[\lim_{t\to\infty} \frac{1}{t} \sum_{s=0}^{t-1} \log f_i(X_s, \xi_{s+1}) = \hat{\lambda}_i(x) \mid X_0 = x\right] = 1$$

- *We have*

$$\int_K \hat{\lambda}_i(x)\mu(dx) = \lambda_i(\mu)$$

  *Furthermore, if $\mu$ is ergodic, then*

$$\hat{\lambda}_i(x) = \lambda_i(\mu) \quad \mu\text{-almost surely}$$

- *If $\mu(\{x \in K : x_i > 0\}) = 1$, then $\lambda_i(\mu) = 0$*

Then Benaïm and al. proved the following criteria for persistence.

**Theorem 2.3.12 :**
*Assume that one of the following equivalent conditions hold:*

*(i) For all invariant probability measures $\mu$ supported on $K_0$,*

$$\lambda_*(\mu) := \max_i \lambda_i(\mu) > 0$$

*(ii) There exists $p \in \Delta$ such that*

$$\sum_i p_i \lambda_i(\mu) > 0$$

*for all ergodic probability measures $\mu$ supported by $K_0$.*

*Then for all $\varepsilon > 0$ there exists $\eta > 0$ such that*

$$\limsup_{t\to\infty} \Pi_t(K_\eta) \leqslant \varepsilon \textit{ almost surely}$$

*whenever $X_0 = x \in K \setminus K_0$.*

*If we additionally assume that $(X_t)$ is irreducible over $K \setminus K_\eta$ for all $\eta > 0$, then there exists a unique invariant probability measure $\pi$ such that $\pi(K_0) = 0$ and the occupation measures $\Pi_t$ converge almost surely to $\pi$ as $t \to \infty$, whenever $X_0 = x \in K \setminus K_0$.*

*In addition, if $(X_t)$ is strongly irreducible over $K \setminus K_\eta$ for all $\eta > 0$. Then the distribution of $X_t$ converges to $\pi$ as $t \to \infty$ whenever $X_0 = x \in K \setminus K_0$; that is*

$$\lim_{t\to\infty} \|\mathbb{P}_x[X_t \in \centerdot] - \pi\|_{TV} = 0 \textit{ for all } x \in K \setminus K_0.$$

*where* $\|\mu - \nu\|_{TV} = \sup_{A \textit{ Borel set}} |\mu(A) - \nu(A)|$

In a current work in progress, Benaïm gave, under some additional hypotheses, a rate of convergence of $\mathbb{P}_x[X_t \in \centerdot]$ to $\pi$.

**Theorem 2.3.13 :**
*Under some additional hypotheses, there exists positive constants $C$, $\theta$ and $\lambda$ such that, for every Borel set $A$ and every $x \in K \setminus K_0$.*

$$\|\mathbb{P}_x[X_t \in A] - \pi(A)\|_{TV} \leqslant C\left(1 + \frac{1}{d(x, K_0)^\theta}\right) e^{-\lambda t}$$

This approach of persistence via invasion rates can be extended to continuous time models. In [46] Schreiber et al. studied the case of a stochastic differential equation on the $d$-dimensional simplex. Namely

$$dX_t^i = X_t^i[F_i(X_t)dt + \sum_{j=1}^m \Sigma_i^j(X_t)dB_t^j], \qquad i = 1, \dots, k.$$

with Lipschitz assumptions on $\Sigma$ and $F$ to ensure strong existence and strong uniqueness and assumptions that both the drift and diffusion terms are in the tangent space $T\Delta$ to ensure $X_t \in \Delta$ a.s.

**Definition 2.3.14:**
The analog of the per-capita growth rate for these continuous time processes is given by

$$\lambda_i(x) = F_i(x) - \frac{1}{2}a_{ii}(x)$$

where

$$a_{ij}(x) = \sum_{k=1}^m \Sigma_i^k(x)\Sigma_j^k(x).$$

When $\lambda_i(x) > 0$, the population tends to increase. When $\lambda_i(x) < 0$, the population tends to decrease. Like in the discrete-time case, we define

$$\lambda_i(\mu) = \int \lambda_i(x)\mu(dx)$$

and

$$\lambda_*(\mu) = \max_i \lambda_i(\mu)$$

Again we have a theorem linking the invasion rates with persistence

**Theorem 2.3.15 :**
*Assume that one of the following equivalent conditions hold:*

*(i) For all invariant probability measures $\mu$ supported on $K_0$,*

$$\lambda_*(\mu) := \max_i \lambda_i(\mu) > 0$$

*(ii) There exists $p \in \Delta$ such that*

$$\sum_i p_i \lambda_i(\mu) > 0$$

*for all ergodic probability measures $\mu$ supported by $K_0$.*

*Then, for all $\varepsilon > 0$ there exists $\eta > 0$ such that, whenever $X_0 = x \in K \setminus K_0$*

$$\limsup_{t\to\infty} \Pi_t(K_\eta) \leqslant \varepsilon \text{ almost surely}$$

*where*

$$\Pi_t = \frac{1}{t}\int_0^t \delta_{X_s} ds$$

*Additionally if we assume that the SDE is non-degenerate, i.e. the column vectors $(x \circ \Sigma)_1(x), \cdots, (x \circ \Sigma)_d(x))$ span $T\Delta$ for all $x \in \mathring{\Delta}$, then, there exists a unique invariant probability $\pi$ such that $\pi(\partial\Delta) = 0$. and*

*(i) The distribution of $X_t$ converges to $\pi$ as $t \to \infty$ whenever $X_0 = x \in \mathring{\Delta}$ that is*

$$\lim_{t\to\infty} \|\mathbb{P}_x[X_t \in \boldsymbol{.}] - \pi\|_{TV} = 0 \text{ for all } x \in \mathring{\Delta},$$

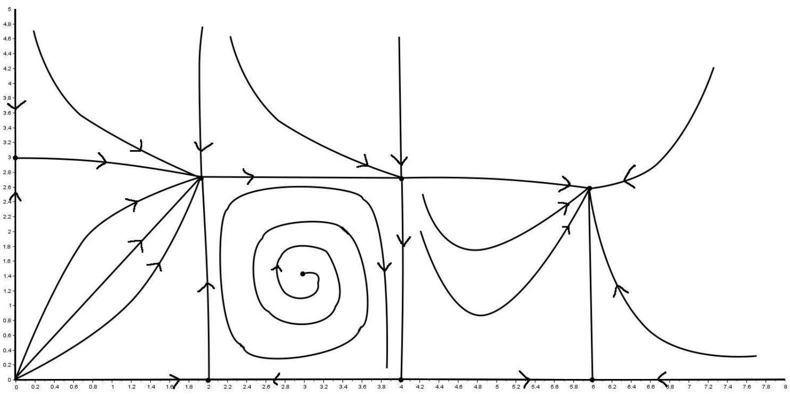
*(ii) The occupation measures $\Pi_t = \frac{1}{t}\int_0^t \delta_{X_s} ds$ converge almost surely to $\pi$, whenever $X_0 = x \in \mathring{\Delta}$.*

These results were expressed for a SDE on $\Delta$ but they also hold for dynamics on a more broad class of state space, e.g. dynamics on $(\mathbb{R}_+)^d$ that are stochastically bounded. A particularly interesting class of examples is randomly perturbed ODE, that is, dynamics of the form

$$dX_t = X_t \circ (F(X_t)dt + \varepsilon dB_t)$$

These systems can lead to surprising behavior, one might for example expect the added "chaos" to make survival harder but this intuition isn't true. The criteria using invasion rates teach us that, to prove stochastic persistence, we don't have to look at the whole

Figure 2.3: Phase portrait of a non-permanent system whose small random perturbations are persistent



dynamics near the border but only at what happens near the support of the ergodic measures

The following phase portrait gives us a very nice example where the deterministic dynamic $\dot{x} = x \circ F(x)$ isn't persistent due to the heteroclinic cycle that stays close to the border for increasingly long time. However the random perturbed dynamic $dX_t = X_t \circ (F(X_t)dt + \varepsilon dB_t)$ might be persistent as we only have to "look at what happens near the border equilibrium". Morally the introduction of the Gaussian noise helps the process escape from the heteroclinic cycle and go into parts of the system that converges to interior attractors.

In the real world, some populations aren't really "persistent" in the sense that, in the very long term they will get extinct due to extreme random effects. However this extinction can happen in so long time that, in our time scale, the population seems to settle in an equilibrium. As the population ultimately goes extinct this "equilibrium" can't be stable, it is "metastable". The theory of quasi-stationary distributions gives a mathematical sense to these behavior of "metastable" equilibria before extinction.

### 2.3.2 Quasi-stationary distributions

Again we won't give the proofs of the results herein, you can find them for example in S. Meleard and D. Villemonais review [32] or in P. Collet, S. Martinez and J. San Martín book [12]. For other references on the subject see Pollett bibliography [39].

Here we will consider a Markov chain $Z_t$ which represents population size, $t$ is either a discrete or a continuous time, Our process state space $E \subset \mathbb{R}^d$ admits an absorbing state, denoted by $\{0\}$. We will denote $E^* = E \setminus \{0\}$, $\mathcal{P}^*$ the set of probability measures whose support lies in $E^*$. We define $T_0$ to be the absorption time.

$$T_0 = \inf\{t > 0; Z_t = 0\}$$

We suppose that, whatever the initial state is, the process will almost surely be absorbed, i.e.

$$\forall z \in E \qquad \mathbb{P}_z[T_0 < \infty] = 1$$

For any probability measure $\mu$ on $E^*$, we denote by $\mathbb{P}_\mu$ (resp. $\mathbb{E}_\mu$) the probability (resp. the expectation) associated with the process $Z$ initially distributed with respect to $\mu$. For any $x \in E^*$, we set $\mathbb{P}_x = \mathbb{P}_{\delta_x}$. We denote by $(P_t)_{t\geq 0}$ the semi-group of the process $Z$ killed at 0. That is, for any $z > 0$ and $f$ measurable and bounded on $E^*$, we define

$$P_t f(z) = \mathbb{E}_z(f(Z_t)\mathbb{1}_{t<T_0})$$

For any finite measure $\mu$ and any bounded measurable function $f$, we set

$$\mu(f) = \int_{E^*} f(x)\mu(dx),$$

and we also define the finite measure $\mu P_t$ by

$$\mu P_t(f) = \mu(P_t f) = \mathbb{E}_\mu(f(Z_t)\mathbb{1}_{t<T_0}).$$

Here the notion of persistence has little relevance as the process is ultimately doomed to extinction. However we can still study its long time behavior and the "survival" of our process is the sense of "What is the law of a non-extinct population at a large time $t$ ?" i.e what is the long term behavior of distribution of $Z_t$ conditioned on survival at time $t$ $\mathbb{P}_\mu[Z_t \in \cdot | T_0 > t]$, where $\mu$ is the initial distribution of the population's size $Z_0$. We can also be interested on the time at which that extinction take place, that is "What is the law of $T_0$ ?".

This type of problematic appears in a lot of population processes, we will illustrate with a classical and historically important example: the Galton-Watson process. It was the study of this process that led Yaglom to the foundation of the theory of QSD, which is the subject of this section.

We recall the setting: The Galton-Watson process is a population dynamic that was introduced by Galton and Watson (see [52]) in order to study the extinction of aristocratic family names. The population at generation $n+1$ consists of the male children (we consider only the male children because only they carry the family name) of generation $n$ males, each individual of generation $n$ having children independently of others individuals. We get a model in discrete time, whose size $(Z_n)_{n\geq 0}$ evolves according to the recurrence formula

$$Z_{n+1} = \sum_{i=1}^{Z_n} \xi_i^{(n)}$$

where $(\xi_i^{(n)})_{i,n}$ is a family of independent random variables, identically distributed following the probability measure $\mu$ on $\mathbb{N}$ with generating function $g$. As defined, $Z_n$ is the size of the $n^{th}$ generation of a population where each individual has a random number of children, chosen following $\mu$ and independently of the rest of the population.

We will assume that $0 < \mu(\{0\}) + \mu(\{1\}) < 1$, i.e. there is a non-zero probability that a noble will have more than one son. When this assumption is not true the process is of almost no interest, it is very easy to show that $Z_n$ will be decreasing with $n$ and either goes almost surely to 0 or stays almost surely constant.

We denote by $m = E(\xi_1^{(0)})$ the average number of sons by individual in our Galton-Watson process. As each individual has offspring independently of the others individuals, we get that, starting from $Z_0$, the process $Z$ is equal to the sum of $Z_0$ independent Galton-Watson processes issued from a single individual. By this branching property, the probability of extinction for the population starting from one individual is obtained as follows:

$$\mathbb{P}_1(\exists n \in \mathbb{N},\ Z_n = 0) = \lim_{n\to+\infty} \mathbb{E}_1(0^{Z_n}) = \lim_{n\to\infty} g \circ \cdots \circ g(0)\ (n \text{ times}).$$

This model is now a very classical example and motivation of the study of branching processes and it is taught in masters course around the globe, the following results can be found in lots of books, e.g. [4] Three different behaviors can occur

**Theorem 2.3.16 :**

- *The sub-critical case $m < 1$: the process becomes extinct in finite time almost surely and the average extinction time $\mathbb{E}(T_0)$ is finite.*
- *The critical case $m = 1$: the process becomes extinct in finite time almost surely, but $\mathbb{E}(T_0) = +\infty$.*
- *The super-critical case $m > 1$: the process is never extinct with a positive probability, i.e. $\mathbb{P}[T_0] = \infty] > 0$ and it yields immediately that $\mathbb{E}(T_0) = +\infty$.*

When viewed with the prism of persistence, the super-critical case is the less interesting, we know that, on the event $\{T_0 = \infty\}$, $Z_n$ goes exponentially fast to infinity

**Proposition 2.3.17 :**
*Suppose that the Galton-Watson process is super-critical, then there exists a random variable $M$ such that*

$$\lim_{n\to\infty} \frac{Z_n}{m^n} = M \ a.s.$$

In cases where extinction is possible we can quantify the extinction probability

**Proposition 2.3.18 :**
*The extinction probability of the Galton-Watson process is the smallest positive solution of the equation*

$$g(s) = s$$

The critical case is very difficult to treat because it can have wildly different behavior depending on the reproduction law $\mu$, in the most extreme case $\mu(\{1\}) = 1$ the process never goes extinct.

The sub-critical case behavior is rich but still treatable, the process goes almost surely to zero but the extinction time can be very large, one can wonder what happens before that extinction ? Does the process stays very close to zero for a long time or does it make long excursions away before coming back ? To answer these question we will need new objects.

Most of the definitions and results presented here come originally from the works of Darroch and Seneta ([15] [14], [47]) .

**Definition 2.3.19:**

1/ Let $\alpha$ be a probability measure on $E^*$. We say that $\alpha$ is a *quasi-limiting distribution* (QLD) for $Z$, if there exists a probability measure $\nu$ on $E^*$ such that, for any measurable set $A \subset E^*$,

$$\lim_{t\to\infty} \mathbb{P}_\nu \left( Z_t \in A | T_0 > t \right) = \alpha(A).$$

2/ We say that $Z$ has a *Yaglom limit* if there exists a probability measure $\alpha$ on $E^*$ such that, for any $x \in E^*$ and any measurable set $A \subset E^*$,

$$\lim_{t\to\infty} \mathbb{P}_x \left( Z_t \in A | T_0 > t \right) = \alpha(A).$$

If there exists a Yaglom limit, then it is unique.

3/ Let $\alpha$ be a probability measure on $E^*$. We say that $\alpha$ is a **quasi-stationary distribution** (QSD) if, for all $t \geq 0$ and any measurable set $A \subset E^*$,

$$\alpha(A) = \mathbb{P}_\alpha \left( Z_t \in A | T_0 > t \right).$$

The next proposition clarifies how these notions are related to one another.

**Proposition 2.3.20 :**
*Every QSD is a QLD and every QLD is a QSD. If the Yaglom limit exists then it is a QSD.*

Existence of a QSD gives great insight about the law of the absorption time, namely

**Proposition 2.3.21 :**
*Suppose that $\mu$ is a QSD for this process $Z_t$. Then there exists a positive real number $\theta(\mu)$, called the exponential rate of survival, such that*

$$\mathbb{P}_\mu[T_0 > t] = e^{-\theta(\mu)t}$$

Thus the law of $T_0$ is uniquely characterized by a single coefficient $\theta(\mu)$. Depending on how close $\theta(\mu)$ is to 1, the absorption time could be very large. Moreover we have the following result.

**Proposition 2.3.22 :**
*Suppose that $\mu$ is a QSD for the process $Z_t$. Then, for any $0 < \gamma < \theta(\mu)$,*

$$\mathbb{E}_\mu(e^{\gamma T_0}) < +\infty.$$

*In particular, there exists $z \in E^*$ such that $\mathbb{E}_z(e^{\gamma T_0}) < +\infty$.*

This Proposition implies that, if the population can escape extinction for too long times with positive probability, then the process has no QSD. This is the case for the critical Galton-Watson process: its extinction time is finite almost surely, but its expectation isn't finite.

Knowing that Yaglom limits and QLD are QSD, we will mostly focus on QSD whose existence is easier to prove. This Proposition equates the existence of QSD with the existence of fixed point for a spectral operator

**Proposition 2.3.23 :**
*Let $\mu$ be a probability measure on $E^*$. We assume that there exists a set $D \subset \mathcal{D}(L)$ such that, for all $t$ $P_t(D) \subset D$ and for any measurable subset $A \subset E^*$, there exists a uniformly bounded sequence $(f_n)$ in $D$ converging point-wisely to $\mathbf{1}_A$.*

*Then $\alpha$ is a quasi-stationary distribution if and only if there exists $\theta(\alpha) > 0$ such that*

$$\alpha(Lf) = -\theta(\alpha)\alpha(f),\ \forall f \in D.$$

We can even have a weaker characterization.

**Proposition 2.3.24** (see [12]) **:**
*Let $\nu \in \mathcal{P}^*$ and let $\beta > 0$ such that $\nu P_1 = \beta\nu$. Then $\beta < 1$ and there exists a QSD $\mu$ whose exponential rate of survival is $\theta = -log(\beta)$*

From that Proposition stem a variety of existence theorem for QSD using fixed points theorem, see e.g. [12]

If we go back to the Galton-Watson process it is pretty obvious that there won't be any QSD in the super-critical case, Proposition 2.3.22 implies that, if there exists a QSD, then $\mathbb{E}[T_0]$ is finite, in particular there can't be a QSD in the critical case. The sub-critical case was treated by Yaglom in 1947 [54].

**Theorem 2.3.25 :**
*Let $(Z_n)_{n\geq 0}$ be a Galton-Watson process with the reproduction generating function $g$. In the sub-critical case, the Yaglom limit exists and is the unique QSD of $Z$. Moreover, its generating function $\hat{g}$ fulfills*

$$\hat{g}(g(s)) = m\hat{g}(s) + 1 - m,\ \forall s \in [0, 1].$$

We now leave the Galton-Watson process behind and move to more general results in the case of finite space, with either continuous or discrete time.

Let $(Z_t)_{t\geq 0}$ be a Markov process (either in discrete or continuous time) with state space $E = \{0, 1, ..., N\}$, $N \geq 1$, we still assume that 0 is its unique absorbing state. We define the semi-group $(P_t)_{t\geq 0}$ as the sub-Markovian semi-group of the killed process and we still denote by $L$ the associated infinitesimal generator. The finiteness of the state space implies that $L$ and $P_t$ are matrices, and a probability measure on the finite space $E^*$ is a vector of non-negative entries whose sum is equal to 1. This will make existence theorem easier, mostly thanks to Perron-Frobenius Theorem.

**Theorem 2.3.26** (Perron-Frobenius Theorem) **:**
*Let $(P_t)$ be a sub-Markovian semi-group on $\{1,\cdots,N\}$ such that the entries of $P_{t_0}$ are positive for $t_0 > 0$. Thus, there exists a unique positive eigenvalue $\rho$, which is the maximum of the modulus of the eigenvalues, and there exists a unique left-eigenvector $\alpha$ such that $\alpha_i > 0$ and $\sum_{i=1}^N \alpha_i = 1$, and there exists a unique right-eigenvector $\pi$ such that $\pi_i > 0$ and $\sum_{i=1}^N \alpha_i \pi_i = 1$, satisfying*

$$\alpha P_{t_0} = \rho\,\alpha \ ; \ P_{t_0}\pi = \rho\,\pi.$$

*In addition, since $(P_t)$ is a sub-Markovian semi-group, $\rho < 1$ and there exists $\theta > 0$ such that $\rho = e^{-\theta}$. Therefore*

$$P_t = e^{-\theta t} A + \vartheta(e^{-\chi t}),$$

*where $A$ is the matrix defined by $A_{ij} = \pi_i \alpha_j$, and $\chi > \theta$ and $\vartheta(e^{-\chi t})$ denotes a matrix such that none of the entries exceeds $Ce^{-\chi t}$, for some constant $C > 0$.*

There are more general version of this theorem, see e.g. Bonsall paper [9] but this one is sufficient for our purpose. From this theorem we can then infer the following existence result.

**Theorem 2.3.27 :**
*Assume that $Z$ is an irreducible and aperiodic process before extinction, which means that there exists $t_0 > 0$ such that the matrix $P_{t_0}$ has only positive entries (in particular, it implies that $P_t$ has positive entries for $t > t_0$). Then the Yaglom limit $\alpha$ exists and is the unique QSD of the process $Z_t$.*

*Moreover, denoting by $\theta(\alpha)$ the extinction rate associated to $\alpha$, there exists a probability measure $\pi$ on $E^*$ such that, for any $i, j \in E^*$,*

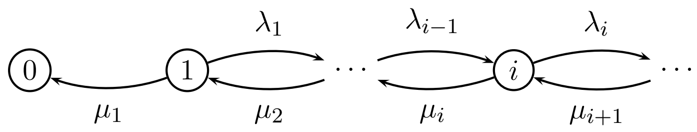
$$\lim_{t\to\infty} e^{\theta(\alpha)t}\,\mathbb{P}_i(Z_t = j) = \pi_i\,\alpha_j$$

*and*

$$\lim_{t\to\infty} \frac{\mathbb{P}_i(T_0 > t+s)}{\mathbb{P}_j(T_0 > t)} = \frac{\pi_i}{\pi_j} e^{-\theta(\alpha)s}.$$

A step beyond finite space Markov chain is the theory of birth and death processes, i.e. pure jump Markov process with only $+1$ or $-1$ jumps. Here we will consider birth and death processes with birth rates $(\lambda_i)_i$ and death rates $(\mu_i)_i$. We will assume both birth and death rates to be non-negative real numbers. More precisely we assume that, for $i > 0$ $\lambda_i > 0$ and $\mu_i > 0$, this is to ensure that our process won't get stuck (if

$\lambda_i = 0$ and $X_0 < i$ then $X_t \leqslant i$ a.s.) A birth and death process works that way: when at state $i$ we wait for an exponential time of parameter $\lambda_i$ before jumping to $i+1$ or independently, will wait for an exponential time of parameter $\mu_i$ before jumping to $i-1$, equivalently we wait for an exponential time of parameter $\lambda_i + \mu_i$ and then jump to $i+1$ with probability $\frac{\lambda_i}{\lambda_i+\mu_i}$ and jump to $i-1$ with probability $\frac{\mu_i}{\lambda_i+\mu_i}$. In population process it makes sense to assume that when everyone is dead then everyone stays dead, there are no immigration, resurrection or zombies. This assumption translates as $\lambda_0 = \mu_0 = 0$, i.e. 0 is an absorbing point. We will now study conditions for almost sure extinction and then study the possible QSD.




From now on, we will assume that $\lambda_i > 0$ and $\mu_i > 0$ for any $i \in \mathbb{N}^*$.

Before studying the conditions under which the process goes extinct we must study the conditions under which the process is "well defined for all times", i.e. the conditions under which it doesn't explode in finite time, such an explosion can occur if there is an accumulation of jumps near a time $T$. We denote by $(\tau_n)_n$ the sequence of the jump times of the process, either births or deaths.

**Theorem 2.3.28 :**
*The birth and death process does not explode in finite time, almost surely, if and only if* $\sum_n r_n = +\infty$, *where*

$$r_n = \frac{1}{\lambda_n} + \sum_{k=1}^{n-1} \frac{\mu_{k+1}\cdots\mu_n}{\lambda_k\lambda_{k+1}\cdots\lambda_n} + \frac{\mu_1\cdots\mu_n}{\lambda_1\cdots\lambda_n}.$$

Assuming that our process doesn't explode in finite time, we can now give the following theorem about its ultimate extinction.

**Theorem 2.3.29 :**
*The birth and death process goes almost-surely to extinction if and only if*

$$\sum_{k=1}^{\infty} \frac{\mu_1\cdots\mu_k}{\lambda_1\cdots\lambda_k} = +\infty.$$

Now that we have a population process that almost surely goes extinct we might ask ourselves what happens to this process before it goes extinct. More precisely we will wonder whether there exists any QSD. Firstly we give a necessary and sufficient condition for a probability measure $(\alpha_j)_{j\geq 1}$ to be a QSD for $Z$, then we will study more precisely the probability measures who verify this property.

**Theorem 2.3.30 :**
*The measure $(\alpha_j)_{j\geq 1}$ is a QSD if and only if*

1. $\alpha_j \geq 0, \ \forall j \geq 1$ *and* $\sum_{j\geq 1} \alpha_j = 1$.
2. $\forall \ j \geq 1$,

$$\lambda_{j-1}\alpha_{j-1} - (\lambda_j + \mu_j)\alpha_j + \mu_{j+1}\alpha_{j+1} = -\mu_1\alpha_1\alpha_j\,;$$
$$-(\lambda_1 + \mu_1)\alpha_1 + \mu_2\alpha_2 = -\mu_1\alpha_1^2.$$

The proof of this theorem is a good illustration of Proposition 2.3.23 in a simple case.

**Proof :**

Suppose that there exists a QSD $\nu$, then, using Proposition 2.3.23 with $D = \mathcal{D}(L)$, we obtain

$$\nu L = -\theta(\nu)\nu$$

where $L$ is the infinitesimal generator of the birth and death process on $\mathbb{N}^*$.

We thus get the linear system

$$-(\lambda_1 + \mu_1)\nu_1 + \mu_2\nu_2 = -\theta\nu_1$$
$$-(\lambda_j + \mu_j)\nu_j + \mu_{j+1}\nu_{j+1} + \lambda_{j-1}\nu_{j-1} = -\theta\nu_j \qquad \text{for all } j \geqslant 2$$

We know that $e^{-\theta t} = \mathbb{P}_\nu[T_0 > t]$, i.e. $\mathbb{P}_\nu[T_0 \leqslant t] = 1 - e^{-\theta t}$ As $t$ goes to 0 we know that the probability that the process makes two jumps become negligeable with respect to the probability that there is one jump. Thus

$$\lim_{t\to 0} \frac{\mathbb{P}_\nu[T_0 \leqslant t]}{t} = \lim_{t\to 0} \frac{\mathbb{P}_\nu[T_0 \leqslant t \text{ and the process makes at most one jump}]}{t}$$
$$= \frac{\mathbb{P}_\nu[\text{the process makes one jump from 1 to 0}]}{t} = \mu_1\nu_1$$

Hence $\theta = \mu_1\nu_1$.

Conversely if there exists a measure $\alpha$ that verify the equations

– $\alpha_j \geq 0, \ \forall j \geq 1$ and $\sum_{j\geq 1} \alpha_j = 1$.
– $\forall \ j \geq 2$,

$$\lambda_{j-1}\alpha_{j-1} - (\lambda_j + \mu_j)\alpha_j + \mu_{j+1}\alpha_{j+1} = -\mu_1\alpha_1\alpha_j\,;$$
$$-(\lambda_1 + \mu_1)\alpha_1 + \mu_2\alpha_2 = -\mu_1\alpha_1^2.$$

Then we can easily remark that $\alpha L = \mu_1 \alpha_1 \alpha$, which means that $\alpha$ is a QSD.

■

As a consequence, we get

**Corollary 2.3.31 :**
*Let us define inductively the sequence of polynomials $(H_n(x))_n$ as follows: For all $x \in \mathbb{R}$ and for $n \geq 2$,*

$$\begin{aligned} \lambda_n \ H_{n+1}(x) &= (\lambda_n + \mu_n - x) \ H_n(x) - \mu_{n-1} \ H_{n-1}(x) \, ; \\ \lambda_1 \ H_2(x) &= \lambda_1 + \mu_1 - x. \\ H_1(x) &= 1 \end{aligned}$$

*Then, any quasi-stationary distribution $(\alpha_j)_j$ satisfies for all $j \geq 1$,*

$$\alpha_j = \alpha_1 \ \pi_j \ H_j(\mu_1 \alpha_1)$$

*where*

$$\pi_1 = 1 \ ; \ \pi_n = \frac{\lambda_1 \cdots \lambda_{n-1}}{\mu_2 \cdots \mu_n}.$$

These two results give us more informations about the possible QSD but don't answer the question "Is there a QSD ?", the following result of Van Doorn and Pollett [49] does.

