



**HAL**  
open science

**Quelques contributions en modélisation aléatoire :  
évolution dans des communautés naturelles, inhibition  
dans des processus de Hawkes, algorithmes stochastiques  
pour des mesures de risques**

Manon Costa

► **To cite this version:**

Manon Costa. Quelques contributions en modélisation aléatoire : évolution dans des communautés naturelles, inhibition dans des processus de Hawkes, algorithmes stochastiques pour des mesures de risques. Probability [math.PR]. Université toulouse 3 Paul Sabatier, 2025. tel-04910664

**HAL Id: tel-04910664**

**<https://theses.hal.science/tel-04910664v1>**

Submitted on 24 Jan 2025

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITÉ PAUL SABATIER TOULOUSE 3



Manuscrit présenté pour l'obtention de  
l'Habilitation à Diriger des Recherches par

---

**Manon COSTA**

QUELQUES CONTRIBUTIONS EN MODÉLISATION ALÉATOIRE :  
*évolution dans des communautés naturelles, inhibition dans des  
processus de Hawkes, algorithmes stochastiques pour des mesures de  
risques*



---

*Soutenue le 23 janvier 2025, à Toulouse,*

*après avis des rapporteurs*

Michel BENAÏM	Université de Neuchâtel (Suisse)
Anton BOVIER	Universität of Bonn (Allemagne)
Eva LOCHERBACH	Université Paris 1

*devant le jury formé de*

Michel BENAÏM	Université de Neuchâtel (Suisse)
Anton BOVIER	Universität of Bonn (Allemagne)
Laure COUTIN	Université Paul Sabatier
Nicolas LOEUILLE	Sorbonne Université
Eva LÖCHERBACH	Université Paris 1
Patricia REYNAUD-BOURET	Université Cote d'Azur
Amandine VÉBER	Université Paris Cité



---

## Remerciements

---

Écrire une HDR est un processus aléatoire, dont l'intensité n'est pas constante et qu'il n'est pas facile de faire converger en temps fini. Toutefois, cela permet de faire un point sur le chemin parcouru et de se retourner sur les différents travaux commencés, sur les idées avancées, sur les choses qui ont fonctionné (ou pas) et celles qui restent à tester. Je voudrais remercier ici celles et ceux qui ont eu un rôle particulier dans l'élaboration de ce manuscrit que ce soit par des discussions au détour d'un couloir, d'un tableau, d'une feuille, d'un café ou d'un verre.

Je voudrais commencer par remercier mes trois rapporteurs, Michel Benaim, Anton Bovier et Eva Löcherbach. Merci du temps que vous avez pris pour lire et évaluer mon travail, cela m'est d'autant plus précieux que vos travaux ont souvent été (et, je pense, seront encore) une source d'inspiration pour mes recherches et des références dans lesquelles apprendre de nouveaux outils. Merci aux membres de mon jury Patricia Reynaud-Bouret, Laure Coutin, Nicolas Loeuille et Amandine Véber, c'est un très grand plaisir que vous ayez accepté de participer à ma soutenance. Amandine qui a suivi mes débuts en probabilités appliquées à l'ENS et m'avais aidé dans ma recherche d'un stage en Nouvelle-Zélande ; Nicolas avec qui j'ai un immense plaisir à collaborer dans différents projets et dont j'admire l'insondable connaissance de la littérature en écologie ; et Patricia qui m'a fait découvrir les processus de Hawkes à Aussois. Merci enfin à Laure d'être la représentante toulousaine de ce jury et dont les conseils et l'oreille bienveillante me sont toujours précieux. Par ailleurs, je tiens à remercier ici Patrick Cattiaux pour son soutien depuis mon arrivée à Toulous et toutes nos discussions passées et j'espère à venir, qui me guident et me poussent tout à la fois.