**Theorem 2.3.32 :**
*We denote $\xi_1$ the limit of the first positive root of $H_n$ as $n$ goes to infinity*

$$\xi_1 = \lim_{n \to \infty} \inf\{x > 0 \mid H_n(x) = 0\}$$

*We have the convergence*

$$\lim_{t \to \infty} \mathbb{P}_i(Z_t = j | T_0 > t) = \frac{1}{\mu_1} \ \pi_j \ \xi_1 \ H_j(\xi_1).$$

*In particular, we obtain*

$$\xi_1 = \lim_{t \to \infty} \mu_1 \mathbb{P}_1[Z_t = 1 | T_0 > t]$$

*and the following alternative :*

*Let*

$$S = \sum_{j \in \mathbb{N}} \frac{1}{\lambda_j \pi_j} \sum_{i=j+1}^{\infty} \pi_i \in \overline{\mathbb{R}}$$

1. *If $\xi_1 = 0$, there is no QSD.*
2. *If $S < \infty$ then $\xi_1 > 0$ and the Yaglom limit is the unique QSD.*
3. *If $S = \infty$ and $\xi_1 \neq 0$, then there is an infinite number of QSD, given by the family $(\beta_j(x))_{0 < x \leq \xi_1}$:*

$$\beta_j(x) = \frac{1}{\mu_1} \ \pi_j \ x \ H_j(x).$$

We will now give some more advanced results of Cattiaux, Collet, Lambert, Martínez, Méléard and San Martín [10]. These results are interesting in themselves but there are also interesting to us because of the similarity of form between the SDE studied by Cattiaux and al. and the SDE we study in the last part of this thesis. More precisely the interesting similarity is the square root term in the diffusion coefficient which gives rise to very interesting behaviors.

We will now study the quasi-stationarity for the logistic Feller diffusion process solution of the equation

$$dZ_t = \sqrt{Z_t}dB_t + (rZ_t - cZ_t^2)dt, \quad Z_0 > 0,$$

where the Brownian motion $B$ and the initial state $Z_0$ are given, and $r$ and $c$ are assumed to be positive.

The main theorem of Cattiaux, Collet, Lambert, Martínez, Méléard and San Martín is.

**Theorem 2.3.33 :**
*Assume that $Z_0$, $r$ and $c$ are positive. Then the Yaglom limit of the process $Z$ exists and is a QLD for $Z$ starting from any initial distribution. As a consequence, it is the unique QSD of $Z$.*

To prove this Theorem, Cattiaux and al. introduced a change of variable by studying the process $(X_t, t \geq 0)$ defined by $X_t = 2\sqrt{Z_t}$. $X$ is still absorbed at 0 and QSDs for $Z$ will be easily deduced from QSDs for $X$. From now on, we focus on the process $(X_t)$.

One can show that $(X_t)$ is the process defined by

$$dX_t = dB_t - q(X_t)dt,$$

where

$$q(x) = \frac{1}{2x} - \frac{rx}{2} + \frac{cx^3}{8}.$$

The graph of $q$ is as follows

$q$ is continuous on $\mathbb{R}_+^*$ but explodes at 0 at infinity.

We introduce the measure $\mu$, defined by

$$\mu(dy) = e^{-Q(y)}dy$$

where $Q$ is given by

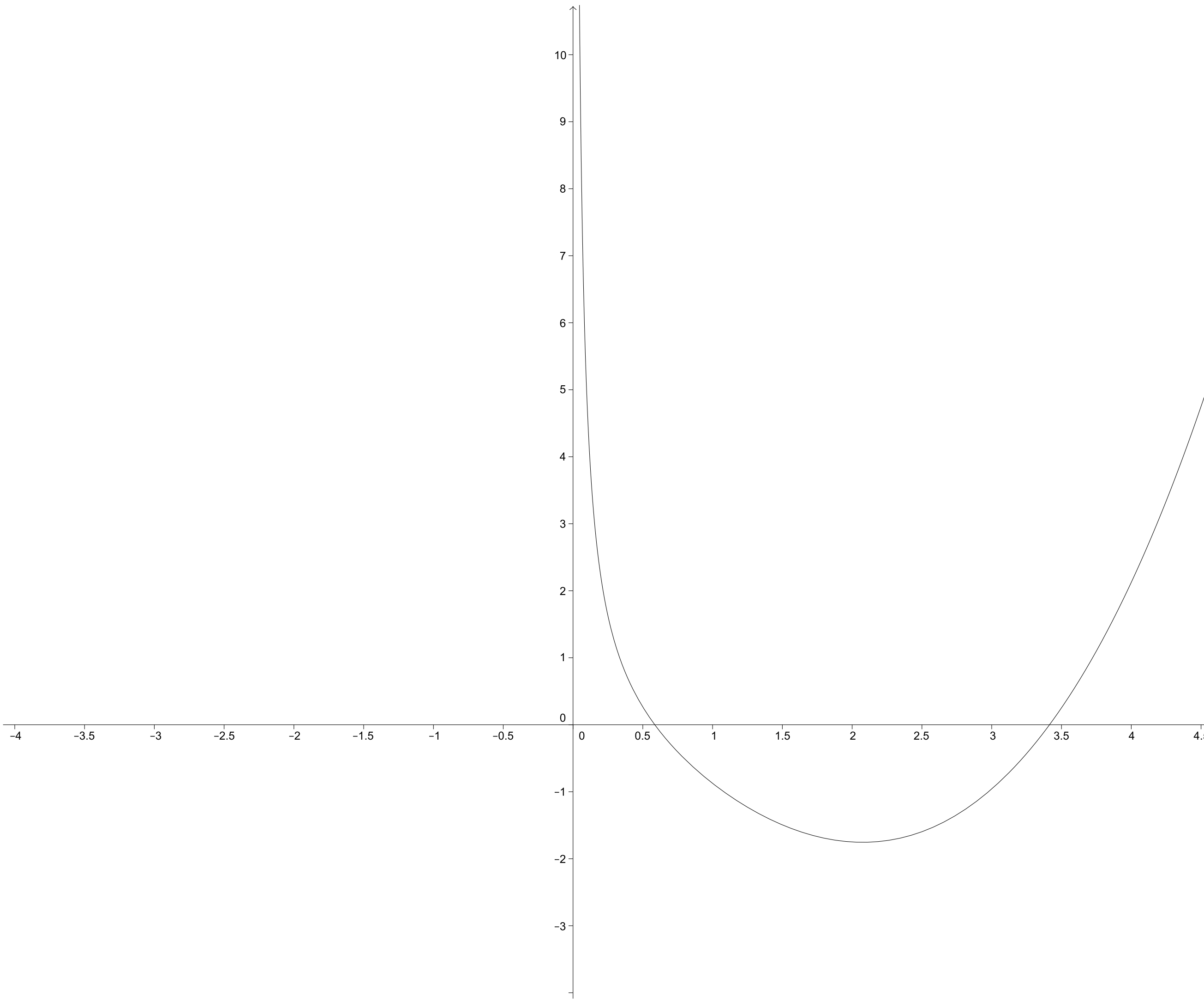
$$Q(y) = \int_1^y 2q(z)dz = \ln y + \frac{r}{2}(1 - y^2) + \frac{c}{16}(y^4 - 1).$$

In particular $-Q/2$ is a potential of the drift $-q$. The following result implies last Theorem but is more precise about the QSD.

**Theorem 2.3.34 :**
*Assume that $X_0$, $r$ and $c$ are positive. Then the Yaglom limit $\alpha$ of the process $X$ exists.*

*Moreover, there exists a positive function $\eta_1 \in \mathbb{L}^2(d\mu)$ such that*

Figure 2.4: Graph of $q(x)$



*1.*

$$\alpha(dx) = \frac{\eta_1(x)e^{-Q(x)}}{\int_{\mathbb{R}_+^*} \eta_1(y)e^{-Q(y)}dy}dx,$$

*2.* $\forall x \in \mathbb{R}_+^*$, $\lim_{t\to\infty} e^{\theta(\alpha)t}\mathbb{P}_x(T_0 > t) = \eta_1(x)$,

*3. there exists* $\chi > 0$ *such that,* $\forall x \in \mathbb{R}_+^*$,

$$\lim_{t\to+\infty} e^{-(\chi-\theta(\alpha))t} \left|\mathbb{P}_x\left(X_t \in A | T_0 > t\right) - \alpha(A)\right| < +\infty.$$

*4. the QSD $\alpha$ attracts all initial distribution, which means that $\alpha$ is a QLD for $X$ starting from any initial distribution. It is thus the unique QSD of $X$*

These results can also be generalized a to multi-type Feller process, namely a process $Z$ taking values in $(\mathbb{R}_+)^k$ defined by

$$dZ_t^i = \sqrt{\gamma_i Z_t^i}dB_t^i + (r_i Z_t^i - \sum_{j=1}^k c_{ij} Z_t^i Z_t^j)\ dt,$$

where all the constants $\gamma_i$, $r_i$ and $c_{i,j}$ are positive real numbers and $(B^i)_{i=1,\cdots,k}$ are independent standard Brownian motions independent of the initial state $Z_0$.

We denote by $T_{Ab}$ the first hitting time of $Ab = \{x \in (\mathbb{R}_+)^k \ |\exists i \text{ s.t. } x_i = 0\}$

**Proposition 2.3.35 :**
*The process $(Z_t)_t$ is well defined on $\mathbb{R}_+$. In addition, for all $x \in (\mathbb{R}_+)^k$,*

$$\mathbb{P}_x(T_{Ab} < +\infty) = 1$$

*and there exists $\lambda > 0$ such that*

$$\sup_{x\in(\mathbb{R}_+)^k} \mathbb{E}_x(e^{\lambda T_{Ab}}) < +\infty.$$

We can remark that, the fact that

$$\sup_{x\in(\mathbb{R}_+)^k} \mathbb{E}_x(e^{\lambda T_{Ab}}) < +\infty.$$

gives us hope that there might exists a QSD. In fact, using the same kind of renormalization of the process, Cattiaux and al. have proved the following theorem

**Theorem 2.3.36 :**
*Under the balance conditions*

$$c_{ij}\gamma_j = c_{ji}\gamma_i, \quad \forall i, j,$$

*there exists a unique quasi-stationary distribution* $\nu$ *for the process* $(X)$ *defined by* $X_t^i = \sqrt{\frac{Z_t^i}{\gamma_i}}$ *and the absorbing set* $Ab = \{x \in (\mathbb{R}_+)^k \; |\exists i \text{ s.t. } x_i = 0\}$*, which is the quasi-limiting distribution starting from any initial distribution: for any* $\mu$ *on* $(\mathbb{R}_+)^k$ *and any* $U \subset (\mathbb{R}_+)^k$*,*

$$\lim_{t \to +\infty} \mathbb{P}_\mu(X_t \in U | T_{Ab} > t) = \nu(U).$$

*Furthermore, there exist* $\lambda > 0$ *and a positive function* $\eta$ *such that*

$$\lim_{t \to +\infty} e^{\lambda t} \, \mathbb{P}_x(X_t \in U; T_{Ab} > t) = \eta(x) \, \nu(U).$$

## 2.4 Random perturbations of deterministic dynamical systems

We saw earlier that, when we want to study a real world problem we simplify it by stuffing everything we don't like or can't control in the "randomness" black box. However we still want our model to give good estimates about the real world problem. It then makes sense to study how far off we can get by using random approximations of our model. Also, as we said in the introduction, both the models we want to study are approximations of a deterministic dynamic. The interest of this is that some objects that are difficult to study when dealing with stochastic process are easier when dealing with deterministic process. We then want to use results that can easily be obtained on deterministic process to obtain informations on the more difficult objects in stochastic process.

### 2.4.1 Stochastic approximation algorithms with constant step size

The theory of stochastic approximation algorithms deals with random discrete time processes with are linked to a deterministic dynamical systems, typically they can be obtained from Euler approximation schemes by adding randomness. If $F$ is a Lipschitz vector field and $u_n$ is defined by $u_0 = y \in \mathbb{R}^d$ and

$$u_{n+1} = u_n + hF(u_n)$$

where $h$ is small, then $u_n$ approximates the trajectory of the flow of the ODE $\dot{x} = F(x)$ with starting point $y$.

A stochastic approximation algorithm will be a discrete time stochastic process of the form

$$x_{n+1} = x_n + \gamma_{n+1} V_{n+1}$$

Typically $\gamma_n$ is a "small" step size and $V_{n+1}$ is a random variable of the form $F(x_n, \xi_{n+1})$. At each time step the system receives a new information $\xi_{n+1}$ that causes $x_n$ to be updated according to a rule characterized by the function $F$.

Here we will concern ourselves with algorithms of the form

$$x_{n+1} = x_n + \gamma_{n+1} \left( F(x_n) + U_{n+1} \right)$$

We recognize something similar to the Euler scheme with step size $\gamma_n$, $x_n$ can now be seen as random approximations of the solution of the ODE $\dot{x} = F(x)$. Now two very different cases can occur:

– $\gamma_n \to 0$ as $n$ goes to $\infty$

– $\gamma_n$ doesn't depend on $n$

Here we will only treat the second case: Stochastic approximation algorithms with constant step size. We will give estimates about the precision of the approximations. Should a curious reader want more information about the first case, we refer him/her to the very thorough course of Benaïm in the Séminaire de Probabilités [2].

This subsection has been inspired by [3] where you can find more information on the subject and complete proofs of the results.

Here we will consider a family of discrete processes $(X_n^N)_{n\in\mathbb{N}}$ taking values in $K$ a compact subset of $\mathbb{R}^m$. The parameter $N$ indexing the processes may take either real or integer values, typically $X_n^N$ is a vector of population frequencies and $N$ represent the total population. We suppose that the $X_n^N$ are defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and denote by $\mathcal{F}_n^N$ the $\sigma$-algebra generated by $\{X_i^N, i = 1, ..., n\}$. We suppose that

$$X_{n+1}^N - X_n^N = \frac{1}{N}(F(X_n^N) + U_{n+1}^N)$$

where

(i) $F : \mathbb{R}^m \to \mathbb{R}^m$ is a locally Lipschitz vector field

(ii) $\mathbb{E}[U_{n+1}^N | \mathcal{F}_n^N] = 0$

(iii) There exists $\Gamma \geqslant 0$ such that $\|U_n^N\| \leqslant \sqrt{\Gamma}$

$F$ will be called the *mean field* associated to $X^N$, we want to compare the behavior of $X^N$ for large $N$ with that of the solutions of $\dot{x} = F(x)$.

We denote by $\{\varphi_t\}$ the flow induced by $F$. So as to compare the trajectory of $\{\varphi_t\}$ with those of $(X_n^N)$ it's convenient to introduce the continuous time process $\hat{X}^N : \mathbb{R} \to \mathbb{R}^m$ defined by

$$\hat{X}^N(k/N) = X_k^N \quad \forall k \in \mathbb{N}$$

and extended by linear interpolation on every $[k/N, (k+1)/N]$.

Let

$$D_N(T) = \max_{0\leqslant t\leqslant T} \|\hat{X}^N(t) - \varphi_t(X_0^N)\|$$

be the variable measuring the distance between the trajectories $t \mapsto \hat{X}^N(t)$ and $t \mapsto \varphi_t(X_0^N)$.

We first have a "law of large numbers" type result stating that, on finite time intervals, $X^N$ is close to $\varphi_.$ with very high probability

**Theorem 2.4.1** (see [3]) **:**
*For every $T > 0$, there exists $c > 0$ (depending only on $F$,$\Gamma$ and $T$) such that, for every $\varepsilon > 0$, and for $N$ large enough :*

$$\mathbb{P}[D_N(T) \geqslant \varepsilon] \leqslant 2me^{-\varepsilon^2 cN}$$

We know that our process stays close to the deterministic trajectories on fixed time intervals but how does it behave in the long run ? In fact our process may and will stray away for the deterministic flow but asymptotically it will stay long time near the forward trajectory of the system. More precisely

**Proposition 2.4.2** (see [3]) **:**
*Suppose $U$ is an open subset of $K$ such that $\overline{\gamma^+(x)} \subset U$ (where $\gamma^+(x) = \{\varphi_t(x)\ ;\ t \geqslant 0\}$ and suppose that $X_0^N$ goes to $x$ as $N$ goes to $\infty$. Then*

$$\mathbb{P}[\lim_{N\to\infty} \tau_{U^c}^N = \infty] = 1$$

*where $\tau_{U^c}^N = Inf\{t \geqslant 0 \mid \hat{X}^N(t) \in U^c\}$*

This proposition ensures that the process will stay for a long time close to the trajectory. If we apply this result by taking $U$ as a fundamental neighborhood of an attractor we will obtain that, with probability approaching one, the process enters $U$ within a finite horizon of time and stays in it for a very long time.

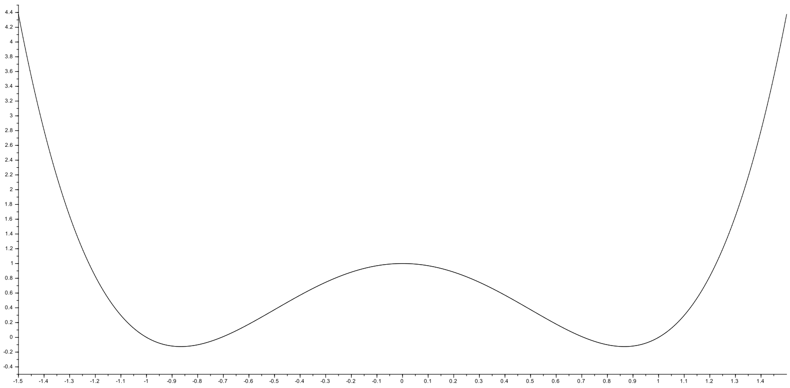
However this doesn't mean that our process will always stay near the omega limit set $\omega(x)$, due to randomness our process can and will take large excursions away from $\omega(x)$. These results only say that the time it will take to make these trips goes to infinity with the population size. A classical illustration of that fact is the case where $F = \nabla V$ and the potential $V$ has two wells, e.g. $V(x) = 2x^4 - 3x^2 + 1$. In that case $\varphi_t(x)$ is always of the sign of $x$ and converges to $sign(x) = \frac{x}{|x|}$ (with $sign(0) = 1$), then the process $X^N$ will quickly enter any neighborhood of $sign(x)$ and stay there for a very long time but, ultimately, it will make excursions between the two wells, staying long time near a well then, after a long time, quickly switch to another.

We now simplify the setting for our purpose, we suppose that $X^N = (X_n^N)_{n\in\mathbb{N}}$ is a Markov chain defined on a countable set $K_N$. We also suppose that $X^N$ admits at least one invariant probability measure $\pi_N$. Under good hypotheses we can expect $X_t^N$ to converge, in some sense, as $t$ goes to infinity, towards its invariant measure. We also know that the process stays close to the limit set of $\varphi$, in fact one can show that the process spends most of its time near the minimal center of attraction $M(\varphi)$, i.e. the closure of the union of the supports of the $\varphi$-invariant measures

$$M(\varphi) = \overline{\bigcup_{\mu\ \varphi\text{-invariant}} Supp(\mu)}$$

This result shows that those two behavior are compatibles

Figure 2.5: Graph of $V(x)$



**Theorem 2.4.3 :**
*The limit set of $\{\pi_N\}$ contains only probability measures that are invariant for the flow $\varphi_t$.*

In our example the limit set of invariant probability measures $\pi_N$ will be the set $\{a\delta_{-1} + (1-a)\delta_1 \mid a \in [0,1]\}$

When our population process gets absorbed by $\{0\}$ it will, of course, converge to the invariant measure $\delta_0$ which is also an invariant measure for $\varphi$. However this is trivial and doesn't give us any insight on the process. We should ask ourselves what happened before the absorption ? Was there some kind of metastable equilibrium ? And if so, can we study it by linking it to $\varphi$ ? In chapter 3 we will give answers to these questions.

When the Markov process approximating the deterministic process isn't a discrete time process but a diffusion process, the problem of studying how "good" the approximation is uses a whole new class of ingredients, this field is often called the Freidlin-Wentzell theory.

### 2.4.2 Freidlin-Wentzell theory

The Freidlin-Wentzell theory is a theory that deals with random perturbations of dynamical systems, in particular we will dwell here on diffusion processes that are approximations of ordinary differential equations, giving estimates about the precision of the approximations and linking the behavior of the diffusion with that of the associated deterministic dynamical system. We refer to Freidlin and Wentzell book [21] for the proofs and additional results.

In $\mathbb{R}^d$ we consider the stochastic differential equation with a small noise term

$$\begin{cases} dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\sigma(X_t^\varepsilon)dB_t \\ X_0^\varepsilon = x \in \mathbb{R}^d \end{cases}$$

**Theorem 2.4.4 :**
*Under the hypotheses that the coefficients of the SDE are Lipschitz and increase no faster than linearly, i.e.*

$$\sum_{i=1}^{d}(b_i(x)-b_i(y))^2+\sum_{i,j=1}^{d}(\sigma_{i,j}(x)-\sigma_{i,j}(y))^2 \leqslant K^2\|x-y\|^2$$

$$\sum_{i=1}^{d}(b_i(x))^2+\sum_{i,j=1}^{d}(\sigma_{i,j}(x))^2 \leqslant K^2(1+\|x\|)$$

*Then, for all $t>0$ and all $\delta>0$, we have*

$$M\|X_t^\varepsilon - x_t\| \leqslant \varepsilon a(t)$$

$$\lim_{\varepsilon\to 0}\mathbb{P}\left[\max_{0\leqslant s\leqslant t}\|X_s^\varepsilon - x_s\| > \delta\right]=0$$

*where $x_t$ is the solution of the ODE $\dot{y}=b(y)$ with the same initial condition $x$ as $X_t^\varepsilon$, $a(t)$ is an increasing function depending on $\|x\|$ and $K$*

If we want a more precise study of the approximations we can study large deviations estimates on finite time intervals.

We first define what will be our rate function:

**Definition 2.4.5:**
Let $\varphi\in\mathcal{C}^0([0;T],\mathbb{R}^d)$ we define:

$$S_{0T}(\varphi)=\begin{cases}\frac{1}{2}\int_0^T|\dot{\varphi}_s-b(\varphi_s)|^2ds \text{ if } \varphi \text{ is absolutly continuous}\\ +\infty \text{ otherwise}\end{cases}$$

Then $S_{0T}$ is a good rate function on $\mathcal{C}^0([0;T],\mathbb{R}^d)$.

We have the following large deviations principle:

**Theorem 2.4.6 :**
*The family of processes $X_t\varepsilon$ in $\mathcal{C}^0([0;T],\mathbb{R}^d)$ satisfies a large deviation principle with good rates function $S_{0T}(\varphi)$ in the sense that:*

$$\forall U \text{ open subset of } \mathcal{C}^0([0;T],\mathbb{R}^d) \quad \liminf_{\varepsilon\to 0}\varepsilon^2 log\mathbb{P}\left[X_t^\varepsilon\in U\right]\geqslant -\inf_{\varphi\in U}S_{0T}(\varphi)$$

$$\forall U \text{ closed subset of } \mathcal{C}^0([0;T],\mathbb{R}^d) \quad \limsup_{\varepsilon\to 0}\varepsilon^2 log\mathbb{P}\left[X_t^\varepsilon\in F\right]\leqslant -\inf_{\varphi\in F}S_{0T}(\varphi)$$

This theorem gives us estimates about the possible exit of the process from a domain $D$. More precisely let $D$ be a domain of $R^d$ and let $\tau^\varepsilon=\inf\{t\mid X_t^\varepsilon\notin D\}$ and define

$$H_D(t,x)=\{\varphi\in\mathcal{C}^0([0;T],\mathbb{R}^d)\mid\varphi_0=x,\ \varphi_t\in D\cup\partial D\}$$

$$\overline{H}_D(t,x)=\{\varphi\in\mathcal{C}^0([0;T],\mathbb{R}^d)\mid\varphi_0=x,\ \exists s\in[0,t]\ \varphi_s\notin D\}$$

We have the following estimates

**Theorem 2.4.7 :**
*Suppose that $\partial D = \partial \overline{D}$ (the boundary of $D$ is the boundary of its closure). Then*

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_x \left[ X_t^\varepsilon \in D \right] = - \inf_{\varphi \in H_D(t,x)} S_{0T}(\varphi)$$

$$\lim_{\varepsilon \to 0} \varepsilon \log \mathbb{P}_x \left[ \tau^\varepsilon \leqslant t \right] = - \inf_{\varphi \in \overline{H}_D(t,x)} S_{0T}(\varphi)$$

*Furthermore, if the minimizer $\hat{\varphi}$ of $S_{0t}(\varphi)$ on $\overline{H}_D(t,x)$ is unique and assumes a value in $\partial D$ at only one value $\hat{s} \in [0,t]$, then,*

$$\lim_{\varepsilon \to 0} \frac{\mathbb{E}_x \left[ \mathbb{1}_{\tau^\varepsilon \leqslant t} g\left(X_{\tau^\varepsilon}^\varepsilon\right) exp\left( \int_0^{\tau^\varepsilon} c\left(\left(X_s^\varepsilon\right) ds\right) \right] }{\mathbb{P}_x \left[ \tau^\varepsilon \leqslant t \right]} = g\left(\overline{\varphi_{\overline{s}}}\right) exp\left( \int_0^s c\left(\hat{\varphi}_s\right) ds \right)$$

When the domain $D$ is the basin of attraction of an equilibrium for the deterministic ODE $\dot{x} = b(x)$ we can refine the results, in particular we can locate the point of exit and give more insight about the time of exit. Let $D$ be an open bounded subset of $\mathbb{R}^d$ with smooth boundary $\partial D$, let $O \in D$ be an asymptotically stable equilibrium of the unperturbed ODE whose basin of attraction contains $D$. We suppose that $\forall x \in \partial D$ $\langle b(x), n(x) \rangle < 0$ where $n(x)$ is the exterior normal to the boundary $\partial D$ in $x$ and $\langle .;. \rangle$ is the classical scalar product in $\mathbb{R}^d$.

**Definition 2.4.8:**
We define the quasi-potential of the system by

$$V(O,x) = \inf\{S_{T_1 T_2}(\varphi);\ \varphi \in \mathcal{C}^0([T_1, T_2]\,; \mathbb{R}^d)\, \varphi_{T_1} = O\, \varphi_{T_2} = x\}$$

**Theorem 2.4.9 :**
*Suppose that there exists an unique point $y_0 \in \partial D$ such that $V(O, y_0) = \min_{y \in \partial D} V(O,y)$. Then*

$$\forall \delta > 0\ \forall x \in D \quad \lim_{\varepsilon \to 0} \mathbb{P}_x \left[ \rho(X_{\tau^\varepsilon}^\varepsilon, y_0) < \delta \right] = 1$$

*where $\rho$ is the Euclidean distance on $\mathbb{R}^d$*

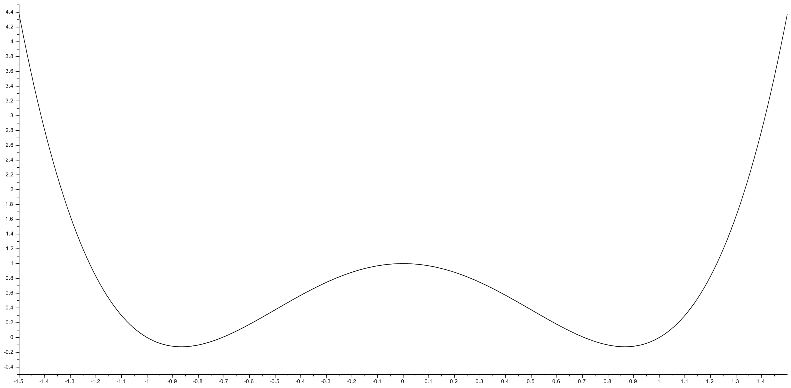
**Theorem 2.4.10 :**
*For all $x \in D$ and all $\alpha > 0$, we have the following estimates:*

$$\lim_{\varepsilon \to 0} \varepsilon^2 \ln \mathbb{E}_x \left[ \tau^\varepsilon \right] = V_0$$

$$\lim_{\varepsilon \to 0} \mathbb{P}_x \left[ e^{\frac{1}{\varepsilon^2}(V_0 - \alpha)} < \tau^\varepsilon < e^{\frac{1}{\varepsilon^2}(V_0 + \alpha)} \right] = 1$$

*where $V_0 = \min_{y \in \partial D} V(O,y)$*

Figure 2.6: Graph of $V(x)$



As an example we could look at the case where our SDE is of the form

$$dX_t = \nabla V(X_t)dt + \sqrt{\varepsilon}dB_t$$

and the potential $V(x) = 2x^4 - 3x^2 + 1$ is the two well potential we saw earlier.

When $b$ is the gradient of a potential function, one can show that the quasi-potential $V(a, b)$ coincides with $|V(a) - V(b)|$ . In our case $\mathbb{R}_+^*$ and $\mathbb{R}_-^*$ are basins of attraction of 1 and $-1$. Freidlin and Wentzell results give us that, starting from 1, our process will end up in the left well after a time of order $exp\left(\frac{V_0}{\varepsilon^2}\right)$, which in our case is $exp\left(\frac{1}{\varepsilon^2}\right)$.

# Chapter 3

# Quasi-stationary distributions for stochastic approximation algorithms with constant step size

This chapter consists of most of the article "Quasi-stationary distributions for stochastic approximation algorithms with constant step size".