Depuis toujours, je me suis intéressée à la biologie dans un sens assez large... si j'ai pris des chemins détournés, elle reste une source quasi inépuisable d'intérêt dans mes recherches, nourrie par cet incroyable rassemblement de chercheur.ses qu'est la Chaire MMB (pour Mathématiques et Modélisation de la Biodiversité). Je profite de cette occasion pour remercier Sylvie Méléard qui l'anime et la porte depuis tant d'années, et celles et ceux qui s'investissent dans l'organisation annuelle des rencontres et de l'école d'Aussois dont les cours de recherche en font chaque année un rendez-vous quasi-incontournable ! Merci en particulier à Diala Abu Awad, Violaine Llaurens, Sylvain Billard pour nos nombreuses discussions, votre enthousiasme et toutes vos explications sur le monde du vivant.

Je voudrais ensuite remercier tous mes co-auteurices avec qui j'ai la chance de travailler dans mes recherches, dans une atmosphère toujours détendue, simple et dans laquelle toutes les questions sont les bienvenues : Carl Graham, Laurence Marsalle et Viet Chi Tran d'avoir

utilisé toutes les craies de couleurs disponibles pour nos couplages de Hawkes ; Anna Bonnet, Félix Cheysson, Céline Duval, Lorick Huang et Eric Luçon, de vous être embarqués avec moi dans l'ANR HAPPY et de m'accueillir régulièrement dans vos bureaux respectifs lors de mes séjours parisiens ; Pete Czuppon et Raphaël Forien pour votre acharnement à finir notre travail entamé depuis bien trop longtemps à cause de systèmes dynamiques féroces ; Jérôme Fehrenbach de venir me chercher pour les questions statistiques ; Sébastien Gadat, Thierry Klein, Lorick Huang et Gersende Fort, avec qui j'améliore mes connaissances en algorithmes stochastiques et optimisation ; Sepideh Mirrahimi pour nos discussions entre EDP, proba et génétique. Merci à Laetitia Colombani et Anthony Muraro d'avoir traversé l'aventure doctorale avec Patrick Cattiaux, Pascal Maillard et moi ; pour ma part je suis ravie d'avoir fait mes débuts d'encadrante avec vous quatre. En tout dernier, je voudrais remercier celles qui me re-motivent depuis les couloirs du CMAP : Camille Coron, Hélène Leman et Charline Smadi, pour toutes nos sessions de travail partagées depuis la thèse, et tous les restos qui les agrémentent. J'espère que nous continuerons de nombreuses années à échanger autour des maths et de la vie.

Cela fait maintenant 9 ans que je suis arrivée à l'Institut de Mathématiques de Toulouse, laboratoire dont j'affectionne les particularités (son isolation thermique notamment) et les petites habitudes. Je voudrais remercier tout le pôle administratif du laboratoire et particulièrement Isabelle Guichard, Karine Wasek, Jean-Marc Bano, Céline Rocca et Tamara Azaiez, qui sont toujours là pour m'aider dans la réalisation de diverses procédures. Merci également au pôle informatique et particulièrement à Mathys Astruc de me déboguer avec bonne humeur. Un grand merci à mes collègues de l'étage, Fanny, Grégory, Jérémy H. (et sa clé USB), Pierre B., Francis, Stefan, Franck, Paul, avec qui je partage mes (nombreuses) railleries ; à Sepideh et maintenant Lorick, d'avoir partagé mon bureau et à l'équipe EDP de m'accepter à déjeuner malgré mes tendances probabilistes. Si mon bureau se trouve du côté déterministe des portes coupe-feu, la salle café est de l'autre côté, merci à Pierre Petit de gérer l'approvisionnement, à Michel Pain de tenir à jour l'indispensable calendrier papier, à Bastien Mallein d'avoir initié l'ANR d'équipe MBAPP et à tous ceux dont les portes ouvertes sur le couloir, invitent régulièrement à la discussion. Parce qu'il n'y a pas que la recherche dans la vie d'un enseignant chercheur, je voudrais remercier tous les collègues avec qui je partage mes enseignements, les TPs avec ou sans IA et l'émerveillement de chaque instant devant les méthodes de calcul des étudiant.es. Enfin, au fil des années, j'ai eu la chance de participer à un certain nombre d'aventures collectives, aussi je voudrais remercier mes collègues d'Opération Postes que j'ai quitté un peu avant d'entamer la rédaction de ce manuscrit ; le comité égalité du laboratoire pour faire bouger doucement les choses et pour toutes nos conversations stimulantes ; et l'équipe d'organisation de la biennale de la SMAI 202? dont la tenue en sortie de Covid demandait une dose certaine d'optimisme.