**Contents**



## 3.1 Introduction

One of the most considered issue in theoretical ecology is to find out under what kind of conditions one can expect a population of interacting species (animals, plants, microorganisms, ...) to survive on the long term with no extinctions. When these conditions are met the interacting populations are said to persist or coexist. In the past, differential equations and nonlinear difference equations have been used to model these phenomena. Famous examples are Lotka [30] and Volterra [50] work on competitive and predator-prey interactions, Thompson [48], Nicholson and Bailey [35] on host-parasite interactions, and

Kermack and McKendrick [27] on disease outbreaks. For these deterministic models, persistence definitions sometimes vary but most authors link persistence with the existence of an attractor bounded away from the extinction states, in which case persistence holds over an infinite time horizon, see e.g. [44]. In order to refine these models and allow for some "roughness" and/or influence of unpredictable outer events, randomness has been added to these models, leading to Markov processes models. However, extinctions being absorbent states and species dying out with positive probabilities, the underlying theory of Markov processes shows that, extinction in finite time happens almost surely. Yet, in the real world, with large sized pools of population, we don't observe that inevitable extinction. This finite extinction time may then be very large and the system may remain in some sort of "metastable state" bounded away from extinction for a long time. In [20], Faure and Schreiber studied this problem for randomly perturbed discrete time dynamical systems, showing that, under the appropriate assumptions about the random perturbations and that there exists a positive attractor (i.e. an attractor which is bounded away from extinction states) for the unperturbed system, when they exist, quasi-stationary distributions concentrate on the positive attractors of the unperturbed system and that, the expected time to extinction for systems starting according to this quasi-stationary distribution grows exponentially with the system size. The aim of this paper is to extend their approach to a class of discrete time Markov process, that, up to a renormalization of time, can be seen as random perturbations of an ordinary differential equation.

In Section 2 we will introduce our setting and give some examples of systems that fall into it. Then, in Section 3 we will show that, under the hypothesis that the deterministic mean dynamic admits an interior attractor, the extinction time grows exponentially with the size of the system and that, when the system size goes to infinity, the limit set of the quasi-stationary distributions of the processes for the weak* convergence consists of invariant measures for the deterministic dynamic. Finally in Section 4 we will study the support of these invariant limiting measures and prove that, under some additional large deviations hypotheses, their support lies within attractors bounded away from the extinction states. To do that we will compare two different notions of chain-recurrence, one given by the large deviations functional and the other a slight variation on Conley's $\delta$-T chain-recurrence. Should the reader need some reminders about the objects used in this paper, he will find some basic properties and some references in the Appendix at the end.

## 3.2 Model, notations and hypotheses

We denote by $\Delta$ the simplex of $\mathbb{R}^d$.

$$\Delta = \{x \in \mathbb{R}^d \ ; \ \forall i = 1 \cdots d \ \ x_i \geqslant 0 \ \ \& \ \ \sum_{i=1}^{d} x_i = 1\}$$

We let $\mathring{\Delta}$ denote the relative interior of $\Delta$ and

$$\Delta_N = \Delta \cap \frac{1}{N}\mathbb{Z}^d$$

$$iD_N = \mathring{\Delta} \cap \frac{1}{N}\mathbb{Z}^d.$$

Let $F : \Delta \to \mathbb{R}^d$ be a locally Lipschitz vector field such that :

$$\forall x \in \Delta \quad \sum_{i=1}^{d} F_i(x) = 0$$

Unless specified otherwise, the topology considered will be the topology induced by the classical $\mathbb{R}^d$ metric topology on $\Delta$. Throughout the paper, if $A$ is a subset of a metric space $(E, d)$, we will denote by $N^\varepsilon(A)$ its $\varepsilon$-neighborhood

$$N^\varepsilon(A) = \{x \in E \ ; \ d(x, A) < \varepsilon\}.$$

We consider a family of Markov chains $(X_n^N)_{n\in\mathbb{N}}$ defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in the $d$-dimensional discrete simplex $\Delta_N$.

We denote by $\mathcal{F}_n^N$ the $\sigma$-algebra generated by $\{X_i^N, i = 1, ..., n\}$. For $A \in \mathcal{F}$ we let $\mathbb{P}_x[A] = \mathbb{P}[A|X_0 = x]$.

Throughout the paper the following hypothesis will always be assumed to hold.

**Standing Hypothesis 3.2.1 :**
*The Markov process $X^N$ has the following properties :*

*(i)* $X_{n+1}^N - X_n^N = \dfrac{1}{N}(F(X_n^N) + U_{n+1}^N)$

*(ii)* $\mathbb{E}[U_{n+1}^N | \mathcal{F}_n^N] = 0$

*(iii) There exists* $\Gamma \geqslant 0$ *such that* $\|U_n^N\| \leqslant \sqrt{\Gamma}$

*(iv) The boundary of the simplex is an absorbing set:*

  *(a) for all* $x \in \partial\Delta$ $\mathbb{P}_x[X_1^N \in \partial\Delta] = 1$

  *(b) for all* $x \in \Delta$ $\mathbb{P}_x[\exists n \ : \ X_n^N \in \partial\Delta] = 1$

*(v)* $X^N$ *restricted to* $\mathring{\Delta}_N$ *is irreducible*

$$\forall x, y \in \mathring{\Delta}_N \quad \mathbb{P}_x[\exists n \ : \ X_n = y] > 0$$

*and aperiodic*

$$\forall x \in \mathring{\Delta}_N \quad gcd(\{n \ ; \ \mathbb{P}_x[X_n = x] > 0\}) = 1$$

**Example 1** (Guiding Thread)**:**
Let $(p_{i,j}(x))_{i,j\in\{1\dots d\}}$ be a family of real-valued continuous functions on $\Delta$ such that, for all $x \in \Delta$ :

$$\forall i \neq j \quad p_{i,j}(x) = 0 \Leftrightarrow x_i x_j = 0, \tag{3.1}$$
$$p_{i,i}(x) = 0, \tag{3.2}$$
$$0 \leqslant p_{i,j}(x) \leqslant 1, \tag{3.3}$$
$$\sum_{i,j=1}^{d} p_{i,j}(x) \leqslant 1. \tag{3.4}$$

Let $(X_k^N)$ be the random walk on $\Delta_N$ defined by:

$$\mathbb{P}\left[X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)|X_k^n = x\right] = p_{i,j}(x)$$

where $(e_i)_{i=1\cdots d}$ is the canonical base of $\mathbb{R}^d$. This type of model often occurs in population games. In this setting $N$ represents the size of the population. Each individual plays a pure strategy $i$ and $X^N$ represents then the vector of proportion of players of each strategy. The jump $X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)$ means that an individual switches his strategy from $i$ to $j$ at time $k$. The conditions on the family $(p_{i,j})$ mean that :

– At each time $k$, it is always possible that a player switches from his strategy $i$ to another strategy $j$ that is currently in use in the population.
– No individual switches to an unused strategy. This makes sense for models based on strategy switching from imitations or models arising from ecology.

We define

$$p_i(x) = \sum_{j=1}^{d} p_{j,i}(x)$$
$$q_i(x) = \sum_{j=1}^{d} p_{i,j}(x)$$

$p(x)$ the vector of coordinates $p_i(x)$ and $q(x)$ the vector of coordinates $q_i(x)$.

In this case $F(x) = p(x) - q(x)$. Hypothesis 3.2.1(i) comes from the Markov property, 3.2.1(ii) and 3.2.1(iii) follow easily from the definition of the chain and 3.2.1(iv) and 3.2.1(v) follow from the fact that the functions $p_{i,j}$ are positive on the relative interior and vanish on the boundary.

A class of examples that falls in Example 1 setting is given by Imitative Protocols games, see e.g. [42]. Consider a population game with $d$ pure strategies, we let $U(x) = (U_1(x), \cdots U_d(x))$ denote the vector of payoffs when the population is in state $x \in \Delta$ and $\overline{U}(x) = \sum_{i=1}^{d} x_i U_i(x)$ denotes the average payoff at population state $x$. Imitative protocols are of the form

$$p_{i,j}(x) = x_i x_j r_{ij}(U(x), x)$$

with some additional assumptions on $r_{ij}$ to ensure that the problem is well-posed. Under such a protocol, at each time we pick an individual uniformly at random among the population and give him a revision opportunity. The opportunity unfolds as such

- The individual picks an opponent uniformly at random among the population (he/she can pick him/her-self) and observes his/her strategy.
- If the individual plays $i$ and the opponent plays $j$ then the individual switches from $i$ to $j$ with probability proportional to $r_{ij}$.

We now give some examples of Imitative Protocols

**Example 2** (Pairwise Proportional Imitation)**:**
After selecting an opponent the agent imitates only if the opponent's payoff is higher than his own, doing so with probability proportional to the payoff difference.

$$p_{i,j}(x) = x_i x_j (U_j(x) - U_i(x))^+$$

where $(y)^+$ stands for $\max(y, 0)$.

The mean dynamic generated by this protocol is

$$\begin{aligned}
\dot{x_i} &= \sum_{j=1}^{d} p_{j,i}(x) - \sum_{j=1}^{d} p_{i,j}(x) \\
&= \sum_{j=1}^{d} x_j x_i \left(U_i(x) - U_j(x)\right)^+ - x_i \sum_{j=1}^{d} \left(U_j(x) - U_i(x)\right)^+ \\
&= x_i \sum_{j=1}^{d} x_j (U_i(x) - U_j(x)) \\
&= x_i (U_i(x) - \overline{U}(x))
\end{aligned}$$

We then get the well known replicator dynamic, a dynamic extensively studied in ecology and evolutionary game theory, see e.g. [24]

**Example 3** (Aspiration and Random Imitation)**:**
A particular case of the former example is the aspiration and random imitation model, see e.g. [3], [6], [8] and [7]. At each time we pick an individual at random in the population and look at his/her "satisfaction", a payoff-like function $u_i(x)$ where $i$ is the type of the drawn individual. If this satisfaction is lower than a certain aspiration level then switch to another type chosen at random in the population, otherwise stay at current type. The morality of this model is that, if your type isn't efficient enough for your tastes then switch to another type. The aspiration levels are independent random variables uniformly distributed on intervals $[a_i(x), b_i(x)]$ with $a_i(x) \leqslant u_i(x) \leqslant b_i(x)$. This model gives us

$$p_{i,j}(x) = x_i x_j \frac{b_i(x) - u_i(x)}{b_i(x) - a_i(x)}$$

If we assume that the aspiration level bounds are not type-dependant, meaning $a_i(x) = a(x)$ and $b_i(x) = b(x)$ for all $i$ then we get a mean field given by

$$\dot{x}_i = x_i \frac{u_i(x) - \sum_{j=1}^d x_j u_j(x)}{b(x) - a(x)}$$

which is a replicator dynamic with fitness functions $f_i(x) = \frac{u_i(x)}{b(x)-a(x)}$

Alternatively we can assume that the aspiration levels follow the type payoff by the relation $b_i(x) = \beta_i u_i(x)$ and $a_i(x) = \alpha_i u_i(x)$ with $\alpha_i < 1 < \beta_i$. In this case we get a dynamic

$$\dot{x}_i = x_i \left( \sum_{j=1}^d x_j v_j - v_i \right)$$

where $v_j = \frac{\beta_j - 1}{\beta_j - \alpha_j}$.

Again this is a replicator dynamics with fitness function $f_i(x) = -v_i$.

**Example 4** (Imitation Driven by Dissatisfaction)**:**
In this protocol, when a $i$ player receives a revision opportunity, he opts to switch strategies with a probability that that is linearly decreasing in his current payoff. Should he decide to change, then he will imitate a randomly selected opponent. This protocol also falls under the former example of aspiration games with constant aspiration levels $a_i(x) = A$, $b_i(x) = B$. This gives the following dynamic

$$\dot{x}_i = \frac{1}{B - A}(F_i(x) - \overline{F}(x))$$

Again we recognize a replicator dynamic.

For other examples see e.g. [42].

## 3.3 Convergence of QSD and absorption time

We denote by $\{\varphi_t\}$ the flow induced by $F$. In order to compare the trajectory of $\varphi_t$ with those of $(X_n^N)$ it's convenient to introduce the continuous process $\hat{X}^N : \mathbb{R} \to \mathbb{R}^m$ defined by

$$\hat{X}^N(k/N) = X_k^N \quad \forall k \in \mathbb{N}$$

and extended on every interval $[k/N, (k+1)/N]$ by piecewise linear interpolation.

Let

$$D_N(T) = \max_{0 \leqslant t \leqslant T} \|\hat{X}^N(t) - \varphi_t(X_0^N)\|$$

be the variable measuring the distance between the trajectories $t \mapsto \hat{X}^N(t)$ and $t \mapsto \varphi_t(X_0^N)$.

We recall this convergence theorem of Benaïm and Weibull.

**Theorem 3.3.1 :**
*For every $T > 0$, there exists $c > 0$ (depending only on $F$,$\Gamma$ and $T$) such that, for every $\varepsilon > 0$, and for $N$ large enough :*

$$\mathbb{P}[D_N(T) \geqslant \varepsilon] \leqslant 2de^{-\varepsilon^2 cN}$$

For a detailed proof of this result, see [3].

We define $T_0^N$ to be the *absorption time.*

$$T_0^N = \inf\{n > 0; X_n^N \in \partial\Delta\}$$

Hypothesis 3.2.1(iv) implies that, whatever the initial state is, the process will almost surely be absorbed, i.e.

$$\forall x \in \Delta_N \qquad \mathbb{P}_x[T_0^N < \infty] = 1.$$

A probability measure $\mu$ on the discrete relative interior of the simplex $\mathring{\Delta}_N$ is said to be a *quasi-stationary distribution*, thereafter referred as QSD, if and only if, for every Borel set $A \subset \mathring{\Delta}_N$ and every $n > 0$,

$$\mathbb{P}_\mu[X_n^N \in A | T_0^N > n] = \mu(A).$$

We remark that, in this case, $\mu$ is a fixed point for the conditional evolution

$$\nu \mapsto \mathbb{P}_\nu[X_n^N \in . | T_0 > n]$$

The following proposition is a classic QSD result and follows easily from the Perron-Frobenius theorem.

**Proposition 3.3.2 :**
*For every $N$, there exists an unique quasi-stationary distribution $\mu^N$ obtained as the only left eigenvector $\mu^N$ of the transition matrix of the Markov chain restricted to $\mathring{\Delta}_N$ verifying*

$$\forall i \in \{1, \cdots d\} \quad \mu_i^N > 0 \quad ; \quad \sum_i \mu_i^N = 1$$

*The corresponding eigenvalue $0 < \rho_N = e^{-\theta_N} < 1$ is such that*

$$\mathbb{P}_{\mu^N}[T_0^N > n] = e^{-\theta_N n}$$

*Hence, starting from $\mu^N$, the expectation of $T_0^N$ is*

$$\mathbb{E}_{\mu^N}[T_0^N] = \frac{1}{1 - \rho_N}$$

**Proof :**
For a detailed proof see e.g. [32].

■

### 3.3.1 Absorption time

A set $A \subset \Delta$ is called an *attractor* for the flow $\{\varphi_t\}$ if

(i) $A$ is compact and invariant, i.e. for every $t \in \mathbb{R}$ $\varphi_t(A) = A$.

(ii) There exists a neighborhood $U$ of $A$, called a fundamental neighborhood, such that

$$\lim_{t \to \infty} d(\varphi_t(x), A) = 0$$

uniformly in $x$ in $U$.

**Theorem 3.3.3 :**
*Starting from $\mu^N$, the law of the absorption time and its expectation are given by Proposition 3.3.2. If we further assume that the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$, then, there exists $\gamma > 0$ such that the following estimate holds :*

$$0 \leqslant 1 - \rho_N \leqslant O\left(\frac{e^{-\gamma N}}{N}\right)$$

*Thus, there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mu^N}[T_0] \geqslant C N e^{\gamma N}$$

**<u>Proof :</u>**
Let $V \subset \mathring{\Delta}$ and let $k \in \mathbb{N}$. By the QSD property we have:

$$\begin{aligned} \rho_N^k \mu^N(V) &= \sum_{x \in \Delta_N} p_N^k(x, V) \mu^N(x) \\ &\geqslant \sum_{x \in V \cap \Delta_N} p_N^k(x, V) \mu^N(x) \\ &\geqslant \inf_{x \in V \cap \Delta_N} p_N^k(x, V) \mu^N(V). \end{aligned}$$

Thus

$$\rho_N^k \geqslant \inf_{x \in V \cap \Delta_N} p_N^k(x, V).$$

Let $U \subset \mathring{\Delta}$ be a fundamental neighborhood of the attractor $A$. We know that $d(\varphi_t(x), A)$ converges uniformly to 0 over $U$. Hence

$$\forall \varepsilon > 0 \quad \exists T(\varepsilon) > 0 \quad \forall t \geqslant T(\varepsilon) \quad \forall x \in U \quad d(\varphi_t(x), A) < \varepsilon.$$

Let $\alpha = d(A, U^c)$, $\varepsilon < \alpha$, $T = T(\varepsilon)$ and $\delta < \alpha - \varepsilon$.

For all $x \in U \cap \Delta_N$

$$\begin{aligned}
p_N^{[NT]}(x, U^c) &\leqslant \mathbb{P}_x[d(X_{[NT]}^N, A) > \alpha] \\
&\leqslant \mathbb{P}_x[d(X_{[NT]}^N, \varphi_T(x)) > \alpha - \varepsilon] \\
&\leqslant \mathbb{P}_x[d(X_{[NT]}^N, \hat{X}^N(T)) + d(\hat{X}^N(T), \varphi_T(x)) > \alpha - \varepsilon] \\
&\leqslant \mathbb{P}_x\left[D_N(T) > \alpha - \varepsilon - (\|F\| + \sqrt{\Gamma})\frac{NT - [NT]}{N}\right] \\
&\leqslant 2de^{-\delta^2 cN} \text{ for } N \text{ large enough (see Theorem 3.3.1)}
\end{aligned}$$

Then

$$\begin{aligned}
\rho_N^{[NT]} &\geqslant \inf_{x \in U \cap \Delta_N} p_N^{[NT]}(x, U) \\
&\geqslant 1 - \max_{x \in U \cap \Delta_N} p_N^{[NT]}(x, U^c) \\
&\geqslant 1 - 2de^{-\delta^2 cN}
\end{aligned}$$

Therefore

$$1 - \rho_N \leqslant 1 - \left(1 - 2de^{-\delta^2 cN}\right)^{\frac{1}{[NT]}}$$

$$\begin{aligned}
\left(1 - 2de^{-\delta^2 cN}\right)^{\frac{1}{[NT]}} &= e^{\frac{1}{[NT]} \log\left(1 - 2de^{-\delta^2 cN}\right)} \\
&= e^{\frac{-1}{[NT]} 2de^{-\delta^2 cN}} + o(e^{-\delta^2 cN})) \\
&= 1 - \frac{2de^{-\delta^2 cN}}{[NT]} + o\left(\frac{2de^{-\delta^2 cN}}{[NT]}\right)
\end{aligned}$$

In conclusion we have

$$0 \leqslant 1 - \rho_N \leqslant O\left(\frac{e^{-\gamma N}}{N}\right)$$

■

### 3.3.2 Convergence of the QSD to an invariant measure

A probability measure $\mu$ on $\Delta$ is called an *invariant measure* for the flow $\{\varphi_t\}$ if, for all $t \in \mathbb{R}$ and all Borel set $A \in \mathcal{B}(\Delta)$, $\mu(\varphi_t^{-1}(A)) = \mu(A)$.

**Theorem 3.3.4 :**
*We suppose that the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$. Then the set of limit points of $\{\mu^N\}$ for the weak* topology is a subset of the set of invariant measures for the flow $\{\varphi_t\}$.*

**Proof :**

Let $f$ be a Lipschitz function from $\Delta$ to $\mathbb{R}$ with constant $L$. We suppose that the sequence $\mu^N$ weakly converges to a measure $\mu$. Let $t > 0$. We want to prove that

$$\lim_{N\to\infty} \int f(x)\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx) = 0$$

The QSD property gives us that, for all $k$

$$\int f(x)\mu^N(dx) = \int \mathbb{E}_x\left[f(X_k^N)\middle|T_0^N > k\right]\mu^N(dx)$$

Let

$$I = \left|\int f(x)\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx)\right|$$

Then, for all $k$,

$$\begin{aligned} I &= \left|\int f(x)\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx)\right| \\ &= \left|\int \mathbb{E}_x\left[f(X_k^N)\middle|T_0^N > k\right]\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx)\right| \\ &= \left|\int \mathbb{E}_x\left[f(X_k^N) - f(\varphi_t(x))\middle|T_0^N > k\right]\mu^N(dx)\right| \end{aligned}$$

In particular, for $k = [Nt]$.

$$I = \left|\int \mathbb{E}_x\left[f(X_{[Nt]}^N) - f(\varphi_t(x))\middle|T_0^N > [Nt]\right]\mu^N(dx)\right|$$

By Theorem 3.3.1, we know that, for $N$ large enough, we have

$$\mathbb{P}_x[D_N(t) > \delta] \leqslant 2de^{-\delta^2 cN}.$$

Thus

$$\mathbb{E}_x[D_N(t)] = \int_0^{+\infty} \mathbb{P}_x[D_N(t) > \delta]d\delta \leqslant \int_0^{+\infty} 2de^{-\delta^2 cN}d\delta = \frac{d\sqrt{\pi}}{\sqrt{cN}}$$

From Theorem 3.3.3, we can infer $1 - \rho_N \leqslant O\left(\frac{e^{-\gamma N}}{N}\right)$.

This implies $e^{\theta_N[Nt]} - 1 \underset{N\to+\infty}{\longrightarrow} 0$, which gives us the boundedness of $e^{\theta_N[Nt]}$.

Hence

$$\begin{aligned}
I &= \left| \int \mathbb{E}_x \left[ f(X^N_{[Nt]}) - f(\varphi_t(x)) \middle| T_0^N > [Nt] \right] \mu^N(dx) \right| \\
&\leqslant \left| \int \frac{\mathbb{E}_x \left[ f(X^N_{[Nt]}) - f(\varphi_t(x)) \right]}{\mathbb{P}_x \left[ T_0^N > [Nt] \right]} \mu^N(dx) \right| \\
&\leqslant \left| \int \frac{\mathbb{E}_x \left[ L | X^N_{[Nt]} - \hat{X}^N(t) + \hat{X}^N(t) - \varphi_t(x) | \right]}{\mathbb{P}_x \left[ T_0^N > [Nt] \right]} \mu^N(dx) \right| \\
&\leqslant \left| \int \frac{\mathbb{E}_x \left[ L (D_N(t) + (Nt - [Nt]) \frac{1}{N} (\|F\| + \Gamma)) \right]}{\mathbb{P}_x \left[ T_0^N > [Nt] \right]} \mu^N(dx) \right| \\
&\leqslant \left| L \left( \frac{d\sqrt{\pi}}{\sqrt{cN}} + \frac{\|F\| + \Gamma}{N} \right) e^{\theta_N [Nt]} \right| \underset{N \to +\infty}{\longrightarrow} 0
\end{aligned}$$

■

## 3.4 Support of the limiting measure

Let $L = L(\{\mu_N\})$ denote the limit set of the sequence $(\mu_N)_{N \in \mathbb{N}}$ for the weak* topology. In view of Theorem 3.3.4, $L$ consists of invariant measures. As the QSD have their support inside $\mathring{\Delta}$, it is natural to study whether the limiting measure also take their support in $\mathring{\Delta}$. However, by the Poincaré Recurrence Theorem, every $\mu \in L$ is supported by the Birkhoff center

$$BC(\varphi) = \overline{\{x \in \Delta \; ; \; x \in \omega(x)\}}$$

Since the boundary of $\Delta$ intersects the Birkhoff center (e.g the vertices of the simplex are equilibria and thus inside the Birkhoff center), knowing that the QSD converges to an invariant measure is not enough, we have to further study the support of the measure $\mu$ to ensure that it is strictly inside the interior of the simplex. For that we will need large deviation assumptions.

**Hypothesis 3.4.1 :**
*For every $\alpha > 0$, there exists a rate function*

$$\begin{aligned}
S_\alpha : V_\alpha \times \mathbb{R} \times \mathcal{C}_x([0,T], V_\alpha) &\longrightarrow \overline{\mathbb{R}}_+ \\
(x, T, \phi) &\longmapsto S_\alpha(x, T, \phi)
\end{aligned}$$

*with the following properties, where $V_\alpha = \Delta \setminus \overline{N^\alpha(\partial\Delta)}$, $\mathcal{C}([0,T], V_\alpha)$ is the set of continuous functions $\psi$ from $[0,T]$ to $V_\alpha$ and $\mathcal{C}_x([0,T], V_\alpha)$ is the set of continuous functions $\psi$ from $[0,T]$ to $V_\alpha$ such that $\psi(0) = x$, both equipped with the topology of uniform convergence.*

– *For every* $s \in ]0,\infty[$ *and* $T > 0$*, the set*

$$\{\phi \in \mathcal{C}_x([0,T], V_\alpha) \ s.t. \ S_\alpha(x,T,\phi) \leqslant s\}$$

*is a compact set*

– *For* $x \in \mathring{\Delta}$ *and* $T > 0$*,* $S_\alpha(x,T,\phi) = 0 \Leftrightarrow \dot{\phi}_s = F(\phi_s) \ \forall s \in [0,T]$

– $\hat{X}^N$ *satisfies a large deviation principle with rate function* $S$ *and speed* $1/N$ *uniformly in* $x$ *on compact subsets of* $V_\alpha$*,*
*i.e. for* $K$ *compact subset of* $V_\alpha$*,* $T > 0$ *and* $A \subset \mathcal{C}_x([0,T], V_\alpha)$

$$-\sup_{x\in K}\inf_{\phi\in\mathring{A}} S_\alpha(x,T,\phi) \leqslant \liminf_{N\to\infty} \frac{1}{N} log \inf_{x\in K} \mathbb{P}_x[\hat{X}^N \in A]$$
$$\limsup_{N\to\infty} \frac{1}{N} log \sup_{x\in K} \mathbb{P}_x[\hat{X}^N \in A] \leqslant -\inf_{x\in K}\inf_{\phi\in\overline{A}} S_\alpha(x,T,\phi)$$

*In particular when* $K = \{x\}$ *we get the "classical" large deviation principle*

$$-\inf_{\phi\in\mathring{A}} S_\alpha(x,T,\phi) \leqslant \liminf_{N\to\infty} \frac{1}{N} log \mathbb{P}_x[\hat{X}^N \in A]$$
$$\limsup_{N\to\infty} \frac{1}{N} log \mathbb{P}_x[\hat{X}^N \in A] \leqslant -\inf_{\phi\in\overline{A}} S_\alpha(x,T,\phi)$$

– $S_\alpha$ *is linear with regards to the concatenation of functions, i.e. if* $\tilde{T} < T$

$$S_\alpha(x,T,\phi) = S(x,\tilde{T},\phi|_{[0,\tilde{T}]}) + S_\alpha(\phi(\tilde{T}), T-\tilde{T}, \phi|_{[\tilde{T},T-\tilde{T}]})$$

– $\lim_{T\to 0} S_\alpha(x,T,\phi) = 0$ *uniformly in* $x \in V_\alpha$ *and* $\phi \in \mathcal{C}_x([0,T], V_\alpha)$*.*

– *For every* $\alpha' < \alpha$*, every* $x \in V_\alpha$*, every* $T$ *and every* $\phi \in \mathcal{C}([0,T], V_\alpha)$ *we have* $S_\alpha(x,T,\phi) = S_{\alpha'}(x,T,\phi)$*. Due to this compatibility property we will omit the* $\alpha$ *from* $S_\alpha$ *and only write* $S(x,T,\phi)$

**Hypothesis 3.4.2 :**
$\forall c > 0 \ \exists U_c$ *an open neighborhood of* $\partial\Delta$ *such that*

$$\lim_{N\to\infty}\inf_{x\in U_c} \frac{1}{N} log \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] \geqslant -c$$

### 3.4.1 Guiding Thread Example

Before going further we will verify that Hypotheses 3.4.1 and 3.4.2 hold for the nearest neighbor random walk model introduced in the Guiding Thread Example 1.

**Example 5:**
To show that Hypothesis 3.4.1 holds we will use Theorem 6.3.3 in [17] with Conditions 6.2.1 and 6.3.1, which will gives us a Laplace principle for $\hat{X}^N|_{[0,T]}$ that holds uniformly on compacts, and Theorem 1.2.3 in [17] will, in turn, give us the desired uniform on compacts large deviation principle.

We only need to show that Conditions 6.2.1 and 6.3.1 holds

When transcribing our model in [17] setting, we get, for $x \in V_\alpha$

$$\mu(dy|x) = \sum_{i \neq j} p_{i,j}(x)\delta_{e_j - e_i}(dy) + \left(1 - \sum_{i \neq j} p_{i,j}(x)\right)\delta_0(dy)$$

We define

$$H_\mu(x, \alpha) = log \int_{\mathbb{R}^d} exp\langle \alpha, y\rangle \mu(dy|x)$$

and

$$L_\mu(x, \beta) = \sup_{\alpha \in \mathbb{R}^d} \{\langle \alpha, \beta\rangle - H_\mu(x, \alpha)\}$$

We refer to Chapter 6.2 in [17] for elementary properties of these functions.