Merci enfin à ceux d'entre vous qui sont devenus plus que des collègues ; à Sébastien qui m'écoute en toutes circonstances et a toujours un DL à l'ordre supérieur à proposer pour me remonter le moral ; à Thierry et notre L2 bientôt validée ; à toutes les LSM improvisées ; à Fanny pour ton oreille toujours attentive ; à Tristan pour les discussions tardives et de m'accompagner dans mes sorties culturelles ; et finalement à Lorick, Guillaume, Jérémy H., Adrien, Pierre N., Juliette, Manue, Max, Bastien, Michel pour toutes les bières partagées.

Je profite aussi de ces lignes pour remercier celles et ceux qui ne font pas partie de ce monde de la recherche, mais qui m'aident à garder un équilibre au quotidien : les crossfiteurs de Kasta et tout particulièrement la team du midi (venez essayer d'ailleurs, si l'envie vous prend :), Amandine, Giacomo, Amaryllis et Yves pour les goûters réguliers ; merci aux HX4 pour les coinches et autres glandouilles partagées depuis tant d'années.

Comme le veut la tradition, je termine par ceux qui sont les plus importants. Merci à ma famille pour votre soutien moral et logistique infallible, et l'intérêt que vous portez à mon travail. Merci à Matthieu, Agnès et Jérémy sans qui rien n'aurait le même sens. Merci de me soutenir, de me supporter au quotidien, d'accepter mes absences régulières, de me suivre dans mes idées folles... J'espère que nous continuerons à construire ensemble des projets petits et grands.



# Contents

<b>Références et organisation générale du manuscrit</b>	<b>9</b>
<b>1 Evolution of mating preferences and their consequences on communities</b>	<b>13</b>
1.1 Introduction - Mating preferences and evolution . . . . .	13
1.2 The effect of assortative mating on speciation . . . . .	14
1.3 Emergence of assortative mating . . . . .	20
1.4 Dissassortative mating and structure of communities . . . . .	27
<b>2 Perspectives on ecological models</b>	<b>37</b>
2.1 Matching types prey-predators models . . . . .	37
2.2 Oncological viruses . . . . .	43
2.3 Perspectives on impact of drift in some population genetics models . . . . .	45
<b>3 Hawkes processes with inhibition</b>	<b>47</b>
3.1 Introduction . . . . .	47
3.2 A renewal structure for inhibited Hawkes processes . . . . .	51
3.3 Stability of a discrete time Hawkes process with inhibition and finite memory	57
3.4 Perspectives . . . . .	62
<b>4 Stochastic algorithms and application to financial data</b>	<b>65</b>
4.1 Introduction . . . . .	65
4.2 Different coupled algorithms . . . . .	68
4.3 CV@R penalized portfolio optimization with biased stochastic mirror descent	74
4.4 Perspectives . . . . .	81





---

## Références et organisation générale du manuscrit

---

Cette habilitation à diriger des recherches présente une synthèse de mes travaux de recherches effectués à l'Institut de Mathématiques de Toulouse, depuis 2016.