**Definition 3.4.3:**
$\mu(dy|x)$ is said to verify Condition A (called 6.2.1 in [17]) if

(i) For each $\alpha \in \mathbb{R}^d$, $\sup_{x \in \mathbb{R}^d} H_\mu(x, \alpha) < \infty$

(ii) The function mapping $x \in \mathbb{R}^d \mapsto \mu(.|x)$ is continuous in the topology of weak convergence.

**Definition 3.4.4:**
$\mu(dy|x)$ is said to verify Condition B ( called 6.3.1 in [17]) if

(i) The relative interior of the convex hull of the support of $\mu(dy|x)$, $Ri(Conv(Supp(\mu(.|x))))$, doesn't depend on $x$.

(ii) $0 \in Ri(Conv(Supp(\mu(.|x))))$

Let $\alpha > 0$, we'll now prove the L.D.P. on $V_\alpha$ The conditions, as they are written, demand that $x$ may take values in all of $\mathbb{R}^d$ while, in our model, we only use $\Delta$. To remedy to that we'll first embed our $d$-dimensional simplex in $\mathbb{R}^{d-1}$ and then extend $\mu$ to a kernel $\eta$ defined on all of $\mathbb{R}^{d-1}$ by taking $\eta(dy|x) = \mu(dy|p_\alpha(x))$, where $p_\alpha$ is the convex projection on $\overline{V_\alpha}$.

This way we have a probability kernel $\eta$ that is defined on all of $\mathbb{R}^{d-1}$, it is then easy to verify that Conditions A and B hold for $\eta$. We thus get a LDP for $\eta$ on all of $\mathbb{R}^{d-1}$ with speed $1/N$ and rate function

$$S_\eta = \begin{cases} \int_0^T L_\eta(\phi(t), \dot{\phi}(t))dt \text{ if } \phi \text{ is uniformly continuous} \\ \infty \text{ otherwise} \end{cases}$$

We now remark that, when $x \in V_\alpha$, $\eta(dy|x) = \mu(dy|x)$ and thus $H_\eta(x,v) = H_\mu(x,v)$ and $L_\eta(x,u) = L_\mu(x,u)$.

From that we deduce that, if $\phi \in \mathcal{C}([0,T], V_\alpha)$, then $L_\eta(\phi,\dot{\phi}) = L_\mu(\phi,\dot{\phi})$ and finally $S_\eta(x,T,\phi) = S_\mu(x,T\phi)$.

We finally get that Hypothesis 3.4.1 holds for our nearest neighbor random walk with the following rate function.

$$S(x,T,\phi) = \begin{cases} \int_0^T L_\mu(\phi(t),\dot{\phi}(t))dt \text{ if } \phi \text{ is uniformly continuous} \\ \infty \text{ otherwise} \end{cases}$$

We still have to verify that the wanted properties holds for this rate function:

- From Proposition 6.2.4 in [17] we get that, for every $s \in ]0,\infty[$ and $T > 0$, the set

$$\{\phi \in \mathcal{C}_x([0,T],\Delta) \text{ s.t. } S(x,T,\phi) \leqslant s\}$$

  is a compact set.
- Let $x \in V_\alpha$, $T > 0$, it is already known that

$$S(x,T,\phi) = 0 \quad \Leftrightarrow \quad \dot{\phi}_s = F(\phi_s) \qquad \forall s \in [0,T]$$

- The LDP comes from Theorem 6.3.3 in [17]
- The linearity of $S(x,T,\phi)$ follows easily from it's definition as an integral.

Let's now prove that Hypothesis 3.4.2 holds too.

**Proposition 3.4.5 :**

*Suppose that, for every couple $(i,j) \in \{1,\cdots d\}^2$ there exists $k \in \mathbb{N}$ such that $\dfrac{\partial^k p_{i,j}(x)}{\partial x_i^k} \neq 0$ on $\{x \in \Delta \ ; \ x_i = 0\}$. Then Hypothesis 3.4.2 holds*

<u>**Proof :**</u>

We want to ensure that, for all $c > 0$ there exists an open neighborhood $U_c$ of $\partial\Delta$ such that

$$\lim_{N\to\infty} \inf_{x\in U_c} \frac{1}{N} log \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] \geqslant -c.$$

We know that $q_i(x)$ and $p_i(x)$ go to 0 as $x_i$ goes to 0 and that there exists $k \in \mathbb{N}$ such that $\dfrac{\partial^k q_i(x)}{\partial x_i^k} > 0$ on $\{x \in \Delta \ ; \ x_i = 0\}$ and thus $\dfrac{\partial^k q_i(x)}{\partial x_i^k} > a$ on a sufficient small neighborhood of $\{x \in \Delta \ ; \ x_i = 0\}$ with $a > 0$.

Let $1 > b > 0$ and let $x \in U_c = \{x \ ; \ \exists i \ x_i < b\}$. We have

$$\begin{aligned} \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] &\geqslant \mathbb{P}_x[X^N_{[Nb]} \in \partial\Delta] \\ &\geqslant \prod_{j=1}^{[Nb]} q_i(x^{(j)}) \end{aligned}$$

where $(x^{(j)})_j$ is a sequence of points in $\Delta_N$ such that $x_i^{(j)} = \frac{j}{N}$.

If $b$ is small enough we have, for $j \leqslant [Nb]$

$$q_i(x^{(j)}) \geqslant a \left(\frac{j}{N}\right)^k$$

Then

$$\prod_{j=1}^{[Nb]} q_i(x^{(j)}) \geqslant a^{[Nb]} \frac{[Nb]!}{N^{[Nb]}}$$

Thus

$$\inf_{x \in U_c} \frac{1}{N} log \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] \geqslant C \left( \frac{[Nb]}{N}(log(a) - 1) + \frac{[Nb]}{N} log \frac{[Nb]}{N} + \frac{1}{2N} log(2\pi[Nb]) \right)$$

where $C$ is a constant

As $N$ goes to infinity, the right-hand term goes to $Cb(log(a) - 1) + Cblog(b)$ which is greater than $-c$ for $b$ small enough. Thus, for $N$ large enough

$$\lim_{N\to\infty} \inf_{x \in U_c} \frac{1}{N} log \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] \geqslant -c$$

Hence our random walk model satisfies Assumption 3.4.2.

■

Finally we showed that both hypotheses holds for the nearest neighbor random walk model introduced in Example 1.

### 3.4.2 Back to the general case

**Definition 3.4.6:**
We define

$$\mathcal{L}(x, y) = \limsup_{T\to\infty} \inf_{\phi \in \mathcal{C}_x([0,T]),\, \phi(T)=y} S(x, T, \phi)$$

$\mathcal{L}(x, y)$ represents the cost for going from $x$ to $y$. We will say that $x$ $\mathcal{L}$-leads to $y$ (denoted by $x \rightsquigarrow_{\mathcal{L}} y$) if, for every $\varepsilon > 0$, there exists a path of points $x = \xi_0, \xi_1, \cdots \xi_{n(\xi)} = y$ such that

$$A_{n(\xi)}(\xi) = \sum_{k=1}^{n(\xi)} \mathcal{L}(\xi_k, \xi_{k+1}) < \varepsilon$$

We define $B_{\mathcal{L}}(x, y) = \inf_{\xi \text{ linking } x \text{ to } y} A_{n(\xi)}(\xi)$.

Thus $x \rightsquigarrow_{\mathcal{L}} y$ iff $B_{\mathcal{L}}(x, y) = 0$

We will say that $x$ is $\mathcal{L}$-chain recurrent if $x \rightsquigarrow_{\mathcal{L}} x$ and will denote by $\mathcal{R}_{\mathcal{L}}(\varphi)$ the set of all $\mathcal{L}$-chain recurrent points.

If $x$ and $y$ are two points of $\mathcal{R}_{\mathcal{L}}$ verifying $x \rightsquigarrow_{\mathcal{L}} y$ and $y \rightsquigarrow_{\mathcal{L}} x$ we will then denote $x \sim_{\mathcal{L}} y$. The equivalence classes for this relation will be called $\mathcal{L}$-basic classes. We define a partial order on these classes by $[x] \prec_{\mathcal{L}} [y]$ if $x \rightsquigarrow_{\mathcal{L}} y$. A maximal $\mathcal{L}$-basic class will be called a $\mathcal{L}$-quasi-attractor.

**Hypothesis 3.4.7 :**
*There is only a finite number of $\mathcal{L}$-basic classes in $\mathring{\Delta}$ denoted by $K_i$, $i = 1 \cdots \nu$. We suppose that they are closed sets and indexed in such a way that the $k$ first $\{K_i\}_{i=1\cdots k}$ are the $\mathcal{L}$-quasi-attractors and the $\nu - k$ others aren't.*

**Proposition 3.4.8 :**
*The function $\mathcal{L}$ has the following properties : for every sequence $(x_n)_{n\in\mathbb{N}} \in V_\alpha$ converging to an $x \in V_\alpha$ and every $y \in V_\alpha$ we have*

$$\lim_{n\to\infty} \mathcal{L}(x_n, y) = \mathcal{L}(x, y) \qquad \lim_{n\to\infty} \mathcal{L}(y, x_n) = \mathcal{L}(y, x)$$

**Proof :**
This proposition follows easily from Hypothesis 3.4.1

■

The following theorem is the main result of this section, giving us more insight in the support of the limiting measures $\mu$.

**Theorem 3.4.9 :**
*We suppose that the flow $\{\varphi_t\}$ associated with the mean dynamic $\dot{x} = F(x)$ has an interior attractor. Under Hypotheses 3.4.1, 3.4.2 and 3.4.7, we have :*

*Every limiting measure $\mu$ has its support in the union of the $\mathcal{L}$-quasi-attractor, i.e. in $\bigcup_{i=1}^{k} K_i$*

We will prove this theorem under an intermediary set of hypotheses then we will prove that the announced hypotheses imply the intermediary hypotheses.

### 3.4.3 Absorption-preserving pseudo-orbit

Here we introduce a different notion of chains for our dynamical system using absorption preserving $\delta, T$ pseudo-orbit, an analog to $\delta, T$ pseudo-orbits introduced by Conley [13], which have been extensively studied in the past.

**Definition 3.4.10:**
Let $\{\varphi_t\}_{t\in\mathbb{R}}$ be a flow given by an ordinary differential equation on $(\Delta, d)$ for which $\partial\Delta$

is an invariant set. We will call $(\delta, T)$ absorption preserving pseudo-orbit ($\delta, T$-ap-pseudo orbit) from $x$ to $y$ a piecewise continuous path

$$x = x_0, \{\varphi_t(x_1)\,;\, t \in [0, t_1]\}, \{\varphi_t(x_2)\,;\, t \in [0, t_2]\}$$
$$\cdots \{\varphi_t(x_1)\,;\, t \in [0, t_k]\}, x_{k+1} = y \quad k \geqslant 1$$

which is uniquely defined by the sequences of points $x_0, \cdots x_{k+1}$ and times $t_1, \cdots t_k$ such that the following hypotheses hold:

$$\begin{cases} d(x, x_1) < \delta \\ d(\varphi_{t_j}(x_j), x_{j+1}) < \delta \quad \forall j = 1 \cdots k \\ t_i \geqslant T \quad \forall i = 1 \cdots k \\ x_j \in \partial\Delta \Rightarrow x_{j+1} \in \partial\Delta \quad \forall j = 0 \cdots k \end{cases}$$

We will denote then $x \rightsquigarrow_{ap,\delta,T} y$

If $x \rightsquigarrow_{ap,\delta,T} y$ for every $\delta > 0$ and every $T > 0$, we will denote $x \rightsquigarrow_{ap} y$

The point $x$ will be said to be ap-chain recurrent if $x \rightsquigarrow_{ap} x$, we define $\mathcal{R}_{ap}(\varphi)$ the set of ap-chain recurrent points.

If $x$ and $y$ are two points of $\mathcal{R}_{ap}(\varphi)$ such that $x \rightsquigarrow_{ap} y$ and $y \rightsquigarrow_{ap} x$ we will write $x \sim_{ap} y$. The equivalence classes for this relation will be called ap-basic classes. We define a partial order on these classes by $[x] \prec_{ap} [y]$ if $x \rightsquigarrow_{ap} y$. A maximal ap-basic class will be called a ap-quasi-attractor.

The ap-chain recurrent points and the ap-basic classes have some interesting properties which will be of use when proving the result on the support of the limit measure $\mu$. We enumerate and prove some of them.

**Proposition 3.4.11 :**
*Let $x \in \mathcal{R}_{ap}$. Then $\overline{[x]_{ap}} \subset [x]_{ap} \cup \partial\Delta$. Moreover, for all $t > 0$ $[x]_{ap}$ is $\varphi_t$ invariant.*

**Proof :**

Let $y \in \overline{[x]_{ap}}$ and let $(y_k)_{k\in\mathbb{N}}$ be a sequence of elements of $[x]_{ap}$ converging to $y$. Suppose $y \notin \partial\Delta$.

Let $\delta > 0$ and $T > 0$ and let $k$ such that $d(y_k, y) < \delta$.

There exists a $\delta, T$ pseudo-orbit $x, (x_1, t_1), \cdots, (x_n, t_n), y_k$ linking $x$ to $y_k$. Thus $x, (x_1, t_1), \cdots, (x_n, t_n), y$ is a $2\delta, T$ pseudo-orbit linking $x$ to $y$.

Hence $x \rightsquigarrow_{ap} y$.

A similar reasoning on $\delta, T$-pseudo-orbit linking $y_k$ to $x$ gives us $y \rightsquigarrow_{ap} x$ and then $y \in [x]_{ap}$.

We show now that $[x]_{ap}$ is an invariant set.

Let $T, T', \varepsilon > 0$. $\varphi_{T'}$ is an uniformly continuous application. Let then $\delta < \varepsilon$ such that $d(x, y) < \delta$ implies $d(\varphi_{T'}(x), \varphi_{T'}(y)) < \varepsilon$.

We know that there exists a $\delta, T$ pseudo-orbit $x, (x_1, t_1), \cdots, (x_n, t_n), x$ linking $x$ to $x$. Hence $x, (x_1, t_1), \cdots, (x_n, t_n + T'), \varphi_{T'}(x)$ is a $\varepsilon, T$ pseudo-orbit linking $x$ to $\varphi_{T'}(x)$.

Before proving the converse, let's just remark that, if $\delta_1 < \delta_2$ and $T_1 > T_2$ then every $\delta_1, T_1$ pseudo-orbit is also a $\delta_2, T_2$ pseudo-orbit.

Let's now suppose that $T > T'$

We consider again our $\delta, T$ pseudo-orbit $x, (x_1, t_1), \cdots, (x_n, t_n), x$ linking $x$ to $x$. One can remark that $t_1 + t_2 - T' > T$.

Thus $\varphi_{T'}(x), (\varphi_{T'}(x_1), t_1+t_2-T), (x_3, t_3), \cdots, (x_n, t_n), x$ is a $\varepsilon, T$ pseudo-orbit linking $\varphi_{T'}(x)$ to $x$ if $\delta$ is chosen small enough.

Then $x \rightsquigarrow_{ap} \varphi_{T'}(x)$.

By composing the $\delta, T$ pseudo-orbits linking $x$ and $\varphi_{T'}(x)$ by $\varphi_{-T'}$ we get that $x \sim_{ap} \varphi_T(x)$ for every $T$ in $\mathbb{R}$.

■

**Proposition 3.4.12 :**
*Let $x \in \Delta$. If $x \in \partial\Delta$ or $\omega(x) \subset \mathring{\Delta}$ then $\omega(x) \subset \mathcal{R}_{ap}$. From this we get that, for every $x$ in $\Delta$, $\omega(x) \cap \mathcal{R}_{ap} \neq \emptyset$.*

**<u>Proof :</u>**

The first point is a well-known result for "classic" chain-recurrence and easily extended to ap-chain-recurrence. Furthermore, if $x \in \mathring{\Delta}$ and $\omega(x) \cap \partial\Delta \neq \emptyset$ then, by taking $y \in \omega(x) \cap \partial\Delta$, we get $\omega(y) \subset \omega(x)$ and $\omega(y) \subset \mathcal{R}_{ap}$.

■

**Proposition 3.4.13 :**
*If $[x]_{ap}$ is maximal, then $x \rightsquigarrow_{ap} z$ if and only if $z \in [x]$. As a consequence we also get that every quasi-attractor is a closed set.*

**<u>Proof :</u>**

This result is trivial as soon as $z \in \mathcal{R}_{ap}$, thus we only have to prove it for $z \notin \mathcal{R}_{ap}$. We know that $\omega(z) \cap \mathcal{R}_{ap} \neq \emptyset$ and that, if $u \in \omega(z) \cap \mathcal{R}_{ap}$ then $z \rightsquigarrow_{ap} u$ thus $x \rightsquigarrow_{ap} u$. Hence $u \in [x]_{ap}$, from that we get $x \rightsquigarrow_{ap} z \rightsquigarrow_{ap} u \rightsquigarrow_{ap} x$ which implies $z \in [x]_{ap} \subset \mathcal{R}_{ap}$.

■

The relation between being an attractor and being a quasi-attractor has been studied in the past, we recall this theorem of [1].

**Proposition 3.4.14 :**
*Let $C$ be a non-empty subset of $\Delta$. The following assertions are equivalent :*

*(i) $C$ is an irreducible attractor i.e. it doesn't contain any proper attractor.*

*(ii) $C$ is an isolated quasi-attractor, i.e. there exists $U$, an open neighborhood of $C$, such that $U \cap \mathcal{R}_{ap} = C$.*

*(iii) $C$ is an isolated connected component of $\mathcal{R}_{ap}$ and*

$$C^+ = \{x \in M \,;\, \exists y \in \mathcal{R}_{ap} \setminus C \text{ s.t. } C \rightsquigarrow_{ap} y \rightsquigarrow_{ap} x\} = \emptyset$$

This result is proved in Part 5 of [1].

The following hypothesis is an analog of Hypothesis 3.4.1 adapted to the context of ap-pseudo-orbits.

**Hypothesis 3.4.15 :**
*There is only a finite number of ap-basic classes in* $\mathring{\Delta}$ *denoted by* $K_i$ , $i = 1 \cdots \eta$*. We suppose that they are closed sets and indexed in such a way that the* $k$ *first* $\{K_i\}_{i=1\cdots k}$ *are the quasi-attractors and the* $\eta - k$ *others aren't.*

**Proposition 3.4.16 :**
*For every* $\theta$ *small enough, there exist* $\delta(\theta) \in [0, \theta[$ *and* $T(\theta) \in ]0, \infty]$ *with* $\delta(\theta), T(\theta) \neq (0, \infty)$ *such that:*

*If there exists a* $\delta, T$ *pseudo-orbit* $\xi_0 \cdots \xi_n$ *verifying,* $\delta < \delta(\theta)$ *or* $T > T(\theta)$ *and, for a certain triplet* $(i, i', j) \in \{1 \cdots \eta\}^3$,

$$d(\xi_0, K_i) < \delta \quad d(\xi_n, K_{i'}) < \delta \quad d(\xi_j, K_i) > \theta$$

*Then* $i \neq i'$ *and* $K_i \prec K_{i'}$*.*

This proposition means that, for $\delta$ small and $T$ large, the $\delta, T$ pseudo-orbits respect the partial order.

**Proof :**

Suppose that, for every $\delta > 0$ and every $T > 0$, there exists a $\delta, T$-pseudo-orbit $\xi_0 \cdots \xi_n$ such that

$$d(\xi_0, K_i) < \delta \quad d(\xi_n, K_{i'}) < \delta$$

then, we can construct $\delta, T$-pseudo-orbits going from $K_i$ to $K_{i'}$ which in turn implies $K_i \rightsquigarrow_{ap} K_{i'}$

Hence, if $K_i \not\rightsquigarrow_{ap} K_{i'}$, there exists $\tilde{\delta} \geqslant 0$ and $\tilde{T} \in ]0, +\infty]$ such that a $\delta, T$ pseudo-orbit $\xi_0 \cdots \xi_n$ verifying

$$d(\xi_0, K_i) < \delta \quad d(\xi_n, K_{i'}) < \delta$$

may only exist if $\delta > \tilde{\delta}$ or $T < \tilde{T}$.

Suppose now that $i = i'$, we will show that there exists $\hat{\delta}$ and $\hat{T}$ such that every $\delta, T$ pseudo-orbit $\xi_0 \cdots \xi_n$ verifying either $\delta < \hat{\delta}$ or $T > \hat{T}$ and

$$d(\xi_0, K_i) < \delta \quad d(\xi_n, K_i) < \delta$$

doesn't contain any point at a distance greater than $\theta$ from $K_i$.

Suppose first that this assertion is false, thus we have real sequences $\delta_l \to 0$ and $T_l \to \infty$ and a sequence of $\delta_l, T_l$ pseudo-orbits verifying

$$d(\xi_0^l, K_i) < \delta \quad d(\xi_{n_l}^l, K_i) < \delta \quad d(\xi_{j_l}^l, K_i) > \theta.$$

$\Delta$ being a compact set, we may suppose that, up to an extraction, as $l$ goes to infinity $x_0^l \to x \in K_i$, $x_{n_l}^l \to z \in K_i$ and $x_{j_l}^l \to y$ with $d(y, K_i) \geqslant \theta$. In that case we get $x \rightsquigarrow_{ap} y \rightsquigarrow_{ap} z$ and thus $y \in K_i$, which is absurd.

■

**Proposition 3.4.17 :**
*For every $\delta > 0$, there exists $T_0 > 0$ such that every $\delta, T_0$ pseudo-orbit intersects $N^\delta(\mathcal{R}_{ap})$.*

**Proof :**

Let $x \in \Delta$ and $\gamma > 0$, we define $T^\gamma(x) = Inf\{t \geqslant 0 \,;\, \varphi_t(x) \in N^\gamma(\mathcal{R}_{ap})\}$

As $\omega(x) \cap \mathcal{R}_{ap} \neq \emptyset$ we get $T^\gamma(x) < +\infty$.

For $\alpha > 0$ we will denote $N_\alpha = \{x \in \Delta \,;\, T^\gamma(x) \geqslant \alpha\}$ the level sets of $T^\gamma$

Let us show that $N_\alpha$ is closed. Let $(x_n)_{n\in\mathbb{N}}$ be a sequence of elements of $N_\alpha$ converging to $y$. By the continuity of $\varphi_t$ we obtain then that, for every $t > 0$, $\lim_{n\to\infty} \varphi_t(x_n) = \varphi_t(y)$. If $t < \alpha$ we get $\varphi_t(x_n) \in (N^\gamma(\mathcal{R}_{ap}))^c$ which is closed. Hence we have $\varphi_t(y) \notin N^\gamma(\mathcal{R}_{ap})$ for all $t < \alpha$ and thus $y \in N_\alpha$.

$T^\gamma(x)$ is then an upper semi-continuous function taking its values in $[0, +\infty[$, $\Delta$ being a compact set, we know then that $T^\gamma(x)$ attains its maximum on $\Delta$ which we will denote $T^\gamma$.

Taking $T_0 > T^\delta$ gives us the result.

■

The following corollary comes easily from the last two propositions.

**Corollary 3.4.18 :**
*For every $\delta > 0$ and every $T > 0$ there exists a family $V_i$ of open neighborhoods of the $K_i$ and positive real numbers $\delta_1$ and $T_1$ such that*

- $\overline{N^{\delta_1}(K_i)} \subset V_i$ *for* $i = 1 \cdots n$
- *Every $\delta_1, T_1$ pseudo-orbit starting from $V_i$ stays in $V_i$ for $i = 1 \cdots k$*
- *If there exists a $\delta_1, T_1$ pseudo-orbit $x, (\xi_1, t_1), \cdots, (\xi_p, t_p), y$ with $x \in N^{\delta_1}(K_i)$ and $y \in N^{\delta_1}(K_j)$ such that*

  $$\exists l \quad \exists \tilde{t} \in [0, t_l] \text{ such that } \varphi_{\tilde{t}}(\xi_l) \notin V_i$$

  *Then $i \neq j$ and $K_i \prec K_j$.*
- *Every $\delta_1, T_1$ pseudo-orbit intersects $N^\delta(\mathcal{R}_{ap})$*

<u>*Remark :*</u> The same conclusions still hold by replacing $\delta_1$ by any $\delta \leqslant \delta_1$

**Definition 3.4.19:**
For $K$ compact subset of $\mathring{\Delta}$ we denote

$$\beta_{\delta,K}(N) = \sup_{x\in\Delta_N\cap K} \mathbb{P}_x[\hat{X}^N(1) \in \Delta \setminus N^\delta(\varphi_1(x))]$$

**Proposition 3.4.20 :**
*If the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$, then, for all $K$ compact subset of $\mathring{\Delta}$ and*

*neighborhood of $A$, there exists $\delta > 0$ such that $\rho_N^N \geqslant 1 - \beta_{\delta,K}(N)$. Moreover, if there exists $U_K$ an open neighborhood of $\partial\Delta$ such that*

$$\lim_{N\to\infty} \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K\cap\Delta_N} \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta]} = 0$$

*Then, for every limiting measure $\mu$, we have $\mu(U_K) = 0$.*

**Proof :**

Let $U$ be an open neighborhood of $A$ and $\delta > 0$ such that $U \subset K$ and for every $t$, $\varphi_t(\overline{U}) \subset U$ and $N^\delta(\varphi_1(\overline{U})) \subset U$.

Then

$$\begin{aligned}
\rho_N^N \mu^N(U) &= \sum_{x\in\Delta_N} p_N^N(x,U)\mu^N(x) \\
&\geqslant \sum_{x\in U\cap\Delta_N} \inf_{x\in U\cap\Delta_N} p_N^N(x,U)\mu^N(x) \\
&\geqslant \mu^N(U)\left(1 - \sup_{x\in U\cap\Delta_N} p_N^N(x,U^c)\right) \\
&\geqslant \mu^N(U)\left(1 - \sup_{x\in U\cap\Delta_N} p_N^N(x,(N^\delta(\varphi_1(\overline{U})))^c)\right) \\
&\geqslant \mu^N(U)\left(1 - \beta_{\delta,K}(N)\right)
\end{aligned}$$

We finally get $\rho_N^N \geqslant 1 - \beta_{\delta,K}(N)$

From this we obtain

$$\begin{aligned}
1 - \beta_{\delta,K}(N) &\leqslant \rho_N^N \mu^N(\mathring{\Delta}) \\
&\leqslant \sum_{x\in\mathring{\Delta}_N} \left(1 - p_N^N(x,\partial\Delta)\right)\mu^N(x) \\
&\leqslant \mu^N(\Delta\setminus U_K) + \mu^N(U_K)\left(1 - \inf_{x\in U_K\cap\Delta_N} p_N^N(x,\partial\Delta)\right)
\end{aligned}$$

Hence

$$\mu^N(U_K) \leqslant \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K\cap\Delta_N} p_N^N(x,\partial\Delta)}$$

$U_K$ being an open set, the weak convergence of the measures $\mu^N$ gives us the desired result.

■

**Hypothesis 3.4.21 :**
*We suppose that the flow $\varphi_t$ admits an attractor $A \subset \mathring{\Delta}$ and that there exists $K$ a compact neighborhood of $A$ in $\mathring{\Delta}$ and $U_K$ an open neighborhood of $\partial\Delta$ such that*

$$\lim_{N\to\infty} \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K\cap\Delta_N} \mathbb{P}_x[\hat{X}^N(1)\in\partial\Delta]} = 0$$

The following assumption is a technical one but we will see later that it is true under the first set of hypotheses.

**Hypothesis 3.4.22 :**
*Let $j \in \{k+1,\cdots,\nu\}$, i.e. such that $K_j$ is not a quasi-attractor. Then*

$$\exists\beta_0 > 0 \quad \forall\beta < \beta_0 \quad \forall\gamma > 0 \quad \exists N_0 \quad \exists\zeta_{\gamma,\beta} : \mathbb{N} \to \mathbb{R}$$

*such that*

$$\lim_{n\to\infty} \zeta_{\gamma,\beta}(n) = 0$$

*and*

$$\forall N \geqslant N_0 \qquad \sup_{x\in N^\beta(K_j)} \mathbb{P}_x\left[t\hat{a}u^N_{N^\beta(K_j)} > e^{N\gamma}\right] \leqslant \zeta_{\gamma,\beta}(N)$$

*where $\hat{\tau}^N_U = Inf\{t \geqslant 0\,;\ \hat{X}^N(t) \notin U\}$.*

We now arrive at the central theorem of this section.

**Theorem 3.4.23 :**
*Under Hypotheses 3.4.15, 3.4.21 and 3.4.22, the limiting measure $\mu$ has its support inside the union of ap-quasi-attractors $K = \bigcup_{i=1}^{k} K_i$*

**<u>Proof :</u>**

Let $\delta > 0$ such that

$$\delta < \delta_1 \qquad \forall 1 \leqslant i \leqslant \eta \quad N^{2\delta}(K_i) \subset V_i$$

Let $(u_N)_{N\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ and $(v_N)_{N\in\mathbb{N}} \in \mathbb{N}^{\mathbb{N}}$ be two sequences verifying the following assumptions:

$$u_N e^{-Min(\delta,\gamma)N} \to 0 \quad \frac{v_N}{u_N} \to 0 \quad \exists\alpha > 0 \ \ s.t.\ v_N \geqslant e^{\alpha N}$$

where $\gamma$ is given by Theorem 3.3.3.

$K$ being an attractor, we know that $\rho_N \to 1$, that $\mu$ is $\varphi$-invariant and has its support inside $\mathcal{R}_{ap}$.

Hypothesis 3.4.21 and Proposition 3.4.20 gives us that $\mu(\partial\Delta) = 0$, hence $\mu(K) = 1$.

It only remains to be shown that, for every $j = k+1\cdots\eta$, there exists an open neighborhood $W_j$ of $K_j$ such that $\mu(W_j) = 0$

We know that

$$\rho_N^{Nu_N}\mu^N(W_j) = \sum_{x\in\Delta_N} \mathbb{P}_x[\hat{X}^N(u_N)\in W_j] = \int_{\Delta_N} \mathbb{P}_x[\hat{X}^N(u_N)\in W_j]d\mu^N(x)$$

We denote $t_N^i = \left[\frac{u_N}{i}\right]$ and define the following events :

$$E_N^\delta = \left\{\text{There exists a } \delta, T_1 \text{ pseudo-orbit closer than } \delta \text{ from } \hat{X}^N(t),\ t\in[0,u_N]\right\}$$

$$E'_N = \left\{\forall i\in\{k+1,\cdots,\eta\}\ , \forall q\geqslant v_N\ \hat{X}^N(p)\in N^{\delta_1}(K_i) \Rightarrow \hat{X}^N(p+q)\notin N^{\delta_1}(K_i)\right\}$$

$E'_N$ is the event "after its first entry in $N^{\delta_1}(K_i)$ the Markov chain $\hat{X}$ will have left it after $v_N$ steps".

Let $W_i$ be open neighborhoods of the $K_i$ such that $N^\delta(W_i)\subset V_i$ for every $i = 1\cdots\eta$. On $E_N^\delta\cap E'_N$ we get,

For $j\geqslant k$ and $N$ large enough (in particular $v_N\leqslant u_N/\eta$ ),

$$\hat{X}^N(u_N)\in W_j \Rightarrow \exists i\in\{2,\cdots\eta\}\quad \hat{X}^N(t_N^i)\notin N^\delta(K)$$

Indeed, should $X^N$ have entered $N^\delta(K)$, then, for some $i$, the pseudo orbit given by $E_N^\delta$ would have entered $N^{2\delta}(K_i)\subset V_i$ (if $\delta$ is chosen small enough). If $i\leqslant k$ we know, by Corollary 3.4.18, that the pseudo orbit would have stayed in $V_i$ and thus couldn't be in $N^\delta(W_j)$ at time $u_N$. Conversely, if $i>k$ then, being on the set $E'_N$ gives us that the chain can't stay for a time greater than $v_n$ in the $\delta$-neighborhood of a non-quasi-attractor. The times $t_N^i$ are large enough that $\hat{X}^N(t_N^i)\in N^\delta(K_l)\Rightarrow \hat{X}^N(t_N^{i+1})\notin N^\delta(K_l)$, as there are only $\eta-k$ non-quasi-attractor, the chain cannot visit $\eta-1$ different non-quasi-attractors before time $u_N$ and be in another different non-quasi-attractor at time $u_N$.

Hence

$$\mathbb{P}[\hat{X}^N(u_N)\in W_j|E_N^\delta\cap E'_N]\leqslant \mathbb{P}\left[\bigcup_{i=2}^{\eta}\{\hat{X}^N(t_N^i)\notin N^\delta(K)\}\right]$$

From Corollary 3.4.18, we can infer that, on $E_N^\delta$, starting from $N^\delta(K_i)$ the chain cannot enter $N^\delta(K_i)$ once it has left $V$ (and also $N^\delta(W_i)$).

Thus

$$\begin{aligned}\mathbb{P}[(E'_N)^c|E_N^\delta] &\leqslant \mathbb{P}[\exists i \text{ such that } \hat{X} \text{ doesn't leave } V_i \text{ before the time } v_N]\\ &\leqslant \sum_{i=k+1}^{\eta}\sup_{x\in N^\delta(K_i)}\mathbb{P}_x[\hat{\tau}_{V_i}^N\geqslant v_N]\\ &\leqslant \sum_{i=k+1}^{\eta}\sup_{x\in N^\delta(K_i)}\mathbb{P}_x[\hat{\tau}_{W_i}^N\geqslant v_N]\end{aligned}$$

where $\hat{\tau}_U^N = Inf\{t\geqslant 0\ ;\ \hat{X}^N(t)\notin U\}$.

These terms will be controlled by Hypothesis 3.4.22 where $\gamma$ will taken as equal to $\alpha$.

We now remark that, if $A, B$ and $C$ are three events, we get

$$\begin{aligned}\mathbb{P}[C] &= \mathbb{P}[C \cap B^c \cap A] + \mathbb{P}[C \cap B \cap A] + \mathbb{P}[C \cap A^c] \\ &\leqslant \mathbb{P}[B^c \cap A] + \mathbb{P}[C|B \cap A] + \mathbb{P}[A^c] \\ &\leqslant \mathbb{P}[B^c|A] + \mathbb{P}[C|B \cap A] + \mathbb{P}[A^c]\end{aligned}$$

From that, we obtain

$$\mathbb{P}_x[\hat{X}^N(u_N) \in W_j] \leqslant \mathbb{P}_x[(E_N^\delta)^c] + \mathbb{P}_x[(E'_N)^c|E_N^\delta] + \mathbb{P}_x[\hat{X}^N(u_N) \in W_j|E_N^\delta \cap E'_N]$$

It only remains to control $\mathbb{P}_x[(E_N^\delta)^c]$. In order to do that we will consider the pseudo-orbit $PO_N(t)$ defined by $x, (x, T_1), (\hat{X}^N(T_1), T_1) \cdots$

For $N$ large enough, classic results on stochastic approximation algorithms (see e.g. [3]) give us the following estimate :

$$\begin{aligned}\mathbb{P}_x[(E_N^\delta)^c] &\leqslant \mathbb{P}[PO_N(t) \text{ isn't a } \delta_1, T_1 \text{ pseudo-orbit}] \\ &\quad + \mathbb{P}_x[\sup_{t\in[0,N]} d(\hat{X}^N(t), PO_N(t)) > \delta] \\ &\leqslant O\left(u_N e^{\delta N}\right)\end{aligned}$$

In conclusion we get

$$\begin{aligned}\mu^N(V_j) &= \frac{1}{\rho_N^{Nu_N}} \int_{\mathring{\Delta}} \mathbb{P}_x[\hat{X}^N(u_N) \in W_j] d\mu^N(x) \\ &\leqslant \frac{1}{\rho_N^{Nu_N}} \int_{\mathring{\Delta}} \mathbb{P}_x[(E_N^\delta)^c] + \mathbb{P}_x[(E'_N)^c|E_N^\delta] d\mu^N(x) \\ &\qquad + \frac{1}{\rho_N^{Nu_N}} \int_{\mathring{\Delta}} \mathbb{P}_x[\hat{X}^N(u_N) \in W_j|E_N^\delta \cap E'_N] d\mu^N(x) \\ &\leqslant \frac{1}{\rho_N^{Nu_N}} \left(O\left(u_N e^{-Min(\delta,\delta_1} N\right) + \zeta_{\beta,\delta}(N) + \sum_{i=1}^{b} \int_{\mathring{\Delta}} \mathbb{P}_x[\hat{X}^N(t_N^i) \notin N^\delta(K)] d\mu^N(x)\right) \\ &\leqslant \frac{1}{\rho_N^{Nu_N}} \left(O\left(u_N e^{\delta N}\right) + \zeta_{\beta,\delta}(N) + \sum_{i=1}^{b} \rho_N^{Nt_N^i} \mu^N((N^\delta(K))^c)\right) \\ &\leqslant \frac{1}{\rho_N^{Nu_N}} \left(O\left(u_N e^{\delta N}\right) + \zeta_{\beta,\delta}(N) + b\mu^N((N^\delta(K))^c)\right)\end{aligned}$$

From Theorem 3.3.3, we can infer that $\frac{1}{\rho_N^{Nu_N}} \to 1$.

Hence we have $\lim_{N\to\infty} \mu^N(V_j) = 0$. As the sets $V_j$ are open neighborhoods of the sets $K_j$ we obtain $\mu(K_j) = 0$.

■

### 3.4.4 Going back to Hypotheses 3.4.1, 3.4.2 and 3.4.7

The aim of this section is to prove that our first set of hypotheses implies the second one, thus proving the announced Theorem 3.4.9. In particular we will show that $\mathcal{L}$-quasi-attractors and ap-quasi attractors are the same. In this section we will assume Hypotheses 3.4.1, 3.4.2 and 3.4.7 to be true.

**Proposition 3.4.24 :**
*For every compact set $K \subset V_\alpha$, there exists a open neighborhood $U_K$ of $\partial\Delta$ such that*

$$\lim_{N\to\infty} \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K\cap\Delta_N} \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta]} = 0$$

*where* $\beta_{\delta,K}(N) = \sup_{x\in K\cap\Delta_N} \mathbb{P}_x[\hat{X}^N(1) \notin N^\delta(\varphi_1(x))]$

**Proof :** Let $K$ be a compact set in $V_\alpha$ and let $\delta_0 > 0$ such that

$$\delta_0 < \inf_{t\in[0,1],x\in K} d(\varphi_t(x), \partial\Delta)$$

Let $c(K) = \frac{1}{4} Inf\{S(x,1,\phi) | x \in K, d(\phi, \varphi(x)) \geqslant \delta_0\} > 0$.
For $x \in K$ we have

$$\begin{aligned}\limsup_{N\to\infty} \frac{1}{N} log\mathbb{P}_x &\left[\hat{X}^N \in \{\phi \in \mathcal{C}_x([0,1]); d(\phi, \varphi(x)) > \delta_0\}\right] \\ &\leqslant - \inf_{\phi; d(\phi,\varphi(x))>\delta_0} S(x,1,\phi) \\ &\leqslant -4c(K)\end{aligned}$$

Thus, for $N$ large enough and for every $x$ in $K$.

$$\frac{1}{N} log\mathbb{P}_x \left[\hat{X}^N \in \{\phi \in \mathcal{C}_x([0,1]); d(\phi, \varphi(x)) > \delta_0\}\right] \leqslant -3c(K)$$

Moreover, there exists $U_K$, an open neighborhood of $\partial\Delta$ such that

$$\lim_{N\to\infty} \inf_{x\in U_K} \frac{1}{N} log\mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] \geqslant -2c(K)$$

Then, for $N$ large enough, we have, for every $x$ in $U_K$

$$\mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta] \geqslant e^{-\frac{2c(K)}{1/N}}$$

Finally we get

$$\lim_{N\to\infty} \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K} \mathbb{P}_x[\hat{X}^N(1) \in \partial\Delta]} \leqslant \lim_{N\to\infty} e^{-\frac{c(K)}{1/N}} = 0$$

■

**Proposition 3.4.25 :**
*Let $T > 0$ and $K$ be a compact subset of $V_\alpha$. Then, for every $\delta > 0$, there exists $\varepsilon(K,T,\delta) > 0$ such that*

$$\forall x \in K \quad S(x,T,\phi) \leqslant \varepsilon \Rightarrow d(\phi, \varphi_.(x))_{[0,T]} \leqslant \delta$$

**Proof :** Let's suppose that this result is false, then

$$\exists \delta > \quad \forall \varepsilon > 0 \quad \exists x \quad \exists \phi_\varepsilon \text{ s.t. } S(x,T,\phi_\varepsilon) < \varepsilon \text{ and } d(\varphi_.(x), \phi_\varepsilon) > \delta$$

We also know that $c = \inf\{S(x,T,\phi)\,; x \in K\,;\ d(\phi, \varphi_.(x)) > \delta\} > 0$
Thus, for every $\varepsilon > 0$

$$\varepsilon > S(X,T,\phi_\varepsilon) \geqslant c > 0$$

which is absurd.

■

**Proposition 3.4.26 :**
*The function $B_{\mathcal{L}} : V_\alpha \times V_\alpha \to \mathbb{R}_+$ is upper semi-continuous.*
**Proof :** Let $x_n$ and $y_n$ two sequences in $V_\alpha$ converging to $x$ and $y$ in $V_\alpha$
For every $\delta > 0$ there exists a path $x = \xi_0, \cdots \xi_{n(\xi)} = y$ such that

$$B_{\mathcal{L}}(x,y) \leqslant A_{n(\xi)}(\xi) \leqslant B_{\mathcal{L}}(x,y) + \delta$$

Let's now consider the path $\xi^n$ given by $x_n = \xi_0^N, \xi_1, \cdots \xi_{n(\xi)-1}, y_n = \xi_{n(\xi)}$
If we take $n$ such that

$$|\mathcal{L}(x_n, \xi_1) - \mathcal{L}(x, \xi_1)| < \delta \text{ and } |\mathcal{L}(\xi_{n(\xi)-1}, y_n) - \mathcal{L}(\xi_{n(\xi)-1}, y)| < \delta$$

We obtain

$$B_{\mathcal{L}}(x_n, y_n) \leqslant A_{n(\xi)}(\xi^N) \leqslant B_{\mathcal{L}}(x,y) + 3\delta$$

Thus, for every $\delta > 0$, $\limsup\limits_{n\to\infty} B_{\mathcal{L}}(x_n, y_n) \leqslant B_{\mathcal{L}}(x,y) + 3\delta$
Finally $\limsup\limits_{n\to\infty} B_{\mathcal{L}}(x_n, y_n) \leqslant B_{\mathcal{L}}(x,y)$

■

**Proposition 3.4.27 :**
*Let $x \in \mathcal{R}_{\mathcal{L}}$ such that $[x]_{\mathcal{L}}$ is a closed subset of $V_\alpha$. Let $\theta > 0$ such that $N^\theta([x]_{\mathcal{L}}) \subset V_\alpha$. Then,*

$$\exists \delta > 0 \quad \exists T > 0 \quad \forall \Psi \textit{ with } \Psi(0) \in [x]_{\mathcal{L}}, \quad \Psi(\tilde{T}) \in [x]_{\mathcal{L}}, \quad \tilde{T} > T,$$

$$\textit{and } S(\Psi(0), \tilde{T}, \Psi) < \delta \quad \forall t \in [0, \tilde{T}] \qquad \Psi(t) \in N^\theta([x]_{\mathcal{L}}).$$

**Proof :** Let $x \in \mathcal{R}_\mathcal{L}$ such that $[x]_\mathcal{L}$ is a closed subset of $V_\alpha$ and let $\theta > 0$ such that $N^\theta([x]_\mathcal{L}) \subset V_\alpha$.

Suppose that the announced result is false, then, there exists a family of functions $\Psi_n \in \mathcal{C}([0, T_n], V_\alpha)$ such that

$$\Psi_n(0) \in [x]_\mathcal{L} \quad \Psi_n(T_n) \in [x]_\mathcal{L} \quad \lim_{n\to\infty} T_n = \infty \quad S(\Psi(0), T_n, \Psi) < 1/n$$

and

$$\forall n \quad \exists t_n > 0 \text{ such that } \Psi_n(t_n) \notin N^\theta([x]_\mathcal{L}).$$

$[x]_\mathcal{L}$ being a compact set, we can assume without loss of generality that $\Psi_n(0) \to u \in [x]_\mathcal{L}$ and $\Psi_n(T_n) \to v \in [x]_\mathcal{L}$.

Let $\tau_n = \inf\{t > 0\,;\, \Psi_n(t) \notin N^\theta([x]_\mathcal{L})\}$. Then we have

$$\Psi_n(\tau_n) \in \overline{V_\alpha} \setminus N^\theta([x]_\mathcal{L})$$

and $\overline{V_\alpha} \setminus N^\theta([x]_\mathcal{L})$ is a compact set.

Thus, without loss of generality, we can assume that

$$\Psi_n(\tau_n) \to w \in V_\alpha \setminus N^\theta([x]_\mathcal{L}).$$

Using the sequential continuity of $\mathcal{L}$ and the fact that $u$ and $v$ are $\mathcal{L}$-chain-recurrent points, we get that $u \rightsquigarrow_\mathcal{L} w$ and $w \rightsquigarrow_\mathcal{L} v$. Thus $w \in [x]_\mathcal{L}$ and we get a contradiction

■

**Corollary 3.4.28 :**
*Suppose $\mathcal{R}_\mathcal{L}$ is a closed subset of $V_\alpha$. Let $\theta > 0$ such that $N^\theta(\mathcal{R}_\mathcal{L}) \subset V_\alpha$ Then,*

$$\exists \delta > 0 \quad \exists T > 0 \quad \forall \Psi \text{ with } \Psi(0) \in \mathcal{R}_\mathcal{L}, \quad \Psi(\tilde{T}) \in \mathcal{R}_\mathcal{L}, \quad \tilde{T} > T$$

$$\text{and } S(\Psi(0), \tilde{T}, \Psi) < \delta \quad \forall t \in [0, \tilde{T}] \qquad \Psi(t) \in N^\theta(\mathcal{R}_\mathcal{L}).$$

**Proposition 3.4.29 :**
*Let $x \in \mathcal{R}_\mathcal{L}$ such that $[x]_\mathcal{L}$ is a closed subset of $V_\alpha$. Then $x \in \mathcal{R}_{ap}$ and $[x]_\mathcal{L} \subset [x]_{ap}$.*

**Proof :** Let $y \in [x]_\mathcal{L}$. Let $(T_n)_{n\in\mathbb{N}}$ be a sequence of positive real numbers and $\Psi_n$ be a sequence of functions from $[0, T_n]$ to $V_\alpha$ such that $\Psi_n(0) = x$, $\Psi_n(T_n) = y$ and $S(x, T_n, \Psi_n) < 1/n$.

$T_n$ is either a bounded sequence or it goes to infinity (up to a sub-sequence).

Let's suppose that $(T_n)_{n\in\mathbb{N}}$ is an increasing sequence of positive real numbers going to infinity and $\Psi_n$ a sequence of functions such that $\Psi_n(0) = x$, $\Psi_n(T_n) = y$ and $S(x, T_n, \Psi_n) < 1/n$

The former proposition gives us a compact set $K \subset V_\alpha$, and positive numbers $\delta$ and $T$ such that, for $T_n > T$ and $1/n < \delta$, $\Psi_n$ lives in $K$

Let $\tilde{T} > T$, $\tilde{\delta} < \delta$ and $\varepsilon = \varepsilon(\tilde{\delta}, \tilde{T}, K)$ given by Proposition 3.4.25.

Then, from Proposition 3.4.25, we can infer that, for $n$ large enough,

$$d(\Psi_n(\tilde{T}), \varphi_{\tilde{T}}(x)) \leqslant \delta$$

As $\Psi_n(\tilde{T}) \in K$, we also have

$$d(\varphi_{\tilde{T}}(\Psi_n(\tilde{T})), \Psi_n(2\tilde{T})) < \delta$$

By iterating this process we get that, for $n$ large enough,

$$(x, \tilde{T}), (\Psi_n(\tilde{T}), \tilde{T}), \cdots$$

is a $\delta, \tilde{T}$ ap-pseudo orbit linking $x$ to $y$.

Hence $x \rightsquigarrow_{ap} y$.

Let's suppose now that $T_n$ is bounded by $\overline{T}$. By taking a sub-sequence we can assume that $T_n \to T$ as $n \to \infty$ for some $T$ with $\overline{T} \geqslant T \geqslant 0$. we continue the function $\Psi_n$ to the time $\overline{T}$ by concatenating the flow $\varphi_{\cdot}$. Then we know that $\Psi_n \to \varphi_{\cdot}(x)$ uniformly over $[0, \overline{T}]$ as $n \to \infty$. Hence $y = \lim_{n\to\infty} \Psi_n(T_n) = \varphi_T(x)$. Thus $T > 0$ and $y \in \gamma^+(x)$.

If $y = x$ we get then that, either $x \rightsquigarrow_{ap} x$ or $x$ is periodic which implies $x \in \mathcal{R}_{ap}$.

Then, if $y \in [x]_{\mathcal{L}}$ we obtain that either $x \rightsquigarrow_{ap} y$ or $y \in \gamma^+(x) \subset [x]_{ap}$. Taking $y = x$ gives us $x \in \mathcal{R}_{ap}$, i.e. $\mathcal{R}_{\mathcal{L}} \subset \mathcal{R}_{ap}$. As the roles of $x$ and $y$ can be exchanged we also get $y \rightsquigarrow_{ap} x$ and in conclusion $[x]_{\mathcal{L}} \subset [x]_{ap}$.

■

**Proposition 3.4.30 :**
*Closed $\mathcal{L}$-classes in $V_\alpha$ are positively invariant sets for the flow $\varphi$*

**Proof :** Let $x \in \mathcal{R}_{\mathcal{L}}$ such that $[x]_{\mathcal{L}}$ is a closed $\mathcal{L}$ basic class in $V_\alpha$. Let $T > 0$ and $\theta$ such that $N^\theta([x]_{\mathcal{L}}) \subset V_\alpha$.

If the paths linking $x$ to itself have bounded length then $x$ is periodic and in this case $\gamma^+(x) \subset [x]_{\mathcal{L}}$.

We now suppose that the paths have unbounded length.

Let $\varepsilon = \varepsilon([x]_{\mathcal{L}}, \tau, \theta)$ and let $\delta > 0$ and $\tau > 0$ given by Proposition 3.4.27, let $\Psi$ be a path of length greater than $\tau$ and of cost smaller than $min(\varepsilon, \delta)$ linking $x$ to itself.

Then, by the triangular inequality we get

$$d(\varphi_T(x), [x]_{\mathcal{L}}) \leqslant d(\varphi_T(x), \Psi(T)) + d(\Psi(T), [x]_{\mathcal{L}}) \leqslant 2\theta$$

Thus, by making $\theta$ go to zero, we obtain $\varphi_T(x) \in \overline{[x]}_{\mathcal{L}} = [x]_{\mathcal{L}}$;

■

**Lemma 3.4.31 :**
*Let $\rightarrowtail$ be a binary, transitive relation on $V_\alpha$, such that, for every $x \in V_\alpha$, every $y \in \alpha(x)$ and every $z \in \omega(x)$,*

$$y \rightarrowtail x \quad x \rightarrowtail z \quad z \rightarrowtail z$$

*Let $[x]_{\rightarrowtail}$ be a maximal, closed and isolated $\rightarrowtail$-basic class.*

*Then $[x]_{\rightarrowtail}$ is an attractor for the flow $\varphi$*

**Proof :** Let $W$ be an isolating open neighborhood of $[x]_{\rightarrowtail}$, i.e. $W$ is an open neighborhood of $[x]_{\rightarrowtail}$ such that the only $\rightarrowtail$-recurrent points of $W$ belong to $[x]_{\rightarrowtail}$.

If $[x]_{\rightarrowtail}$ isn't an attractor then there exists $p \in \partial\overline{W}$ such that

$$\gamma_{-}(p) \subset W \text{ and } \alpha(p) \subset [x]_{\rightarrowtail}$$

Let $y \in \omega(p)$, we have $p \rightarrowtail y$ and $y \rightarrowtail y$.

Similarly let $z \in \alpha(p)$, we have $z \rightarrowtail p$ and thus $z \rightarrowtail y$.

Let's define

$$[x]^{+}_{\rightarrowtail} = \{z \ , \exists y \in \mathcal{R}_{\rightarrowtail} \setminus [x]_{\rightarrowtail} \text{ and } x \rightarrowtail y \rightarrowtail z\}$$

As $[x]_{\rightarrowtail}$ is maximal, we have $[x]^{+}_{\rightarrowtail} = \emptyset$.

Yet we have $x \rightarrowtail y$ and $y \in \mathcal{R}_{\rightarrowtail}$.

Hence $y \in [x]_{\rightarrowtail}$, and, as $x \rightarrowtail p \rightarrowtail y$ we get $p \in [x]_{\rightarrowtail}$ which is absurd.

■

**Proposition 3.4.32 :**
*Let $[x]_{\mathcal{L}}$ be a closed isolated quasi-attractor. Then $[x]_{\mathcal{L}}$ is an attractor.*

**Proposition 3.4.33 :**
*Let $K$ be a compact subset of $V_\alpha$.*

*For $\eta > 0$, $\delta > 0$, $\tilde{T} > 0$ there exists $N_0 > 0$ such that, for all $N > N_0$, all $x \in K$, all $T < \tilde{T}$ and all $\Psi \in \mathcal{C}_x([0,T], K)$*

$$\mathbb{P}_x[d(\Psi, \hat{X}^N) < \eta] \geqslant exp\left(-\frac{S(x,T,\Psi)+\delta}{1/N}\right)$$

**Proof :** Let $\xi^K_{\gamma,T} = sup\{|S(x,T,\Psi) - S(x,T,\phi)| \,;\, x, y \in K \,,\, d(\Psi,\phi) < \gamma\}$. The inferior semi-continuity of $S$ gives us that $\lim_{\gamma\to 0} \xi^K_{\gamma,T} = 0$

Let $\eta > 0$, $\delta > 0$, $\tilde{T} > 0$ and let $\gamma < \eta$ such that $\xi^K_{\gamma,\tilde{T}} < \delta$ and $N^\gamma(K) \subset V_\alpha$

The large deviation principle gives us that, for $\phi \in \mathcal{C}_x([0,T], K)$,

$$\liminf_{N\to\infty} \frac{1}{N} log \inf_{x\in K} \mathbb{P}_x[\hat{X}^N \in N^\gamma_x(\phi)]] \geqslant -\sup_{x\in K} \inf_{\Psi\in N^\gamma(\phi)} S(x,T,\Psi)$$

where $N^\gamma(\phi) = \{\Psi \in \mathcal{C}([0,T],K) \ ; \ \|\Psi - \phi\| < \gamma\}$. Thus, there exists $g_K(N)$ a function which goes to zero as $N$ goes to infinity such that

$$\frac{1}{N} log \inf_{x\in K} \mathbb{P}_x[\hat{X}^N \in N^\gamma_x(\phi)]] \geqslant -\sup_{x\in K} \inf_{\Psi\in N^\gamma(\phi)} S(x,T\Psi) - g_K(N)$$

Hence

$$\begin{aligned}\mathbb{P}_x[d(\Psi,\hat{X}^N)<\eta] \geqslant \quad & \mathbb{P}_x[d(\Psi,\hat{X}^N)<\gamma] \geqslant \inf_{x\in K}\mathbb{P}_x[\hat{X}\in N^\gamma(\Psi)] \\ & \geqslant exp\left(\frac{-\sup_{x\in K}\inf_{\phi\in N^\gamma(\Psi)} S(x,T,\phi)-g_K(N)}{1/N}\right) \\ & \geqslant exp\left(\frac{-\xi^K_{\gamma,T}-S(x,T,\Psi)-g_K(N)}{1/N}\right)\end{aligned}$$

Taking $N$ large enough such that $\xi^K_{\gamma,T}+g_K(N)<\delta$ gives us the result

■

**Proposition 3.4.34 :**
*We denote by $BC(\varphi)=\overline{\{x\in\Delta\,;\,x\in\omega(x)\}}$ the Birkhoff center of the flow $\varphi$. Then*

$$V_\alpha\cap BC(\varphi)\subset\mathcal{R}_\mathcal{L}$$

**Proof :** We know that $BC(\varphi)=\overline{Rec(\varphi)}=\overline{\{x\in\Delta\,;\,x\in\omega(x)\}}$. It's apparent that $Rec(\varphi)\subset\mathcal{R}_\mathcal{L}$. Hypothesis 3.4.7 allow us to conclude.

■

**Corollary 3.4.35 :**
*Let $\mu$ be an invariant measure for the flow $\varphi$ whose support $S$ lies within $V_\alpha$. Then $S\subset\mathcal{R}_\mathcal{L}$.*

We arrive at the main theorem of this section, linking $\mathcal{L}$-chain recurrence with ap-chain recurrence.

**Theorem 3.4.36 :**
*Under the hypotheses 3.4.1, 3.4.2 and 3.4.7 we have*

$$\mathcal{R}_{ap}\cap V_\alpha=\mathcal{R}_\mathcal{L}\cap V_\alpha$$

$$\forall x\in\mathcal{R}_\mathcal{L}\cap V_\alpha \qquad [x]_\mathcal{L}=[x]_{ap}$$

**Proof :** By Proposition 3.4.29 we already know that, if $[x]_\mathcal{L}$ is a closed $\mathcal{L}$-basic class in $\mathring{\Delta}$, then $x\in\mathcal{R}_{ap}$ and $[x]_\mathcal{L}\subset[x]_{ap}$.

The function $B_\mathcal{L}$ is upper semi-continuous. Thus

$$\forall\gamma>0 \quad \exists U^i_\gamma \text{ neighborhood of } K_i \text{ s.t. } \forall(a,b)\in(U^i_\gamma)^2 \quad B_\mathcal{L}(a,b)<\gamma$$

**Lemma 3.4.37 :**
*Let $U$ be a neighborhood of $\mathcal{R}_\mathcal{L}\cap\mathring{\Delta}$ such that $K=\overline{U}$ is a compact subset of $\mathring{\Delta}$. Thus there exists $T>0$ such that, if $\gamma^+(x)\subset K$ then $\{\varphi_t(x)\,;\,t\in[0,T]\}\cap U\neq\emptyset$*

**Proof :** Due to the fact that, if $\gamma_+(x) \subset K$, $\omega(x) \subset U$ we can't have $\gamma^+(x) \subset K \setminus U$. Let's define $\nu(x) = inf\{t\,;\, \varphi_t(x) \notin K \setminus U\} < +\infty$

We will show that $\nu(x)$ is upper semi-continuous.