Mes travaux s'intéressent à l'étude de différents processus stochastiques trouvant leurs motivations dans des questions aussi variées que l'étude des communautés écologiques et leur structuration; la modélisation de l'inhibition dans la répartition des occurrences temporelles de certains phénomènes, et l'estimation de mesures de risques via des algorithmes stochastiques. Si ces questions ont l'air éloignées (et elles le sont), elles m'ont permis d'apprendre une large variété de techniques probabilistes et de les appliquer afin d'apporter des réponses aux problématiques initiales.

Ce manuscrit est composé de 4 chapitres regroupant chacun un ensemble de travaux sur une thématique précise et des perspectives de recherche associées.

Une partie importante de mes travaux de recherche se situe à l'interface entre les mathématiques et l'écologie. Au cours de ma thèse, je me suis focalisée sur les interactions de type proies-prédateurs et sur la façon dont la sélection naturelle modifie la structure des communautés en agissant sur les phénotypes des proies et prédateurs qui définissent la force de leurs interactions. Si ces questions sont toujours présentes dans mes recherches, j'ai depuis étudié d'autres types de d'interactions.

Le chapitre 1, concerne la question de l'évolution des préférences sexuelles dans des populations naturelles. Je me suis particulièrement intéressée à des phénomènes d'attachement préférentiel (*assortative mating* et *disassortative mating*) dans des populations ayant une reproduction sexuée et à l'impact de l'évolution de ces stratégies sur la structure des communautés. Les travaux suivants seront présentés dans ce chapitre :

- [CCLS18] A stochastic model for speciation by mating preference, C. Coron, M.C., H. Leman et C. Smadi. *Journal of Mathematical Biology*, 76:1421-1463 (2018)
- [CCL<sup>+</sup>21] Emergence of homogamy in a two-loci stochastic population model , C. Coron, M.C., F. Laroche, H. Leman and C. Smadi. *ALEA*, (Volume 18, 2021).
- [CCL<sup>+</sup>23] Origin and persistence of polymorphism in loci targeted by disassortative preference: a general model , C. Coron, M.C., H. Leman, V. Llaurens, C. Smadi. *Journal of Mathematical Biology*, 86(1):4 (2023)

Le deuxième chapitre regroupe des questions sur lesquelles je travaille ou que je souhaiterai étudier par la suite sur différents modèles motivés par la biologie en générale. Plus particulièrement:

- un modèle de coévolution de communautés proies-prédateurs avec des phénotypes "appariés" et mutations rapides, sur lequel je travaille avec Pete Czuppon (WWU Münster) et Raphaël Forien (INRAE Avignon).
- un modèle d'évolution d'une infection virale dans des cellules en expansion dans la continuité de l'article [CCF22], utilisant des processus de croissance-frangmentation aléatoires,
- des questions sur l'impact de la stochasticité dans les modèles de génétique des populations et plus particulièrement la mesure de l'écart entre les modèles en taille de population infinie et leurs versions incluant un tirage aléatoire à chaque reproduction, avec Sepideh Mirrahimi (IMAG), Sylvain Gandon (CEFE) d'une part et Thomas Aubier (EDB) d'autre part.

Le troisième chapitre 3 regroupe des résultats sur l'étude de processus de Hawkes non linéaires incluant de l'inhibition. La question centrale de ces travaux est de mieux appréhender le rôle de l'inhibition dans la stabilité de ces processus. Ces questions ont été l'occasion de pour moi de co-diriger deux thèses, celle de Laetitia Colombani avec Patrick Cattiaux et celle d'Anthony Muraro avec Pascal Maillard. Je présenterai à la fin de ce chapitre des perspectives de travail qui forment une partie du programme du projet ANR HAPPY (JCJC, 2024-2028) dont je suis la porteuse.