Let $x \in K$ and $x_n \in K^{\mathbb{N}}$ such that $x_n \to x$.

The continuity of $\varphi$ gives us $d(\varphi_{T_n}(x_n), \varphi_T(x)) \to 0$ with the convention that $\varphi_\infty(x) = \omega(x)$. Let $\tilde{T} = \nu(x)$, there exists a sequence $\varepsilon_m \to 0$ such that $\varphi_{\tilde{T}+\varepsilon_m}(x) \notin K \setminus U$.

Yet we have $\varphi_{\tilde{T}+\varepsilon_m}(x_n) \to \varphi_{\tilde{T}+\varepsilon_m}(x)$. Hence, for $n$ large enough we get $\nu(x_n) \leqslant \tilde{T} + \varepsilon_m$

We now get $\limsup_{n\to\infty} \nu(x_n) \leqslant \tilde{T} + \varepsilon_m$

Then $\limsup_{n\to\infty} \nu(x_n) \leqslant \nu(x)$, i.e. $\nu$ is u.s.c.

Thus $T = \max_{x \in K\setminus U} \nu(x)$ exists and we get, if $\gamma^+(x) \subset K$ then

$$\{\varphi_t(x)\,;\, t \in [0, T+\varepsilon]\} \cap U \neq \emptyset$$

■

Let $x \in \mathcal{R}_{ap} \cap \mathring{\Delta}$ and let $y \in [x]_{ap}$, there exists two sequences of positive real numbers $\delta_k \to 0$, $T_k \to \infty$ and a sequence of $\delta_k, T_k$ ap-pseudo-orbits linking $x$ to $y$ denoted $\Psi^k$.

Lemma 3.4.37 gives us $K \subset \mathring{\Delta}$ a compact neighborhood of $\mathcal{R}_{ap}$ such that, for $k$ large enough, every $\delta_k, T_k$ ap-pseudo-orbit stays in $K$.

We know that $\mathcal{R}_{\mathcal{L}} \subset \mathcal{R}_{ap} \subset K$. Let $\gamma$ such that $\bigcup_i N^\gamma(K_i) \subset K$. Let then $T$ given by the former lemma. For $k$ large enough, we have $T_k > T$ and thus every continuous part of the $\delta_k, T_k$ ap-pseudo-orbit intersect $U_\gamma$.

Let $\varepsilon < \gamma$ and $V = \bigcup_i N^\varepsilon(K_i)$, we assume $\varepsilon$ to be small enough such that the $\varepsilon$-neighborhood of the $K_i$ are disjoint.

We define two sequences of times $(\sigma_i(k))_{i=1\cdots p_k}$ and $(\tau_i(k))_{i=1\cdots q_k}$ by

$$\sigma_0(k) = 0 \qquad \tau_0(k) = min\{t\,;\, \Psi^k_t \notin V\}$$

$$\sigma_{i+1}(k) = min\{t > \tau_i(k)\,;\, \Psi^k_t \in V\} \quad \tau_{i+1}(k) = min\{t > \sigma_{i+1}(k)\,;\, \Psi^k_t \notin V\}$$

By Lemma 3.4.37 we know that $\sigma_{i+1}(k) - \tau_i(k) < 2T$.

If our pseudo-orbit $\psi^k$ enters more than once the same $N^\varepsilon(K_i)$ then we truncate what happens between the first entry and the last exit and we will keep the same name for the new path (which may no more be a pseudo-orbit). This way we get $q_k, p_k < \eta$.

If $y \in \bigcup_i K_i$ then $q_k = p_k + 1$, else $q_k = p_k$.

In the first case we get

$$B_{\mathcal{L}}(x,y) \leqslant \sum_{i=0}^{q_k-1} B_{\mathcal{L}}(\Psi^k_{\sigma_i(k)}, \Psi^k_{\tau_i(k)}) + \sum_{i=0}^{p_k} B_{\mathcal{L}}(\Psi^k_{\tau_i(k)}, \Psi^k_{\sigma_{i+1}(k)}) + B_{\mathcal{L}}(\Psi^k_{\tau_{q_k}}, y)$$

In the second case, we get

$$B_{\mathcal{L}}(x,y) \leqslant \sum_{i=0}^{q_k-1} B_{\mathcal{L}}(\Psi^k_{\sigma_i(k)}, \Psi^k_{\tau_i(k)}) + \sum_{i=0}^{p_k} B_{\mathcal{L}}(\Psi^k_{\tau_i(k)}, \Psi^k_{\sigma_{i+1}(k)}) + B_{\mathcal{L}}(\Psi^k_{\sigma_{p_k}}, y)$$

In either case we obtain

$$B_{\mathcal{L}}(x,y) \leqslant (1+\nu)\alpha + (1+\nu) Sup \left\{ B_{\mathcal{L}}(a,b) \; ; \; d(b, \varphi_{[0,2T]}(a)) < \varepsilon_k \right\}$$

where

$$\varepsilon_k = \sup_{d(x,y)<\delta_k \, , t\in[0,2T]} d(\varphi_t(x), \varphi_t(y))$$

From the continuity of $\varphi$, the positive invariance of $[x]_{ap}$ and Proposition 3.4.8 we infer that

$$\lim_{k\to 0} Sup \left\{ B_{\mathcal{L}}(a,b) \; ; \; d(b, \varphi_{[0,2T]}(a)) < \varepsilon_k \right\} = 0$$

Hence we have, for every $\alpha > 0$, $B_{\mathcal{L}}(x,y) \leqslant (\nu+1)\alpha$.

In conclusion $x \rightsquigarrow_{\mathcal{L}} y$. Similarly $y \rightsquigarrow_{\mathcal{L}} x$, i.e. $y \in [x]_{\mathcal{L}}$

■

*Remark :* This proof gives us a little more, it proves that, if $y \in \mathcal{R}_{ap}$ and $x \rightsquigarrow_{ap} y$, then $x \rightsquigarrow_{\mathcal{L}} y$.

When we take $\alpha$ such that $\mathring{\Delta} \cap \mathcal{R}_{\mathcal{L}} = V_\alpha \cap \mathcal{R}_{\mathcal{L}}$, this proposition proves that the first set of hypotheses implies Hypothesis 3.4.22.

**Proposition 3.4.38 :**
*Let $j \in \{k+1, \cdots, \eta\}$, i.e. such that $K_j$ is not a ap-quasi-attractor. Then*

$$\exists \lambda_0 > 0 \quad \forall \lambda < \lambda_0 \quad \forall \gamma > 0 \quad \exists N_0 \quad \exists \zeta_{\gamma,\lambda} : \mathbb{N} \to \mathbb{R} \textit{ with } \lim_{n\to\infty} \zeta_{\gamma,\lambda}(n) = 0 \textit{ such that}$$

$$\sup_{x\in N^\lambda(K_j)} \mathbb{P}_x \left[ \tau^N_{N^\lambda(K_j)} > e^{\gamma N} \right] \leqslant \zeta_{\gamma,\lambda}$$

**Proof :** $K_j$ isn't a quasi-attractor, thus $\exists \lambda_0 > 0$ such that $\overline{N^{2\lambda_0}(K_j)} \subset \mathring{\Delta}$ and, for all $\gamma > 0$, all $\lambda < \lambda_0$ and all $x \in N^\lambda(K_j)$, there exists $T^\gamma$ and $\Psi^\gamma$ such that

$$\Psi^\gamma(0) = x, \quad \Psi^\gamma(T^\gamma) = y^\gamma \notin N^{2\lambda}(K_j), \quad S(x, T^\gamma, \Psi^\gamma) < \gamma$$

Let then $U = N^{2\lambda}(K_j)$, $V = \omega(\overline{U})$. Let $r > 0$ and let $K$ be a compact subset of $\mathring{\Delta}$ such that $N^r(U) \subset K$.

As $\Psi^\gamma$ starts in $U$ and ends outside of $U$ it must pass through $K \setminus U$. Without loss of generality we may suppose that $\Psi^\gamma$ stays in $K$ and $y^\gamma \in K \setminus U$.

Proposition 3.4.33 says that :

For $\lambda > 0$, $\delta > 0$, $\tilde{T} > 0$ , there exists $N_0(K) > 0$ such that, for all $N > N_0$, all $T < \tilde{T}$, all $x \in K$ and all $\Psi \in \mathcal{C}_x([0,T],K)$

$$\mathbb{P}_x[d(\Psi,\hat{X}^N) < \lambda] \geqslant exp\left(-\frac{S(x,T,\Psi)+\delta}{1/N}\right)$$

Applying this to our case gives us $N_0$ such that, $\forall N > N_0$

$$\mathbb{P}_x[d(\Psi^\gamma,\hat{X}^N) < \lambda] \geqslant exp\left(-\frac{S(x,T^\gamma,\Psi^\gamma)+\delta}{1/N}\right) \geqslant exp\left(-\frac{\gamma+\delta}{1/N}\right)$$

Yet $\{d(\hat{X}^N,\Psi^\gamma) < \lambda\}$ implies that $\hat{X}^N$ leaves $N^\lambda(K_j)$ before the time $T^\gamma$

Hence

$$\forall x \in K \quad \forall N \geqslant N_0(K) \quad \mathbb{P}_x[\tau_{N^\lambda(K_j)} > T^\gamma] < 1 - e^{-\frac{\gamma+\delta}{1/N}}$$

Then

$$\begin{aligned}
\mathbb{P}_x\left[\tau_{N^\eta(K_j)} > e^{\frac{\gamma}{1/N}}\right] &< \left(1 - e^{-\frac{\gamma+\delta}{1/N}}\right)^{\left[\frac{e^{\frac{2\gamma}{1/N}}}{T^\gamma}\right]} \\
&< e^{\left[\frac{e^{\frac{2\gamma}{1/N}}}{T^\gamma}\right] ln\left(1-e^{-\frac{\gamma+\delta}{1/N}}\right)} \\
&< e^{\left[\frac{e^{\frac{2\gamma}{1/N}}}{T^\gamma}\right] -e^{-\frac{\gamma+\delta}{1/N}}} \\
&< e^{-e^{-\frac{\gamma-\delta}{1/N}}\frac{1}{T^\gamma}}
\end{aligned}$$

Taking $\delta < \gamma$ and $\zeta_{\gamma,\lambda} = e^{-e^{-\frac{\gamma-\delta}{1/N}}}$ allows us to conclude.

■

We linked ap-basic classes with $\mathcal{L}$-basic classes but, unless the partial orders on the class are similar, we might not have the same quasi-attractors for both partial orders. This proposition shows that the quasi-attractors are the same.

**Proposition 3.4.39 :**
*Let $[x]$ be a basic-class. $[x]$ is a $\mathcal{L}$-quasi-attractor if and only if $[x]$ is an ap-quasi-attractor.*

**Proof :**

Theorem 3.4.36 already gives us that, if $y \in \mathcal{R}_{ap}$ and $x \rightsquigarrow_{ap} y$, then $x \rightsquigarrow_{\mathcal{L}} y$.

Hence, if $[x]$ is a $\mathcal{L}$-quasi-attractor, it is also an ap-quasi-attractor.

Suppose now that $[x]$ is an ap-quasi-attractor.

Let $y \in \mathcal{R}_{\mathcal{L}}$ such that $x \rightsquigarrow_{\mathcal{L}} y$. If the path linking $x$ to $y$ have bounded length we get $y \in \gamma^+(x)$. As $\mathcal{L}$-basic classes are positive invariant sets we get then $y \in [x]$.

Let us now suppose that the paths linking $x$ to $y$ have unbounded length. Propositions 3.4.28 and 3.4.8 give us a compact set containing the paths linking $x$ to $y$ for $T$ large enough. Using the same technique as in the proof of the proposition 3.4.29 gives us $x \rightsquigarrow_{ap} y$, i.e. $y \in [x]$.

■

## Acknowledgements

The author wishes to thank Michel Benaïm and Mathieu Faure for their advices and proofreading on this work.

# Chapter 4

# Long time behaviour of 1/2 Hölder population diffusion processes

This chapter consists of most of the article ”Long time behaviour of 1/2 Hölder population diffusion processes”.

**Contents**



## 4.1 Introduction

In the past 20 years the issue of the long-term survival of interacting populations has received an ever increasing attention in the field of populations biology. This lead to the introduction of the concepts of persistence and permanence for both deterministic models and stochastic models. In deterministic models, such as differential equations, persistence is often equated with the existence of an attractor bounded away from the extinction states, permanence also called uniform persistence requires that attractor to be global. For the past 30 years there has been an extensive literature on methods for verifying permanence and or persistence. These models provided great insight in the behavior of population models but remained rigid. In order to refine these models and allow for some ”roughness” and/or influence of unpredictable outer events, randomness has been added to these models, leading to models with much more varied behavior and, one might hope, more realistic ones too. However, stochastic models such as stochastic differential equations introduced new difficulties in the notions of persistence and permanence. The requirement that trajectories stay bounded away from the extinction states is too strong as population trajectories in stochastic models can and often will wander arbitrarily close to the extinction states. These models are then said to be stochastically persistent if there is a positive probability to remain away from extinction, see [45] for a review on the subject.

Again these models where there is a positive probability to remain away from extinction give great insight but do not allow to study the whole variety of possible behaviors. When studying finite population stochastic models, the underlying theory of Markov processes shows that extinction in finite time happens almost surely. Yet, in the real world, with large sized pools of population, we don't observe that inevitable extinction. This finite extinction time may then be very large and the system may remain in some sort of ”metastable state” bounded away from extinction for a long time. These mathematical models have been corroborated by biologists who remarked that some interacting populations, while doomed to ultimately settle on an ”extinction state” with some of the species going extinct, seem to settle in some some kind of population equilibrium.

In [20], Faure and Schreiber studied this problem for randomly perturbed discrete time dynamical systems, showing that, under the appropriate assumptions about the random perturbations and that there exists a positive attractor (i.e. an attractor which is bounded away from extinction states) for the unperturbed system, when they exist, quasi-stationary distributions concentrate on the positive attractors of the unperturbed system and that, the expected time to extinction for systems starting according to this quasi-stationary distribution grows exponentially with the system size. In [31] their approach was extended to a class of discrete time Markov process, that, up to a renormalization of time, can be seen as random perturbations of an ordinary differential equation.

The aim of this paper is to obtain similar results as those of [31] for the long time behavior of some diffusion processes and their quasi-stationnary distributions.
In Section 2 we will introduce our setting and give some examples of systems that fall into it. Then, in Section 3, we will show that our stochastic dynamic get almost surely absorbed by the extinction states in finite time, then, under the hypothesis that the deterministic mean dynamic admits an interior attractor, we will give a speed at which the extinction time grows with the size of the system and prove that, when the system size goes to infinity, the limit set of the quasi-stationary distributions of the processes for the weak* convergence consists of invariant measures for the deterministic dynamic. Finally, in Section 4 we will prove some Freidlin-Wentzell type results of our SDE, namely

a weak law of large number and a large deviations principle.

## 4.2 Setting

In [31], we studied a class of discrete time Markov process, that, up to a renormalization of time, can be seen as random perturbations of an ordinary differential equation. A simple yet rich model of such a Markov process is a $(X_k^N)$ the random walk on $\Delta_N = \Delta \cap (1/N\mathbb{Z})^d$ defined by:

$$\mathbb{P}\left[X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)|X_k^n = x\right] = p_{i,j}(x)$$

where $(e_i)_{i=1\cdots d}$ is the canonical base of $\mathbb{R}^d$ and $\Delta$ is the simplex in $\mathbb{R}^d$. This type of model often occurs in population games. In this setting $N$ represents the size of the population. Each individual plays a pure strategy $i$ and $X^N$ represents then the vector of proportion of players of each strategy. The jump $X_{k+1}^N = X_k^N + \frac{1}{N}(e_j - e_i)$ means that an individual switches his strategy from $i$ to $j$ at time $k$. Typically the coefficients $p_{i,j}(x)$ will take the form $p_{i,j}(x) = x_i x_j \lambda_{i,j}(x)$ with $\lambda_{i,j}(x) > 0$. This makes sense for models based on strategy switching from imitations or models arising from ecology.

Depending on the coefficients $p_{i,j}$ this models shows interesting behavior, in particular the chain will ultimately rest in one of the extinction states, that is the vertices of the simplex. In [31], results on the long time and/or large population behavior of this model were proved by comparing its behavior with that of the mean-field ordinary differential equation which can be obtained by taking the first order term in the expansion in $N$ of $\mathbb{E}[f(X_{k+1}^N)|X_k = x]$. Indeed

$$\begin{aligned}\mathbb{E}[f(X_{k+1}^N)|X_k = x] &= \mathbb{E}[f(X_{k+1}^N) - f(x)|X_k^N = x] + f(x) \\ &= \sum_{i,j}\left(f(x + \frac{e_j - e_i}{N}) - f(x)\right)p_{i,j}(x) + f(x)\end{aligned}$$

Taking $G_i(x) = \sum_j (p_{j,i}(x) - p_{i,j}(x))$ and $a(x)$ such that $a_{i,j}(x) = -(p_{j,i}(x) + p_{i,j}(x))$ and $a_{i,i}(x) = \sum_j (p_{j,i}(x) + p_{i,j}(x))$ we obtain

$$\mathbb{E}[f(X_{k+1}^N)|X_k^N = x] = f(x) + \frac{1}{N} < \nabla f(x), G(x) > + \frac{1}{2N^2}Tr(D^2 f(x)a) + o\left(\frac{1}{N^2}\right)$$

If we only take into account the first term in the expansion we obtain an Euler scheme for approximating the ODE $\dot{x} = G(x)$. If we now take into account the second order term we recognize the infinitesimal generator of a stochastic differential equation of the following form.

$$dX_t^{(N)} = G(X_t^{(N)})dt + \frac{1}{\sqrt{N}}\gamma(X_t^{(N)})dB_t$$

where $\circ$ stands for the component by component product in $\mathbb{R}^d$ and $a = \gamma\gamma^*$.

Typically the coefficients $p_{i,j}(x)$ take the form $p_{i,j}(x) = x_i x_j \lambda_{i,j}(x)$. In that case we would obtain a SDE of the form

$$dX_t^{(N)} = X_t^{(N)} F(X_t^{(N)})dt + \frac{1}{\sqrt{N}} \sqrt{X_t^{(N)}} \circ \sigma(X_t^{(N)})dB_t \tag{4.1}$$

This is the type of SDE we will be studying here.

In [46], Schreiber, Benaïm and Atchadé gave criteria for the persistence of a class of SDE on the $d$-dimensional simplex of the following form

$$dX_t = X_t \circ F(X_t)dt + X_t \circ \sigma(X_t)dB_t$$

The main difference between their model and (4.1) is the lack of the Lipschitz property of the diffusion term. This seemingly small difference will lead to a whole different behavior. We will prove that our model will be absorbed in finite time by the boundary, whereas Schreiber, Benaïm and Atchadé model remains in the relative interior of the simplex for all times.

### 4.2.1 Notations and standing hypotheses

We denote by $\Delta$ the $d$-dimensional simplex.

$$\Delta = \{x \in \mathbb{R}^d \ ; \ \forall i = 1 \cdots d \ \ x_i \geqslant 0 \ \ \& \ \ \sum_{i=1}^{d} x_i = 1\}$$

We let $\mathring{\Delta}$ denote the relative interior of $\Delta$. We will denote by $\circ$ the component by component product in $\mathbb{R}^d$.

$$(x_1, x_2, \cdots x_d) \circ (y_1, y_2, \cdots y_d) = (x_1 y_1, x_2 y_2, \cdots, x_d y_d)$$

We consider a family of Markov processes $(X_t^N)_{t \in \mathbb{R}_+}$ on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ taking values in $\Delta$ defined by

$$dX_t^{(N)} = X_t^{(N)} F(X_t^{(N)})dt + \frac{1}{\sqrt{N}} \sqrt{X_t^{(N)}} \circ \sigma(X_t^{(N)})dB_t \tag{4.2}$$

Throughout this chapter, these hypotheses will always be assumed to hold

**Standing Hypothesis 4.2.1 :**

*(i)* $F : \Delta \to \mathbb{R}^d$ *is a L-Lipschitz vector field*

*(ii)* $\forall x \in \Delta \quad \sum_{i=1}^{d} x_i F_i(x) = 0$

*(iii)* $\sigma$ *is a continuously derivable application from* $\Delta$ *to* $\mathcal{M}_{d,l}(\mathbb{R})$

*(iv)* $\forall x \in \Delta$ *and* $\forall j \in \{1, \cdots, l\}$ $\quad \sum_{i=1}^{d} \sqrt{x_i}\sigma_{i,j}(x) = 0$

*(v) For all* $i \in \{1 \cdots d\}$ *and all* $x \in \Delta$*, we have* $(\sigma\sigma^*)_{ii}(x) > \varepsilon$

**Proposition 4.2.2 :**
*For all* $N \geqslant 1$ *the SDE 4.2 admits a weakly unique weak solution.*

This proposition is a consequence of Theorem 2.1.25, i.e. Theorem 4.22 in [25]

Unless specified otherwise, the topology considered will be the topology induced by the classical $\mathbb{R}^d$ metric topology on $\Delta$. If $A$ is a subset of a metric space $(E, d)$, we will denote by $N^\varepsilon(A)$ its $\varepsilon$-neighborhood

$$N^\varepsilon(A) = \{x \in E \,;\, d(x, A) < \varepsilon\}.$$

We denote by $\mathcal{F}_t^N$ the $\sigma$-algebra generated by $\{X_s^N, s \leqslant t\}$. For $A \in \mathcal{F}$ we let $\mathbb{P}_x[A] = \mathbb{P}[A|X_0 = x]$.

From the assumptions on the drift and diffusion terms and the fact that they vanish on the boundary we get that $X_t^N \in \Delta$ a.s. We will compare the solutions of the SDE with those of the ODE

$$\dot{x}_t = x_t F(x_t)$$

**Definition 4.2.3:**
We denote by $L^{(N)}$ the infinitesimal generator of the diffusion $X_\cdot^{(N)}$, that is, the operator defined by

$$\forall f \in \mathcal{D}(L) \quad Lf(x) = \langle x \circ F(x); \nabla f \rangle + \frac{1}{2N} Tr(D^2 f(x)\Sigma(x))$$

where $\Sigma(x) = \sqrt(x) \circ \sigma(x)(\sigma(x) \circ \sqrt{x})*$

<u>*Remark :*</u>

The factor $\frac{1}{\sqrt{N}}$ in the diffusion term doesn't impact the qualitative behavior of the SDE such as its absorption by the border or the existence of quasi-stationary distributions. Thus, when only interested in qualitative behavior, most of the time we will assume $N = 1$ and simply write $X_t$ instead of $X_t^{(1)}$ and $L$ instead of $L^{(1)}$ to simplify notations.

## 4.3 Border absorption in finite time

When studying SDE of the form

$$dX_t = X_t \circ F(X_t)dt + X_t \circ \sigma(X_t)dB_t$$

with $F$ and $\sigma$ Lipschtiz, a simple exponential martingale argument or the use of the strong uniqueness property show that, whenever $X_0 \in \mathring{\Delta}$, $X_t \in \mathring{\Delta}$ almost surely for all

$t$. Such a behavior is no more true when the diffusion term is no more Lipschitz, in fact we get that

**Theorem 4.3.1 :**
*Let* $\tau = \inf\{t > 0\ ;\ X_t \in \partial\Delta\}$.
*Then* $\mathbb{P}_x[\tau < \infty] = 1$

**Proof :**
Let $V_i(x) = -x_i log(x_i)$ and let $U_\delta^i = \{x \in \Delta\ ;\ x_i < \delta\}$
We have

$$LV(x) = (-log(x_i) - 1)x_i F_i(x) - \frac{1}{2}\frac{1}{x_i}x_i \sum_j \sigma_{ij}^2(x)$$
$$= V_i(x)F_i(x) - x_i F_i(x) - \frac{1}{2}(\sigma\sigma^*)_{ii}^2(x)$$

Thus, if $x \in U_\delta^i$ we get

$$LV(x) \leqslant \|F\|(-\delta \log(\delta) + \delta) - \frac{1}{2}\sum_j \sigma_{ij}^2(x)$$

As $(\sigma\sigma^*)_{ii} > \varepsilon$ we get, for $\delta$ small enough and $0 < \alpha < \frac{\varepsilon}{2}$, that

$$LV(x) \leqslant -\alpha$$

Hence, if $x_0 \in U_\delta^i$, and $\tau_{i,\delta} = Inf\{t > 0\ ;\ X_t \notin U_\delta^i\}$

$$V(X_{t\wedge\tau_{i,\delta}}) = V(x_0) + \int_0^{t\wedge\tau_{i,\delta}} LV(X_s)ds + M_{t\wedge\tau_{i,\delta}} \leqslant V(x_0) - \alpha t \wedge \tau_{i,\delta} + M_{t\wedge\tau_{i,\delta}}$$

where $M_t$ is a local martingale.
Then

$$0 \leqslant \mathbb{E}[V(X_{t\wedge\tau_{i,\delta}})] \leqslant V(x_0) - \alpha\mathbb{E}[t \wedge \tau_{i,\delta}]$$

which, in turn, gives

$$\mathbb{E}[t \wedge \tau_{i,\delta}] \leqslant \frac{V(x_0)}{\alpha}$$

In particular we have $\mathbb{P}[\tau_{i,\delta} < \infty] > 0$.
Let us now decompose $\tau_{i,\delta}$ whether the chain exits in the direction of $\partial\Delta$ or in the direction of the interior, we define

$$\tau_{i,\delta}^1 = Inf\{t > 0\ ;\ X_t \notin U_\delta^i\ \&\ X_t \in \partial\Delta\}$$

$$\tau_{i,\delta}^2 = Inf\{t > 0\ ;\ X_t \notin U_\delta^i\ \&\ X_t \notin \partial\Delta\}$$

We naturally get $\tau_{i,\delta} = \tau_{i,\delta}^1 \wedge \tau_{i,\delta}^2$.
Then

$$\mathbb{E}[V(X_{\tau_{i,\delta}})] \leqslant V(x_0) < -\delta log(\delta)$$

$$-\delta log(\delta)\mathbb{P}[\tau_{i,\delta} = \tau_{i,\delta}^2] + \mathbb{E}[\mathbb{1}_{\tau_{i,\delta}=\tau_{i,\delta}^1} V(X_{\tau_{i,\delta}})] < -\delta log(\delta)$$

Thus $\mathbb{P}[\tau_{i,\delta} = \tau_{i,\delta}^2] < 1$, i.e. $\mathbb{P}[\tau_{i,\delta} = \tau_{i,\delta}^1] > 0$

Define $U_\delta = \cup_i U_\delta^i$ the former argument gives us that, for $x \in U_\delta$ we have $\mathbb{P}_x[\tau_i < \infty] > 0$.

If we show that, for all $x \in \delta$, $\mathbb{P}_x[\exists t > 0, X_t \in U_\delta] > 0$ we would then obtain, via the Markov property, that, for all $x \in \delta$, $\mathbb{P}_x[\tau < \infty] > 0$.

To do that we will make use of the Lemma 5.7.4 in [26] on the domain $\Delta \setminus U_{\delta/2}$.

**Lemma 4.3.2 :**
*Let $D$ be an open subset of $\mathbb{R}^d$ and consider a stochastic differential equation on $\overline{D}$ with drift term $b$ and diffusion term $s$ such that*

*(i) $b$ and $s$ don't depend on $t$*

*(ii) $b$ and $s$ are continuous and satisfy the linear growth condition on $\overline{D}$*

$$\|b(x)\|^2 + \|s(x)\|^2 \leqslant K^2(1 + \|x\|^2)$$

*(iii) The SDE admits a weak solution for every starting point in $D$ and this solution is unique in the sense of probability law*

*(iv) for some $1 \leqslant i \leqslant d$ we have*

$$\min_{x\in\overline{D}} (ss^*)_{ii}(x) > 0$$

*Then, for all $x \in D$,*

$$\mathbb{E}_x[\tau_D] < \infty$$

*where $\tau_D = Inf\{t \geqslant 0\ ;\ X_t \notin D\}$*

In our case $b = xF(x)$ and $s = \sqrt{x}\sigma(x)$. On $D = \Delta \setminus U_\delta$ these functions verify Assumptions $(i)$, $(ii)$, and $(iii)$ and on $\Delta \setminus U_\delta$ we have

$$(ss^*)_{ij}(x) = \sqrt{x_i}\sqrt{x_j} \sum_k \sigma_{ik}\sigma_{jk} > \delta\varepsilon^2$$

Hence we can apply Lemma 4.3.2.

We finally get that, for all $x \in \Delta$, $\mathbb{P}_x[\tau < \infty] > 0$. The only step remaining is proving that, in fact

$$\forall x \in \Delta \qquad \mathbb{P}_x[\tau < \infty] = 1$$

We know that

$$\forall x \in U_\delta \qquad \mathbb{P}_x[\tau < \infty] > \frac{\delta log(\delta) - V(x)}{\delta log(\delta)}$$

Then, if $x \in U_{\delta/2} \subset U_\delta$ we get

$$\mathbb{P}_x[\tau < \infty] > \frac{\delta log(\delta) - \delta/2 log(\delta/2)}{\delta log(\delta} > c > 0$$

where $c$ is a positive constant.

Thus, for $x \in \mathring{\Delta}$

$$\begin{aligned}\mathbb{P}_x[\tau < \infty] &= \mathbb{E}_x[\mathbb{1}_{\tau<\infty}] \\ &= \mathbb{E}_x[\mathbb{E}[\mathbb{1}_{\tau<\infty}|\mathcal{F}_{\tau_{U_{\delta/2}}}] > c\end{aligned}$$

Hence, for all $t > 0$

$$E_x[\mathbb{1}_{\tau<\infty}|\mathcal{F}_t] > c$$

As $t$ goes to infinity, $E_x[\mathbb{1}_{\tau<\infty}|\mathcal{F}_t]$ goes to $\mathbb{E}[\mathbb{1}_{\tau<\infty}|X_0 = x]$ a.s. .
Thus $\mathbb{E}[\mathbb{1}_{\tau<\infty}|X_0 = x] > c > 0$ a.s. , i.e. $\mathbb{E}[\mathbb{1}_{\tau<\infty}|X_0 = x] = 1$ a.s.
Finally we get that, for all $x \in \mathring{\Delta}$, $\mathbb{P}_x[\tau < \infty] = 1$.

■

## 4.4 Quasi-stationary Distributions

**Definition 4.4.1:**
Let $\tau_N = Inf\{t > 0\ ;\ X_t^{(N)} \in \partial\Delta\}$. A probability measure $\mu_N$ on the relative interior of the simplex $\mathring{\Delta}$ is said to be a *quasi-stationary distribution* for the process $X^{(N)}$, thereafter referred as QSD, if and only if, for every Borel set $A \subset \mathring{\Delta}$ and every $t > 0$,

$$\mathbb{P}_\mu[X_t^{(N)} \in A|\tau_N > t] = \mu(A).$$

We remark that, in this case, $\mu$ is a fixed point for the conditional evolution

$$\nu \mapsto \mathbb{P}_\nu[X_t^{(N)} \in \centerdot|\tau_N > t]$$

For more information on QSD see e.g. [32],[39] and [12].

### 4.4.1 Existence

First we will give a result about the regularity of the process, namely that the process is strongly Feller, this property will be needed later for the proof of the existence of a QSD.

**Theorem 4.4.2 :**
*The process $X^{(N)}$, up to the time $\tau_N$ where it exits $\mathring{\Delta}$, is a strongly Feller process. That is, for all measurable function $f$ from $\mathring{\Delta}$ to $\mathbb{R}$ and all $t > 0$*

$$\mathbb{E}_x[f(X_t^{(N)})\mathbb{1}_{t<\tau_N}]$$

*is a continuous function of $x$ over $\mathring{\Delta}$. We may remark that, if $f_{|\partial\Delta} = 0$ then*

$$E_x[f(X_t^{(N)})] = \mathbb{E}_x[f(X_t^{(N)})\mathbb{1}_{t<\tau_N}] + \mathbb{E}_x[f(X_t^{(N)})\mathbb{1}_{t\geqslant\tau_N}] = \mathbb{E}_x[f(X_t^{(N)})\mathbb{1}_{t<\tau_N}]$$

*Thus $E_x[f(X_t^{(N)})]$ is also a continuous function of $x$*

This result is a consequence of a theorem announced by Girsanov in [23] about the regularity of multidimensional diffusion process, he never proved said theorem due to his untimely death, a proof of this result and of another Girsanov theorem about the strong Feller property of limits of compatible strong Feller process (result which could also be used here to prove the strong Feller property) can be found in [34].

**Theorem 4.4.3 :**
*For all $N$ there exists a QSD $\mu^N$ for the process $X_\cdot^{(N)}$*

**Proof :** The factor $N$ is not altering the long time behavior of the system, thus we only have to prove the existence of a QSD for the process $X_t = X_t^{(1)}$.

$$dX_t = X_t \circ F(X_t)dt + \sqrt{X_t} \circ \sigma(X_t)dB_t$$

We will prove the existence of a quasi-stationary distribution by making use of Lemma 2.9 in [12] rewritten in our setting.

**Lemma 4.4.4** ( Lemma 2.9 in [12]) **:**
*Let $\mu$ be a probability measure on $\mathring{\Delta}$ such that, for all continuous function $f$*

$$\mathbb{E}_\mu[f(X_\alpha)|\tau > \alpha] = \beta\mu f$$

*Then $\beta > 1$ and there exists a QSD $\nu$ whose exponential rate of survival is $\theta = -\log(\beta) > 0$.*

Let $\Delta^\varepsilon = \{x \in \Delta \ ; \ d(x, \partial\Delta) \geqslant \varepsilon\}$, $\tau_\varepsilon = \inf\{t > 0 \ ; \ X_t \notin \Delta^\varepsilon\}$ and $X^\varepsilon$ be the process $X_t$ killed when it exits $\Delta^\varepsilon$, that is the process defined by $X_t^\varepsilon = X_t$ for $t \in [0, \tau_\varepsilon]$ and $X_t^\varepsilon = \partial$ for $t > \tau_\varepsilon$, where $\partial$ is a cemetery state. As $\Delta^\varepsilon$ is a compact set, we know, from Proposition 2.10 in [12], that $X^\varepsilon$ admits a QSD $\mu^\varepsilon$ with associated parameter $\theta(\varepsilon)$. The measures $\mu^\varepsilon$ are probability measures with support in the compact set $\Delta$, thus, up to a sub-sequence, they converge, in the weak* limit sense, as $\varepsilon$ goes to zero, to a measure $\mu$.

Let $\alpha > 0$ We have $e^{-\alpha\theta(\varepsilon)} = \mathbb{P}_{\mu^\varepsilon}[\tau^\varepsilon > \alpha]$

$$\begin{aligned}\mathbb{P}_\mu[\tau > \alpha] &= \mathbb{P}_\mu[\tau > \alpha] - \mathbb{P}_{\mu^\varepsilon}[\tau > \alpha] + \mathbb{P}_{\mu^\varepsilon}[\tau > \alpha] - \mathbb{P}_{\mu^\varepsilon}[\tau^\varepsilon \geqslant \alpha] + \mathbb{P}_{\mu^\varepsilon}[\tau^\varepsilon \geqslant \alpha] \\ &= \mathbb{P}_{\mu^\varepsilon}[\tau > \alpha \ , \ \tau^\varepsilon < \alpha] + \mathbb{P}_\mu[X_\alpha \in \mathring{\Delta}] - \mathbb{P}_{\mu^\varepsilon}[X_\alpha \in \mathring{\Delta}] + \mathbb{P}_{\mu^\varepsilon}[\tau^\varepsilon \geqslant \alpha]\end{aligned}$$

Due to the ellipticity of $X_t$ on the set $\Delta^{\varepsilon/2}$ we know that (for the definition of ellipticity and related properties we refer to [26] Chapter 5 Section 7)

$$\mathbb{P}[\text{ there exists an open interval } I \text{ such that } \forall t \in I \ X_t \in \partial\Delta^\varepsilon]$$

Thus the exit time of $\mathring{\Delta}^\varepsilon$ is equal to the exit time of $\Delta^\varepsilon$.

Then the function $\mathbb{P}_x[\tau^\varepsilon \geqslant \alpha]$ on $\Delta$ is continuous in $x$ by virtue of the strong Feller property (see Schilling and Wang Theorem 3.4 [43] and Dynkin book [19]). From that we can also deduce the continuity of the function $\mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha]$

Furthermore, the sets $\{\tau > \alpha\ ,\ \tau^\varepsilon < \alpha\}$ are a decreasing family of sets with void intersection. Thus, $(\mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha])_{\varepsilon>0}$ is a decreasing family of continuous functions that verify for all $x \in \Delta$ $\lim_{\varepsilon\to 0} \mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha] = 0$. As $\mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha] = 0$ when $x \in \partial\Delta$ and $\Delta$ is a compact set, we get, using Dini Theorem, that $\mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha]$ goes uniformly to 0 as $\varepsilon$ goes to 0. Hence, there exists $g(\varepsilon)$ such that $\lim_{\varepsilon\to 0} g(\varepsilon) = 0$ and $\mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha] \leqslant g(\varepsilon)$, hence $0 \leqslant \int \mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha]\mu^\varepsilon(dx) \leqslant g(\varepsilon)$. By the Feller property we also get that $\lim_{\varepsilon\to 0} \mathbb{P}_\mu[X_\alpha \in \mathring{\Delta}] - \mathbb{P}_{\mu^\varepsilon}[X_\alpha \in \mathring{\Delta}] = 0$.

We know that, starting from the QSD $\mu^\varepsilon$, the absorption time $\tau^\varepsilon$ has an exponential distribution. Hence, it has no atoms and

$$\mathbb{P}_{\mu^\varepsilon}[\tau^\varepsilon \geqslant \alpha] = \mathbb{P}_{\mu^\varepsilon}[\tau^\varepsilon > \alpha] = e^{-\alpha\theta(\varepsilon)}$$

Finally we obtain that

$$\lim_{\varepsilon\to 0} e^{-\alpha\theta(\varepsilon)} = \mathbb{P}_\mu[\tau > \alpha]$$

We still must prove that there exists an $\alpha$ such that $\mathbb{P}_\mu[\tau > \alpha] > 0$.

Let $V \subset \mathring{\Delta}^\varepsilon$ and let $t \in \mathbb{R}$. By the QSD property we have:

$$\begin{aligned} e^{\alpha\theta(\varepsilon)}\mu^\varepsilon(V) &= \mathbb{P}_{\mu^\varepsilon}[X^\varepsilon_\alpha \in V] \\ &= \int_{\Delta_\varepsilon} \mathbb{P}_x[X^\varepsilon_\alpha \in V]\mu^\varepsilon(dx) \\ &\geqslant \int_V \mathbb{P}_x[X^\varepsilon_\alpha \in V]\mu^\varepsilon(dx) \\ &\geqslant \inf_{x\in V} \mathbb{P}_x[X^\varepsilon_\alpha \in V]\mu^\varepsilon(V). \end{aligned}$$

As the diffusion $X^\varepsilon_t$ is uniformly elliptic on $\Delta^\varepsilon$ the QSD $\mu_\varepsilon$ give a positive weight on all set of positive Lebesgue measure. Thus there exists a set $V \subset \Delta^\varepsilon$ such that $\mu^\varepsilon(V) > 0$ for all $\varepsilon$

Hence

$$e^{\alpha\theta(\varepsilon)} \geqslant \inf_{x\in V} \mathbb{P}_x[X^\varepsilon_\alpha \in V].$$

The left hand term goes to $\mathbb{P}_\mu[\tau > \alpha]$ as $\varepsilon$ goes to 0. From the Feller property we get that $x \mapsto \mathbb{P}_x[X^\varepsilon_\alpha \in V]$ is a continous function.

Here the functions $\mathbb{P}_x[X^\varepsilon_\alpha \in V]$ converge monotonously to $\mathbb{P}_x[X_\alpha \in V]$ as $\varepsilon$ goes to 0. The Dini theorem implies then that the convergence is uniform and thus that

$$\lim_{\varepsilon\to 0} \inf_{x\in V} \mathbb{P}_x[X^\varepsilon_\alpha \in V] = \inf_{x\in V} \mathbb{P}_x[X_\alpha \in V]$$

From that we obtain that

$$\mathbb{P}_\mu[\tau > \alpha] \geqslant \inf_{x \in V} \mathbb{P}_x[X_\alpha \in V].$$

And the second term is clearly positive due to the ellipticity of the process $X$ on $\mathring{\Delta}$.

From now on we will take $\alpha$ such that $\mathbb{P}_\mu[\tau > \alpha] = \beta > 0$. Let $f$ be a continuous function from $\Delta$ to $\mathbb{R}$ and let

$$I = \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha] - \beta f(x)\mu(dx) \right|$$

We will show that $I = 0$

$$\begin{aligned}
I &= \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha] - \beta f(x)\mu(dx) \right| \\
&\leqslant \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha]\mu(dx) - \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha]\mu^\varepsilon(dx) \right| \\
&\quad + \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha] - \mathbb{E}_x[f(X_\alpha)|\tau_\varepsilon > \alpha]\mu^\varepsilon(dx) \right| \\
&\quad + \left| \int \mathbb{E}_x[f(X_\alpha)|\tau_\varepsilon > \alpha] - e^{-\alpha\theta(\varepsilon)} f(x)\mu^\varepsilon(dx) \right| \\
&\quad + \left| e^{-\alpha\theta(\varepsilon)} \int f(x)\mu^\varepsilon(dx) - \beta \int f(x)\mu(dx) \right|
\end{aligned}$$

We will define

$$\begin{aligned}
I_1 &= \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha]\mu(dx) - \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha]\mu^\varepsilon(dx) \right| \\
I_2 &= \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha] - \mathbb{E}_x[f(X_\alpha)|\tau_\varepsilon > \alpha]\mu^\varepsilon(dx) \right| \\
I_3 &= \left| \int \mathbb{E}_x[f(X_\alpha)|\tau_\varepsilon > \alpha] - e^{-\alpha\theta(\varepsilon)} f(x)\mu^\varepsilon(dx) \right| \\
I_4 &= \left| \int e^{-\alpha\theta(\varepsilon)} f(x)\mu^\varepsilon(dx) - \int \beta f(x)\mu(dx) \right|
\end{aligned}$$

From the QSD property of $\mu^\varepsilon$ we get that $I_3 = 0$. As $\mu^\varepsilon \rightharpoonup \mu$, we get that $\lim_{\varepsilon \to 0} I_4 = 0$ and, as our process is strongly Feller, we also get $\lim_{\varepsilon \to 0} I_1 = 0$.

Only $I_2$ remains to be controlled.

For that we will first see what happens should $f$ equals $\mathbb{1}_A$ with $A \subset \mathring{\Delta}$ a Borel set.

In that case we get

$$
\begin{aligned}
I_2 &= \left| \int \mathbb{E}_x[f(X_\alpha)|\tau > \alpha] - \mathbb{E}_x[f(X_\alpha)|\tau_\varepsilon > \alpha]\mu^\varepsilon(dx) \right| \\
&= \left| \int \mathbb{P}_x[X_\alpha \in A|\tau > \alpha] - \mathbb{P}_x[X_\alpha \in A|\tau_\varepsilon > \alpha]\mu^\varepsilon(dx) \right| \\
&= \left| \int \frac{\mathbb{P}_x[X_\alpha \in A]}{\mathbb{P}_x[\tau > \alpha]} - \frac{\mathbb{P}_x[X_\alpha \in A\ ,\ \tau^\varepsilon > \alpha]}{\mathbb{P}_x[\tau^\varepsilon > \alpha]}\mu^\varepsilon(dx) \right| \\
&= \left| \int \mathbb{P}_x[X_\alpha \in A\ ,\ \tau^\varepsilon > \alpha] \left( \frac{1}{\mathbb{P}_x[\tau > \alpha]} - \frac{1}{\mathbb{P}_x[\tau^\varepsilon > \alpha]} \right) + \mathbb{P}_x[X_\alpha \in A\ ,\ \tau^\varepsilon < \alpha]\mu^\varepsilon(dx) \right| \\
&= \left| \int \mathbb{P}_x[X_\alpha \in A\ ,\ \tau^\varepsilon > \alpha] \left( \frac{\mathbb{P}_x[\tau^\varepsilon > \alpha] - \mathbb{P}_x[\tau > \alpha]}{\mathbb{P}_x[\tau > \alpha]\mathbb{P}_x[\tau^\varepsilon > \alpha]} \right) + \mathbb{P}_x[X_\alpha \in A\ ,\ \tau^\varepsilon < \alpha]\mu^\varepsilon(dx) \right| \\
&\leqslant \left| \int \frac{\mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha]}{\mathbb{P}_x[\tau > \alpha]\mathbb{P}_x[\tau^\varepsilon > \alpha]} + \mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha]\mu^\varepsilon(dx) \right|
\end{aligned}
$$

However, the sets $\{\tau > \alpha\ ,\ \tau^\varepsilon < \alpha\}$ are a decreasing family of sets with void intersection, thus $\lim_{\delta \to 0} \mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha] = 0$ and, by monotonous convergence, we also get $\lim_{\delta \to 0} \int \mathbb{P}_x[\tau > \alpha\ ,\ \tau^\varepsilon < \alpha]\mu(dx) = 0$. Thus, if $f = \mathbb{1}_A$, we get $\lim_{\delta \to 0} I_2 = 0$. The same conclusion will hold for a linear combination of such functions. Finally, when $f$ is only supposed continuous, for all $\gamma > 0$ we may take $g_\gamma$ a simple function such that $\|f - g_\gamma\|_\infty < \gamma$ and obtain that

$$\limsup_{\varepsilon \to 0} I_2 < \gamma$$

Finally we obtain that $I = 0$, that is $\mathbb{E}_\mu[f(X_\alpha)|\tau > \alpha] = \beta \int f(x)\mu(dx)$. Lemma 2.9 in [12] allows us to conclude that $\beta < 1$ and that there exists a QSD for the process $X_t$.

■

It might comes as a surprise that the dynamic induced by $\dot{x} = x \circ F(x)$ doesn't impact on the existence of a QSD: whether there exists an interior attractor for the dynamical system $\dot{x} = x \circ F(x)$(that is the system is permanent) or the dynamic $\dot{x} = x \circ F(x)$ goes quickly to the border, there still exists a QSD. In some simple case we might even compute it.

**Example 6:**
We study here the one-dimensionnal SDE

$$dX_t = X_t(1 - X_t)dt + \sqrt{X_t(1 - X_t)}dB_t$$

The deterministic dynamic $\dot{x} = x(1 - x)$ has a very simple behavior: For all $x \neq 0$, the solution of the ODE $\varphi_t(x)$ with initial condition converges to 1 as $t$ goes to infinity. Let us look for a QSD for the process $X_t$. For that we look for a probability measure $\mu$ such that

$$\mu L = \lambda \mu \tag{4.3}$$

with $\lambda > 0$, and $L$ the infinitesimal generator associated with the semi-group $P_t f = \mathbb{E}[f(X_t)\mathbb{1}_{\tau>t}]$ To simplify the problem we will only search among probability measure of the form $\mu(dx) = g(x)dx$ with $g$ of class $\mathcal{C}^2$.

In that case (4.3) can be rewritten as $L^* g = \lambda g$ where $L^*$ is the adjoint of the operator $L$. This leads to the ODE

$$\frac{1}{2}((x(1-x)g(x))'' - ((x(1-x)g(x))' = \lambda g$$

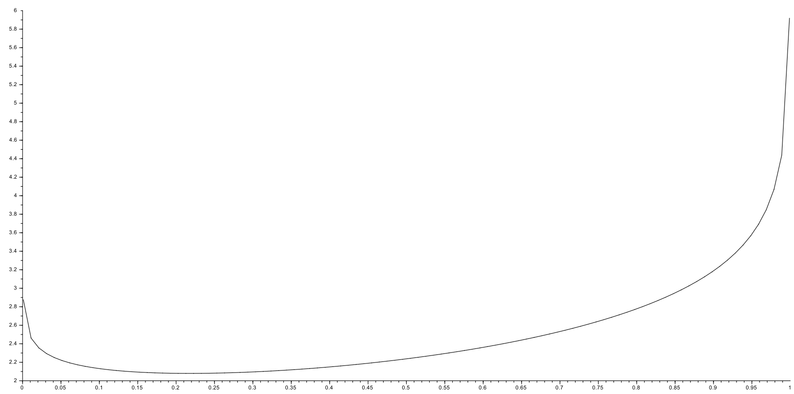
Defining $h(x) = ((x(1-x)g(x))$ we obtain

$$\frac{h''(x)}{2} - h'(x) = \frac{\lambda h(x)}{(x(1-x)}$$

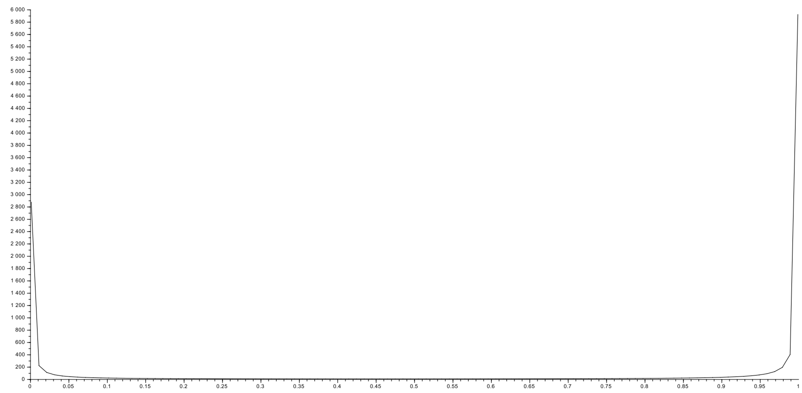
Such an ODE is easily solved and the solution takes the form

$$\begin{pmatrix} h'(x) \\ h(x) \end{pmatrix} = C \exp \lambda \int_{x_0}^{x} \begin{pmatrix} \frac{1}{2} & \frac{1}{u(1-u)} \\ 1 & 0 \end{pmatrix} du = C \exp \lambda \begin{pmatrix} \frac{x-x_0}{2} & \log\left(\frac{x}{1-x}\right) - \log\left(\frac{x_0}{1-x_0}\right) \\ x - x_0 & 0 \end{pmatrix}$$

where $C$ is a $1 \times 2$ constant vector. We skip the tedious calculations and give the graph of the function $h$



which in turn gives us the graph of $g$



### 4.4.2 Absorption time

We recall a classical result about QSD and absorption time, see e.g. [32]

**Proposition 4.4.5 :**
*Suppose that $\mu$ is a QSD for this process $Z_t$. Then there exists a positive real number $\theta(\mu)$ such that*

$$\mathbb{P}_\mu[\tau_N > t] = e^{-\theta(\mu)t}$$

A set $A \subset \Delta$ is called an *attractor* for the flow $\{\varphi_t\}$ if

(i) $A$ is compact and invariant, i.e. for every $t \in \mathbb{R}$ $\varphi_t(A) = A$.

(ii) There exists a neighborhood $U$ of $A$, called a fundamental neighborhood, such that

$$\lim_{t\to\infty} d(\varphi_t(x), A) = 0$$

uniformly in $x$ in $U$.

Let

$$D_N(T) = \max_{0\leqslant t\leqslant T} \|X_t^{(N)} - \varphi_t(X_0^{(N)})\|$$

be the variable measuring the distance between the trajectories $t \mapsto X_t^{(N)}$ and $t \mapsto \varphi_t(X_0^{(N)})$. We have the following estimate on $D_N(T)$, we refer to Section 4 for the proof in a more general setting.

**Proposition 4.4.6 :**

$$\forall \delta > 0 \quad \mathbb{P}[D_N(T) \geqslant \delta] \leqslant \frac{T\|\sigma\|_\infty}{N\delta}$$

*In particular, we get*

$$D_N(T) \xrightarrow[N\to\infty]{\mathbb{P}} 0$$

**Theorem 4.4.7 :**
*Starting from $\mu^N$, the law of the absorption time and its expectation are given by Proposition 4.4.5. If we further assume that the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$, then, the following estimate holds :*

$$0 \leqslant 1 - e^{\theta_N} \leqslant O\left(\frac{1}{N}\right)$$

*where $\theta_N = \theta(\mu^N)$.*

*Thus, there exists a constant $C > 0$ such that*

$$\mathbb{E}_{\mu^N}[\tau] \geqslant CN$$

**Proof :**

Let $V \subset \mathring{\Delta}$ such that $\mu^N(V) > 0$ for all $N$, and let $t \in \mathbb{R}$. By the QSD property we have:

$$\begin{aligned} e^{t\theta_N}\mu^N(V) &= \mathbb{P}_{\mu^N}[X_t^{(N)} \in V] \\ &= \int_\Delta \mathbb{P}_x[X_t^{(N)} \in V]\mu^N(dx) \\ &\geqslant \int_V \mathbb{P}_x[X_t^{(N)} \in V]\mu^N(dx) \\ &\geqslant \inf_{x\in V} \mathbb{P}_x[X_t^{(N)} \in V]\mu^N(V). \end{aligned}$$

Thus

$$e^{t\theta_N} \geqslant \inf_{x\in V} \mathbb{P}_x[X_t^{(N)} \in V].$$

Let $U \subset \mathring{\Delta}$ be a compact fundamental neighborhood of the attractor $A$. We know that $d(\varphi_t(x), A)$ converges uniformly to 0 over $U$. Hence

$$\forall \varepsilon > 0 \quad \exists T(\varepsilon) > 0 \quad \forall t \geqslant T(\varepsilon) \quad \forall x \in U \quad d(\varphi_t(x), A) < \varepsilon.$$

Let $\alpha = d(A, U^c)$, $\varepsilon < \alpha$, $T = T(\varepsilon)$ and $\delta < \alpha - \varepsilon$.

For all $x \in U$

$$\begin{aligned} \mathbb{P}_x[X_T^N \in U^c] &\leqslant \mathbb{P}_x[d(X_T^N, A) > \alpha] \\ &\leqslant \mathbb{P}_x[d(X_T^N, \varphi_T(x)) > \alpha - \varepsilon] \\ &\leqslant \mathbb{P}_x\left[D_N(T) > \alpha - \varepsilon\right] \\ &\leqslant \frac{CTe^{LT}}{\delta^2 N} \quad \text{for } N \text{ large enough (see Theorem 4.5.1)} \end{aligned}$$

We need to show that $\mu^N(U) > 0$. However $\mu^N(U) = 0$ implies that

$$\forall t > 0 \quad \mathbb{P}_\mu^N[X_t^{(N)} \in U | \tau > t] = 0$$

Which, due to the property 4.2.1($v$) of the diffusion term is clearly absurd. Then

$$\begin{aligned} e^{T\theta_N} &\geqslant \inf_{x\in U} \mathbb{P}_x[X_t^N \in U] \\ &\geqslant 1 - \max_{x\in U} \mathbb{P}_x[X_t^N \in U^c] \\ &\geqslant 1 - \frac{CTe^{LT}}{\delta^2 N} \end{aligned}$$

Therefore

$$1 - e^{T\theta_N} \leqslant 1 - \left(1 - \frac{CTe^{LT}}{\delta^2 N}\right)^{\frac{1}{T}}$$

In conclusion we have

$$0 \leqslant 1 - e^{\theta_N} \leqslant O\left(\frac{1}{N}\right)$$

■

### 4.4.3 Convergence of the QSD to an invariant measure

A probability measure $\mu$ on $\Delta$ is called an *invariant measure* for the flow $\{\varphi_t\}$ if, for all $t \in \mathbb{R}$ and all Borel set $A \in \mathcal{B}(\Delta)$, $\mu(\varphi_t^{-1}(A)) = \mu(A)$.

**Theorem 4.4.8 :**
*The set of limit points of $\{\mu^N\}$ for the weak* topology is a subset of the set of invariant measures for the flow $\{\varphi_t\}$.*

*Remark* In [31], we needed the existence of an attractor to ensure the convergence of the QSD to invariant measures. This was linked to a renormalization of time for the process $X_k^N$ and the subsequent need to ensure that $e^{-N\theta_N}$ converges to zero. Here we don't have to make such a rescaling, thus the existence of an attractor is not needed to ensure the convergence of the QSD to invariant measures.

**Proof :**
Let $f$ be a Lipschitz function from $\Delta$ to $\mathbb{R}$ with constant $L$. We suppose that the sequence $\mu^N$ weakly converges to a measure $\mu$. Let $t > 0$. We want to prove that

$$\lim_{N\to\infty} \int f(x)\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx) = 0$$

The QSD property gives us that, for all $k$

$$\int f(x)\mu^N(dx) = \int \mathbb{E}_x\left[f(X_T^N)\middle|\tau_N > T\right]\mu^N(dx)$$

Let

$$I = \left| \int f(x)\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx) \right|$$

Then, for all $k$,

$$\begin{aligned} I &= \left| \int f(x)\mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx) \right| \\ &= \left| \int \mathbb{E}_x \left[ f(X_T^N) \middle| \tau_N > T \right] \mu^N(dx) - \int f(\varphi_t(x))\mu^N(dx) \right| \\ &= \left| \int \mathbb{E}_x \left[ f(X_T^N) - f(\varphi_t(x)) \middle| \tau_N > T \right] \mu^N(dx) \right| \end{aligned}$$

In particular, for $T = t$.

$$I = \left| \int \mathbb{E}_x \left[ f(X_t^N) - f(\varphi_t(x)) \middle| \tau_N > t \right] \mu^N(dx) \right|$$

By Theorem 4.5.1, we know that, for $N$ large enough, we have

$$\mathbb{P}_x[D_N(t) > \delta] \leqslant \frac{Cte^{Lt}}{\delta^2 N}.$$

Thus

$$\mathbb{E}_x[D_N(t)] = \int_0^{+\infty} \mathbb{P}_x[D_N(t) > \delta] d\delta \leqslant \int_0^{+\infty} Min\left(1, \frac{Cte^{Lt}}{\delta^2 N}\right) d\delta = \frac{Kte^{Lt}}{N}$$

with $K$ a constant.