- [CGMT20] Renewal in Hawkes processes with self-excitation and inhibition , M.C., C. Graham, L. Marsalle and V.C. Tran. *Advances in Applied Probability* (Volume 52, Issue 3, 2020)
- [CCC22] Limit Theorems for Hawkes processes including inhibition , P. Cattiaux, L. Colombani, M.C. *Stochastic Processes and their Applications* 149:404-426 (2022)
- [CMM24a] Complete characterization of stability of a two-parameter discrete-time Hawkes process with inhibition, M.C., P. Maillard, A. Muraro. *Journal of Applied Probability*, published online, 2024:1-22
- [CMM24b] Stability of discrete-time Hawkes process with inhibition: towards a general condition, M.C., P. Maillard, A. Muraro. Arxiv preprint 2409.01660.

Le quatrième chapitre s'articule autour de différents résultats sur des algorithmes stochastiques de type Robbins Monro. Tout d'abord, je présenterai des résultats théoriques sur des algorithmes permettant l'estimation conjointe du quantile et du super-quantile d'une distribution de probabilité. Ensuite, je détaillerai une application de ces méthodes à un problème d'optimisation stochastique et des perspectives sur l'utilisation de ces algorithmes en analyse de sensibilité. Ce chapitre est basé sur les travaux suivants

- [BCG21] Stochastic approximation algorithms for superquantile estimation , B. Bercu, M.C. and S. Gadat. *Electron. J. Probab.*, 26:1-29 (2021)
- [CG21] Non asymptotic controls on a recursive superquantile approximation, M.C., S. Gadat. *Electron. J. Statist.* 15(2): 4718-4769 (2021).
- [CGH24] CV@R penalized portfolio optimization with biased stochastic mirror descent, M.C., S. Gadat, L. Huang. to appear in *Mathematical Finance* (2024).

Pour conclure, je voudrais dire quelques mots sur les travaux suivants que j'ai choisi de ne pas présenter en détail afin de limiter la taille de ce manuscrit et de préserver la cohérence thématique des travaux présentés. Beaucoup de ces travaux proviennent de collaborations interdisciplinaires avec des biologistes, écologues ou médecins. Cette interdisciplinarité est une part de mon travail qui m'apporte au moins autant de plaisir qu'elle nécessite de temps pour apprendre à échanger entre chercheurs de formations différentes.

- [GCA<sup>+</sup>19] Multiparametric analysis of CD8+ T cell compartment phenotype in chronic lymphocytic leukemia reveals a signature associated with progression toward therapy , P. Gonnord, M.C., A. Abreu, M. Peres, L. Ysebaert, S. Gadat, S. Valitutti. *OncoImmunology*, (2019)
- [CGGR19] Cytometry inference through adaptive atomic deconvolution, M.C., S. Gadat, P. Gonnord et L. Risser. *Journal of Nonparametric Statistics* 2019, Vol. 31, No. 2, pp506-547
- [CEM21] Survival criterion for a population subject to selection and mutations ; Application to temporally piecewise constant environments , M.C., C. Etchegaray and S. Mirrahimi. *non linear Analysis: Real Worls Applications* (Volume 59, June 2021)
- [CCF22] Slow-fast model and therapy optimization for oncolytic treatment of tumours, P. Cordelier, M.C., J. Fehrenbach. *Bulletin of Mathematical Biology*, 2022 May 10;84(6):64
- [CCC23] Asymptotic deviation bounds for cumulative processes , P. Cattiaux, L. Colombani, M.C. *Stochastic Processes and Their Applications* (2023) 163:85–103
- [NCL21] Implication of drift and rapid evolution on negative niche construction , Phuong Linh Nguyen, M.C., N. Loeuille. *Soumis*.
- [BCF<sup>+</sup>23] Influence of sex-limited mimicry on extinction risk in Aculeata: a theoretical approach, M. Boutin, M.C., C. Fontaine, A. Perrard, V. Llaurens. *Peer Community Journal*, Volume 3 (2023), article no. e113.

Enfin, les articles suivants, écrits au cours de ma thèse ne seront pas présentés.