Hence

$$\begin{aligned} I &= \left| \int \mathbb{E}_x \left[ f(X_t^N) - f(\varphi_t(x)) \middle| \tau_N > t \right] \mu^N(dx) \right| \\ &\leqslant \left| \int \frac{\mathbb{E}_x \left[ f(X_t^N) - f(\varphi_t(x)) \right]}{\mathbb{P}_x\left[\tau_N > t\right]} \mu^N(dx) \right| \\ &\leqslant \left| \int \frac{\mathbb{E}_x \left[ L|X_t^N - \varphi_t(x)| \right]}{\mathbb{P}_x\left[\tau_N > t\right]} \mu^N(dx) \right| \\ &\leqslant \left| \int \frac{\mathbb{E}_x \left[ L(D_N(t)) \right]}{\mathbb{P}_x\left[\tau_N > t\right]} \mu^N(dx) \right| \\ &\leqslant \left| L \frac{Kte^{Lt}}{N} e^{\theta_N t} \right| \underset{N \to +\infty}{\longrightarrow} 0 \end{aligned}$$

■

**Definition 4.4.9:**
For $K$ compact subset of $\mathring{\Delta}$ we denote

$$\beta_{\delta,K}(N) = \sup_{x\in K} \mathbb{P}_x[X_1^{(N)} \in \Delta \setminus N^\delta(\varphi_1(x))]$$

**Proposition 4.4.10 :**
*If the flow $\{\varphi_t\}$ admits an attractor $A \subset \mathring{\Delta}$, then, for all $K$ compact subset of $\mathring{\Delta}$ and neighborhood of $A$, there exists $\delta > 0$ such that $e^{-\theta_N} \geqslant 1 - \beta_{\delta,K}(N)$. Moreover, if there exists $U_K$ an open neighborhood of $\partial\Delta$ with*

$$\lim_{N\to\infty} \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K} \mathbb{P}_x[X_1^{(N)} \in \partial\Delta]} = 0$$

*is unbounded. Then, for all limiting measure $\mu$, we have $\mu(U_{K,T}) = 0$.*

**Proof :**

As our system evolve in a compact space we know, see e.g. Conley [13] I 7.2, that there exists a Lyapunov function $g$ for the attractor $A$, i.e. $A = g^{-1}(0)$ and, for $x$ in the basin of attraction of $A$, $t \mapsto g(\varphi_t(x))$ is strictly decreasing. Thus there exists $U$ an open neighborhood of $A$ such that $\overline{U} \subset B(A) \cap K$ where $B(A)$ is the basin of attraction of $A$ and $\varphi_1(\overline{U}) \subset U$. Let $\delta < d(\varphi_1(\overline{U}), U^c)$. Then $N^\delta(\varphi_1(\overline{U})) \subset U$.

Thus

$$\begin{aligned}
e^{-\theta_N}\mu^N(U) &= \int_\Delta \mathbb{P}_x[X_1^{(N)} \in U]\mu^N(dx) \\
&\geqslant \int_U \inf_{x\in U} \mathbb{P}_x[X_1^{(N)} \in U]\mu^N(dx) \\
&\geqslant \mu^N(U)\left(1 - \sup_{x\in U} \mathbb{P}_x[X_1^{(N)} \in U^c]\right) \\
&\geqslant \mu^N(U)\left(1 - \sup_{x\in U} \mathbb{P}_x[X_1^{(N)} \in N^\delta(\varphi_1(\overline{U}))^c]\right) \\
&\geqslant \mu^N(U)\left(1 - \beta_{\delta,K}(N)\right)
\end{aligned}$$

We finally get $e^{-\theta_N} \geqslant 1 - \beta_{\delta,K}(N)$

From this, as $\mu^N(\mathring{\Delta}) = 1$, we obtain

$$\begin{aligned}
1 - \beta_{\delta,K}(N) &\leqslant e^{-\theta_N}\mu^N(\mathring{\Delta}) \\
&\leqslant \int_{\mathring{\Delta}} \left(1 - \mathbb{P}_x[X_1^{(N)} \in \partial\Delta]\right)\mu^N(dx) \\
&\leqslant \mu^N(\Delta \setminus U_K) + \mu^N(U_K)\left(1 - \inf_{x\in U_K} \mathbb{P}_x[X_1^{(N)} \in \partial\Delta]\right)
\end{aligned}$$

Hence

$$\mu^N(U_K) \leqslant \frac{\beta_{\delta,K}(N)}{\inf_{x\in U_K} \mathbb{P}_x[X_1^{(N)} \in \partial\Delta]}$$

$U_K$ being an open set, the weak convergence of the measures $\mu^N$ gives us the desired result.

■

## 4.5 Asymptotic behavior of the system

### 4.5.1 Law of Large Numbers

In this section we will take more general diffusion processes, namely we study $X_t^\varepsilon$ such that

$$dX_t^\varepsilon = b(X_t^\varepsilon)dt + \sqrt{\varepsilon}\Sigma(X_t^\varepsilon)dB_t \tag{4.4}$$

where $X_0^\varepsilon = x \in \Delta$, $t \in [0,1]$, $b$ is Lipschitz continuous with constant $L$ and $\Sigma$ is non negative, 1/2 Hölder continuous with the same constant $L$. Both $\Sigma$ and $b$ vanish on $\partial\Delta$ such that $X_t^\varepsilon \in \Delta$ a.s. We want to compare the trajectories of the SDE with those of the ODE.

$$\dot{x}_t = b(x_t)$$

As our diffusion term is not Lipschitz continuous, classical large numbers or large deviations results like the ones that can be found in [21] cannot be applied here. By extending their approach to our weaker hypotheses we will obtain similar results, however the lack of regularity on the diffusion term forces our results to be much weaker than the classical Lipschitz results.

**Theorem 4.5.1 :**
*Let $m_\varepsilon(T) = \sup_{0\leqslant s\leqslant T} \|X_s^\varepsilon - x_s\|$. Then*

$$\forall\delta > 0 \quad \mathbb{P}[m_\varepsilon(T) \geqslant \delta] \leqslant \frac{T\varepsilon\|\Sigma\|_\infty}{\delta}$$

*In particular, we get*

$$m_\varepsilon(T) \xrightarrow[\varepsilon\to 0]{\mathbb{P}} 0$$

<u>*Remark*</u>
If we go back to our first SDE 3.4.9 we have $D_N(T) = m_{\frac{1}{N}}(T)$.

<u>**Proof :**</u> We have

$$\begin{aligned} m_\varepsilon(t) &= \sup_{0\leqslant s\leqslant t} \|\int_0^s b(X_u^\varepsilon) - b(x_u)du + \frac{1}{\sqrt{N}}\int_0^s \Sigma(X_u^\varepsilon)dB_u\| \\ &\leqslant L\int_0^t m(s)ds + \sup_{0\leqslant s\leqslant t} \|\sqrt{\varepsilon}\int_0^s \Sigma(X_u^\varepsilon)dB_u\| \end{aligned}$$

We denote $Z_t^\varepsilon = \sup_{0\leqslant s\leqslant t} \|\sqrt{\varepsilon}\int_0^s \Sigma(X_u^\varepsilon)dB_u\|$.
By the Gronwall Lemma we get $m_\varepsilon(T) \leqslant e^{LT} Z_T$.
Let $Y_t^\varepsilon$ be defined by $dY_t^\varepsilon = \sqrt{\varepsilon}\Sigma(X_t^\varepsilon)dB_t$.
Then

$$\begin{aligned} d\|Y_t^\varepsilon\|^2 &= 2\varepsilon < Y_t^\varepsilon, dY_t^\varepsilon > + \frac{\varepsilon}{2} Tr(2Id\ d < Y_t^\varepsilon >) \\ &= 2\varepsilon < Y_t^\varepsilon, \Sigma(X_t^\varepsilon) > dB_t + \varepsilon Tr(x_t^\varepsilon \circ \Sigma\Sigma^*(X_t^\varepsilon))dt \end{aligned}$$

Thus $\mathbb{E}[\|Y_t^\varepsilon\|^2|\mathcal{F}_s] = \|Y_s^\varepsilon\|^2 + \varepsilon\int_s^t Tr(X_t^\varepsilon \circ \Sigma\Sigma^*(X_t^\varepsilon))dt \leqslant \|Y_s^\varepsilon\|^2$
$Y_t^\varepsilon$ is then a sub-martingale. Using a theorem of Doob we get, for $\delta > 0$

$$\mathbb{P}[\sup_{0\leqslant t\leqslant T} \|Y_t^\varepsilon\|^2 \geqslant \delta] \leqslant \frac{\mathbb{E}[\|Y_T^\varepsilon\|]}{\delta} \leqslant \frac{T\varepsilon\|\Sigma\|_\infty}{\delta}$$

Hence the announced result.

■

### 4.5.2 Large deviations principle

This large deviations principle may appear to be inapplicable in any practical case due to the complexity of its rate function, however, as seen in [31], sometimes we only need to know that there exists a large deviations principle to derive more practical results.

**Theorem 4.5.2 :**
*$X_t^\varepsilon$ satisfies a large deviation principle on $\mathcal{C}([0,1])$ with rate function*

$$I(y) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B(y,\delta)} I_m(z)$$

*where $I_m(x) = Inf\{J(x)\,;\, y = F^m(x)\}$, $J(g) = \int_0^1 \|\dot{g}(t)\|^2 dt$ and $F^m$ is the application that maps $g \in \mathcal{C}([0,1])$ to $h$ uniquely defined by*

$$h(t) = h(\frac{k}{m}) + b(h(\frac{k}{m}))(t - \frac{k}{m}) + \Sigma(h(\frac{k}{m}))(g(t) - g(\frac{k}{m}))$$

*for $t \in [\frac{k}{m}, \frac{k+1}{m}]$ and $h_i(t) = 0$ if $\exists s$ s.t. $h_i(s) = 0$.*

**Proof :**

In order to simplify notations we will take $x = 0$ and move the simplex such that $0 \in \Delta$.

We will approximate $X^\varepsilon$ by $X^{\varepsilon,m}$ defined as the strong solution of

$$\begin{aligned} dX_t^{\varepsilon,m} &= b(X^{\varepsilon,m}_{\frac{[mt]}{m}})dt + \sqrt{\varepsilon}\Sigma(X^{\varepsilon,m}_{\frac{[mt]}{m}})dW_t \\ d(X_t^{\varepsilon,m})_i &= 0 \text{ if } (X_t^{\varepsilon,m})_i = 0 \end{aligned}$$

where $W_t$ is the Brownian motion for which there exists a solution to (4.4).

**Lemma 4.5.3 :**
*For all* $\delta > 0$

$$\lim_{m\to\infty} \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}[\|X^{\varepsilon,m} - X^{\varepsilon}\| > \delta] = -\infty$$

We will prove this Lemma later on.

Let $F^m$ be the application that maps $g \in \mathcal{C}([0,1])$ to $h$ uniquely defined by

$$h(t) = h(\frac{k}{m}) + b(h(\frac{k}{m}))(t - \frac{k}{m}) + \Sigma(h(\frac{k}{m}))(g(t) - g(\frac{k}{m})$$

for $t \in [\frac{k}{m}, \frac{k+1}{m}]$ and $h_i(t) = 0$ if $\exists s$ s.t. $h_i(s) = 0$.

We remark that $X^{\varepsilon,m} = F^m(\sqrt{\varepsilon}W_.)$. The assumptions on $b$ and $\Sigma$ ensure the continuity of $F^m$ for the norm $\|.\|_\infty$.

**Lemma 4.5.4 :**
*Let* $g \in H^1$

$$\limsup_{m\to\infty} \sup_{g;\|g\|_{H^1}\leqslant\alpha} d(F^m(g), F(g)) = 0$$

*where* $d$ *is the distance induced by* $\|.\|_\infty$ *and* $F$ *is the application which maps* $g$ *to the set of functions* $f$ *that satisfy*

$$f(t) = \int_0^t b(f(s))ds + \int_0^t \Sigma(f(s))\dot{g}(s)ds$$

*if* $g \in H^1$ *and* $\infty$ *otherwise.*

We will prove this Lemma later on.

The following Theorem is Theorem 4.2.16 is [16]

**Theorem 4.5.5 :**
*Let* $\{\mu_{\varepsilon,m}\}$ *be a family of probability measures which, for every* $m$*, satisfies a large deviations principle with rate function* $I_m$*.*

*Then, every family of probability measures* $\{\tilde{\mu}_\varepsilon\}$ *for which* $\mu_{\varepsilon,m}$ *are exponential approximations satisfies a weak large deviations principle with rate function*

$$I(y) = \sup_{\delta>0} \liminf_{m\to\infty} \inf_{z\in B_{y,\delta}} I_m(z)$$

*where* $B_{y,\delta}$ *is the ball centered in* $y$ *with radius* $\delta$*.*

*Moreover, if* $I(\cdot)$ *is a good rate function and, for every closed set* $F$

$$\inf_{y\in F} I(y) \leqslant \limsup_{m\to\infty} \inf_{y\in F} I_m(y)$$

*Then* $\{\tilde{\mu}_\varepsilon\}$*satisfies a strong large deviations principle.*

The Lemma 4.5.3 tells us that $X^{\varepsilon,m}$ are exponential approximations of $X^{\varepsilon}$. From Schilder Theorem we can infer a large deviations principle for $\varepsilon B_.$ with good rate function $J(g) = \int_0^1 \|\dot{g}(t)\|^2 dt$ if $g \in H^1$ and $\infty$ otherwise, the Contraction Principle will in turn give us the LDP for $\mu_{\varepsilon,m}$ with good rate function $I_m(x) = Inf\{J(x)\,;\, y = F^m(x)\}$.

Combining said results gives us the aforementioned Theorem.

■

**Proof :** [Proof of Lemma 4.5.3]

Let $\delta > 0$ and let $Z_t = X_t^{\varepsilon,m} - X_t^{\varepsilon}$.

Let $\rho > 0$, $\tau_1 = Inf\{t;\ \|X_t^{\varepsilon,m} - X_{\frac{[mt]}{m}}^{\varepsilon,m}\| \geqslant \rho\} \wedge 1$ and $\tau_2 = Inf\{t;\ \|Z_t\| \geqslant \delta\} \wedge \tau_1$.

The following inclusion holds :

$$\{\|X^{\varepsilon,m} - X^{\varepsilon}\|_\infty > \delta\} \subset \{\tau_1 < 1\} \cup \{\sup_{t \in [0\tau_1]} \|X_t^{\varepsilon,m} - X_t^{\varepsilon}\| \geqslant \delta\}$$

We denote $dZ_t = b_t dt + \sqrt{\varepsilon}\sigma_t dB_t$. We know that $b_t$ and $\sigma_t$ satisfy the following assertions

$$\begin{aligned} \|\sigma_t\| &\leqslant M(\rho + \|Z_t\|)^{1/2} \\ \|b_t\| &\leqslant B(\rho + \|Z_t\|) \end{aligned}$$

We define $\phi(y) = (\rho + \|y\|)^{1/\varepsilon}$ and $U_t = \phi(Z_t)$. Then $dU_t = g_t dt + \tilde{\sigma}_t dB_t$ where

$$|g_t| \leqslant \frac{K}{\varepsilon} U_t(\|Z_t\|^2 + \|Z_t\| + d\varepsilon)$$

$$|\tilde{\sigma}_t| \leqslant \frac{K}{\sqrt{\varepsilon}} U_t \|Z_t\|$$

Thus $\tilde{\sigma}_t$ is bounded on $[0;\tau_2]$ which, in turn, implies that $U_t - \int_0^t g_s ds$ is a martingale on $[0;\tau_2]$.

By Doob Inequality we then get

$$\begin{aligned} \mathbb{E}[U_{t\wedge\tau_2}] &= U_o + \mathbb{E}\left[\int_0^{t\wedge\tau_2} g_s ds\right] \\ &\leqslant U_0 + \frac{K}{\varepsilon}\mathbb{E}\left[\int_0^{t\wedge\tau_2} U_s(\|Z_s\|^2 + \|Z_s\| + d\varepsilon)ds\right] \\ &\leqslant U_0 + \frac{K'}{\varepsilon}\mathbb{E}\left[\int_0^{t\wedge\tau_2} U_s ds\right] \\ &\leqslant U_0 + \frac{K'}{\varepsilon}\int_0^{t\wedge\tau_2} \mathbb{E}\left[U_s\right] ds \end{aligned}$$

Then, by use of the Gronwall Lemma

$$\mathbb{E}[U_{\tau_2}] = \mathbb{E}[U_{1\wedge\tau_2}] \leqslant U_0 e^{K'/\varepsilon} \leqslant \phi(Z_0) e^{K'/\varepsilon}$$

Given that $\phi$ is a non-negative and growing function, we get by Tchebychev inequality

$$\mathbb{P}[\|Z_{\tau_2}\| \geqslant \delta] = \mathbb{P}[\phi(Z_{\tau_2}) \geqslant \phi(\delta)] \leqslant \frac{\mathbb{E}[\phi(Z_{\tau_2})]}{\phi(\delta)} \leqslant \frac{\mathbb{E}[U_{\tau_2}]}{\phi(\delta)}$$

Thus

$$\varepsilon \log \mathbb{P}[\|Z_{\tau_2}\| \geqslant \delta] \leqslant K' + \log\left(\frac{\rho + \|Z_0\|}{\rho+\delta}\right)$$

We know that $\{\sup_{t\in[0,\tau_1]} \|Z_t\| > \delta\} = \{\|Z_{\tau_2} > \delta\}$

We then get

$$\varepsilon \log \mathbb{P}[\sup_{t\in[0\tau_1]} \|Z_t\| \geqslant \delta] \leqslant K' + \log\left(\frac{\rho + \|Z_0\|}{\rho+\delta}\right)$$

$K'$ doesn't depend on $\varepsilon, \delta, \rho$ and $m$. Hence

$$\lim_{\rho\to 0} \sup_{m\geqslant 1} \limsup_{ep\to 0} \varepsilon \log \mathbb{P}[\sup_{t\in[0\tau_1]} \|Z_t\| \geqslant \delta] = -\infty$$

$$\|X_t^{\varepsilon,m} - X_{\frac{[mt]}{m}}^{\varepsilon,m}\| \leqslant C[\frac{1}{m} + \sqrt{\varepsilon} \max_{k=0\ldots m-1} \sup_{0\leqslant s\leqslant 1/m} B_{s+k/m} - B_{k/m}]$$

where $C \geqslant max(\|\sigma\|_\infty, \|b\|_\infty)$

We get that, according to Lemma 5.2.1 in [16], $\forall m \leqslant C/\rho$

$$\mathbb{P}[\sup_{t\in[0,1]} \|X_t^{\varepsilon,m} - X_{\frac{[mt]}{m}}^{\varepsilon,m}\| \geqslant \rho] \leqslant m\mathbb{P}[\sup_{0\leqslant s\leqslant 1/m} B_s \geqslant \frac{\rho - C/m}{\sqrt{\varepsilon}C}]$$
$$\leqslant 4dme^{-m\frac{(\rho-C/m)^2}{2d\varepsilon C^2}}$$

Thus $\forall \rho > 0$

$$\lim_{n\to\infty} \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}[\tau_1 < 1] = -\infty$$

Finally we get that

$$\lim_{m\to\infty} \limsup_{\varepsilon\to 0} \varepsilon \log \mathbb{P}[\|X^{\varepsilon,m} - X^\varepsilon\| > \delta] = -\infty$$

■

**Proof :** [Proof of Lemma 4.5.4]

Let $g \in H^1([0,1])$ (we can restrict ourselves to $g \in H^1$ because we only care about $g$ such that $J(g) < \infty$) and $f_m = F^m(g)$, $m \in \mathbb{N}$.

Let $t, u \in [0,1]$

$$
\begin{aligned}
\|f_m(t)-f_m(u)\| &= \|\int_u^t b(f_m(\frac{[ms]}{m}))ds + \Sigma(f_m(\frac{[mt]}{m}))(g(t)-g(\frac{[mt]}{m})\\
&\quad + \sum_{k=[mu]+1}^{[mt]-1} \Sigma(f_m(\frac{k}{m}))(g(\frac{k+1}{m})-g(\frac{k}{m}))\\
&\quad + \Sigma(f_m(\frac{[mu]}{m})(g(\frac{[mu]}{m})-g(u))\|\\
&\leqslant \|b\|_\infty(t-u) + \|\Sigma(f_m(\frac{[mt]}{m}))\|\|g\|_{H^1}|t-\frac{[mt]}{m}|\\
&\quad + \sum_{k=[mu]+1}^{[mt]-1} \|\Sigma(f_m(\frac{k}{m}))\|\|g\|_{H^1}|\frac{k+1}{m}-\frac{k}{m}|\\
&\quad + \|\Sigma(f_m(\frac{[mu]}{m})\|\|g\|_{H^1}|\frac{[mu]}{m})-u|\\
&\leqslant (\|b\|_\infty + \|\Sigma\|_\infty\|g\|_{H^1})(t-u)
\end{aligned}
$$

The family $f_m$ is equi-Lipschitz, thus equi-continuous. $\{f_m(t); m \in N\} \subset \Delta$ is relatively compact.

By Ascoli-Arzela Theorem,we get the relative compactness of the family $f_m$ for the infinity norm topology. Let $f$ be a limit point of the family $f_m$, by the continuity of $b$ and $\Sigma$, we will get $f \in F(g)$.

As the limit points of the family $f_m$ stay in $F(g)$, we have

$$\forall\varepsilon\exists N\forall n \geqslant N \quad d(f_n, F(g)) \leqslant \varepsilon$$

i.e. $\limsup d(f_m, F(g)) = 0$. The compactness of $\{g, J(g) \leqslant \alpha\}$ allows us to conclude.

■

## Acknowledgments

The author would like to thank Yoann Offret and Michel Benaïm for their guidance and advice.

# Index

# Bibliography


[1] M. Benaïm. Recursive algorithms, urn processes, and the chaining number of chain recurrent sets. Ergodic Theory and Dynamical Systems, 18:53–87, 1998.

[2] M. Benaïm. Dynamics of stochastic approximation algorithms. Séminaire de probabilités de Strasbourg, 33:1–68, 1999.

[3] M. Benaïm and J.W. Weibull. Deterministic approximation of stochastic evolution in games. Econometrica, 71:873–903, 2003.

[4] Michel Benaïm and Nicole El Karoui. Promenade Aléatoire: Chaînes de Markov et Simulations; Martingales et Stratégies. Éditions de l'École Polytechnique, Palaiseau Cedex, 2004.

[5] Michel Benaïm, Josef Hofbauer, and William H. Sandholm. Robust permanence and impermanence for stochastic replicator dynamics. Journal of Biological Dynamics, 2:180–195, 2008.

[6] Ken Binmore, John Gale, and Larry Samuelson. Learning to be imperfect: The ultimatum game. Games and Economic Behavior, 8:56–90, 1995.

[7] Ken Binmore and Larry Samuelson. Muddling through: Noisy equilibrium selection. Journal of Economic Theory, 74:235–265, 1997.

[8] Jonas Björnerstedt and Jörgen W. Weibull. Nash equilibrium and evolution by imitation. In Kenneth J. Arrow et al., editors, The Rational Foundations of Economic Behavior, pages 155–181. St. Martin's Press, New York, 1996.

[9] F. F. Bonsall. Linear operators in complete positive cones. Proc. London Math. Soc., 1958.

[10] P. Cattiaux, P. Collet, A. Lambert, S. Martinez, S. Méléard, and J. San Martin. Quasi-stationary distributions and diffusion models in population dynamics. Ann. Probab, 37(5):1926–1969, 2009.

[11] Patric Cattiaux. Long time behavior of markov processes.

[12] Pierre Collet, Servet Martínez, and Jaime San Martín. Quasi-stationary distributions. Markov chains, diffusions and dynamical systems. Springer, 2013.

[13] C.C. Conley. Isolated Invariant Sets and the Morse Index. American Mathematical Society, 1978.

[14] J. N. Darroch and E. Seneta. On quasi-stationary distributions in absorbing continuous-time finite Markov chains. J. Appl. Probability, 4:192–196, 1967.

[15] JN Darroch and E. Seneta. On quasi-stationary distributions in absorbing discrete-time finite Markov chains. Journal of Applied Probability, 2(1):88–100, 1965.

[16] Amir Dembo and Ofer Zeitouni. Large Deviations Techniques and Applications. Springer, New York, second edition, 1998.

[17] P. Dupuis and R. S. Ellis. A Weak Convergence Approach to the Theory of Large Deviations. Wiley, New York, NY, 1997.

[18] Richard Durrett. Stochastic Calculus: A Practical Introduction. CRC Press, Boca Raton, FL, 1996.

[19] E.B. Dynkin. Markov processes. Number vol. 2 in Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen mit besonderer Berücksichtigung der Anwendungsgebiete. Springer-Verlag, 1965.

[20] M. Faure and S. J. Schreiber. Quasi-stationary distributions for randomly perturbed dynamical systems. arXiv:1101.3420v1, Submitted, 2011.

[21] M. I. Freidlin and A. D. Wentzell. Random Perturbations of Dynamical Systems. Springer, New York, second edition, 1998.

[22] J.F.L. Gall. Mouvement Brownien, Martingales Et Calcul Stochastique. Mathématiques et Applications. Springer Berlin Heidelberg, 2012.

[23] I. Girsanov. Strongly-feller processes i. general properties. Theory of Probability and Its Applications, 5(1):5–24, 1960.

[24] J. Hofbauer and K. Sigmund. Evolutionary Dynamics and Population Dynamics, 1998.

[25] Nobuyuki Ikeda and Shinzo Watanabe. Stochastic Differential Equations and Diffusion Processes. North Holland, Amsterdam, 1981.

[26] Ioannis Karatzas and Steven E. Shreve. Brownian motion and stochastic calculus. Springer, 2 edition, 1991.

[27] W. O. Kermack and A. G. McKendrick. A contribution to the mathematical theory of epidemics. Proc. Roy. Soc. Lond. A, 115:700–721, 1927.

[28] Y. Kifer. Random perturbations of dynamical systems. 1988.

[29] Smith H. L. and Thieme H. R. Dynamical Systems and Population Persistence. AMS Graduate Studies in Mathematics, 2010.

[30] A. J. Lotka. Elements of Physical Biology. Williams and Witkins, Baltimore, 1925.

[31] B. Marmet. Quasi-stationary distributions for stochastic approximation algorithms with constant step size. arXiv:1303.7081, Submitted, 2013.

[32] S. Méléard and D. Villemonais. Quasi-stationary distributions and population processes. arXiv:1112.4732v1, Submitted, 2011.

[33] S. P. Meyn and R. L. Tweedy. Markov Chains and Stochastic Stability. Springer, Berlin, second edition, 2005.

[34] S. Molchanov. Strong feller property of diffusion processes on smooth manifolds. Theory of Probability and Its Applications, 13(3):471–475, 1968.

[35] A. J. Nicholson and V. A. Bailey. The balance of animal populations. Proc. Zool. Soc. Lond., pages 551–598, 1935.

[36] J. R. Norris. Markov Chains. Cambridge University Press, Cambridge, 1997.

[37] Bernt Oksendal. Stochastic Differential Equations (3rd Ed.): An Introduction with Applications. Springer-Verlag New York, Inc., New York, NY, USA, 1992.

[38] Isle Royale National Park. Wolfs and mooses of isle royale.

[39] P. K. Pollett. Quasi-stationary distributions: A bibliography. 2012.

[40] P.E. Protter. Stochastic Integration and Differential Equations: Version 2.1. Applications of mathematics. U.S. Government Printing Office, 2004.

[41] D. Revuz and M. Yor. Continuous Martingales and Brownian Motion. Grundlehren der mathematischen Wissenschaften. Springer, 2001.

[42] William H. Sandholm. Population games and evolutionary dynamics. Economic Learning and Social Evolution. MIT Press, Cambridge, MA, 2010.

[43] R. L. Schilling and J. Wang. Strong Feller Continuity of Feller Processes and Semigroups. ArXiv e-prints, June 2010.

[44] S. J. Schreiber. Persistence despite perturbations for interacting populations. Journal of Theoretical Biology, 242:844–52, 2006.

[45] S. J. Schreiber. Persistence for stochastic difference equations: A mini-review, 2011.

[46] S. J. Schreiber, M. Benaïm, and K. A. S. Atchadé. Persistence in fluctuating environments. in review.

[47] E. Seneta. Quasi-stationary behaviour in the random walk with continuous time. Austral. J. Statist., 8:92–98, 1966.

[48] W. R. Thompson. La theorie mathematique de l'action des parasites entomophages et le facteur du hasard. Annales Faculte des Sciences de Marseille, 2:69–89, 1924.

[49] E. A. van Doorn and P. K. Pollett. Quasi-stationary distributions. Memorandum 1945, Department of Applied Mathematics, University of Twente, Enschede, June 2011.

[50] V. Volterra. Fluctuations in the abundance of a species considered mathematically. Nature, 118:558–560, 1926.

[51] P. Walters. An introduction to ergodic theory. Springer Verlag, 2000.

[52] H. W. Watson and F. Galton. On the probability of the extinction of families. Journal of the Anthropological Institute of Great Britain, 4:138–144, 1875.

[53] J.W. Weibull. Evolutionary game theory. Cambridge, Mass, MIT Press, 1995.

[54] A. M. Yaglom. Certain limit theorems of the theory of branching random processes. Doklady Akad. Nauk SSSR (N.S.), 56:795–798, 1947.