- [CHLM16] Stochastic eco-evolutionary model of a prey-predator community, M.C., N. Loeuille, C. Hauzy et S. Méléard. *Journal of Mathematical Biology*, (2016), 72(3) pp573-622
- [Cos16] A piecewise deterministic model for prey-predator communities, M.C. *Annals of Applied Probability* 2016, Vol. 26, No. 6, 3491-3530



## Chapter 1

---

# Stochastic models for Ecology I

## Evolution of mating preferences and their consequences on communities

---

An important part of my research aims at constructing probabilistic models and associated mathematical techniques in order to understand the evolutionary dynamics of complex ecological or biological systems. For these questions I work both in collaboration with biologists, ecologists or doctors in order to understand the question of interest and to build adequate models, and with mathematicians aiming at deriving theoretical properties of the previous models (or of simplified versions). My work focuses mainly on individual-based models and their scaling limits, as they allow to observe the impact of basic mechanisms on the population at different scales.

In this first chapter, I present a series of three works in collaboration with Camille Coron, H el ene Leman and Charline Smadi and two ecologists Fabien Laroche and Violaine Llaurens, dedicated to the evolution of mating preferences in communities.

### 1.1 Introduction - Mating preferences and evolution

Evolution, from a darwinian perspective, is based on three mechanisms: heredity whereby characteristic of individuals are transferred to the offspring, mutations which generates individual variations and natural selection. Natural selection acts on individuals through their ability to reproduce and survive in their environment. It also creates a constant feedback loop between individuals and their environment.

Sexual selection plays an important role in evolutionary theory [JR09]. Originally, the existence of sexual selection was the main hypothesis to explain the survival of rare or disadvantageous phenotypes, such as bulky or colourful ornamental features. Today, these mechanisms are more widely studied, as their role in evolution seems to be more important, particularly for trait selection and species diversification [Lan81, Ros17].

Mate choice is a process that leads to a bias in reproductive investment towards individuals

with certain traits, known as 'mating cues'. These selection biases can be expressed before or after reproduction and involve more diverse mechanisms than just mate discrimination, such as mechanical constraints, molecular mechanisms, temporal synchrony (flowering dates, different migrations) or spatial synchrony.

Mate choice also comes at a high cost for the individual, generating direct selection on sexual preferences and limiting their evolution. For example, choosing breeding partners can lead to a loss of fitness, as many reproductive opportunities are lost. In addition, the implementation of choice mechanisms requires physiological, morphological or behavioural changes that generate metabolic costs (e.g. incompleteness) [Ros17]. However, despite these costs, sexual preferences are maintained in many species, suggesting the existence of direct or indirect positive effects.

This chapter focuses mainly on the issue of assortative and disassortative mating preferences, i.e. when individuals prefer to mate with similar (or different) individuals, and their implications for community structure. In the first section, we examine the conditions under which assortative mating preferences at a neutral locus can lead to sympatric speciation, the process by which new species evolve from a single ancestral species while sharing the same space and resources. We then consider how assortative mating can arise in a random mating community when the mating cue is evolutionarily neutral. In the third section, we consider disassortative mating behaviour and study the emergence of phenotypic diversity at loci targeted by preference. In these different works, we mainly study evolution through the question of invasion of a given mutant individual. This view corresponds to a single step of evolution in the adaptive dynamics framework introduced by Geritz, Metz et al, Dieckmann and Law [MGM<sup>+</sup>96, DL96]. In this framework, the time scale of demographic dynamics is much faster than the time scale of mutations. As a consequence, when a mutation occurs, the single mutant individual evolves in a resident population at its demographic equilibria. Mathematically, we use the approach developed by Champagnat [Cha06] which is well adapted for the study of scaling limits of birth and death processes in a large population with rare mutations. The main idea is to split the dynamics into different phases: first the study of the population in a large population limit without mutation allows to define the resident equilibrium state by comparison with a dynamical system. Then once a mutant appears, its population can be compared with a (multi-type) branching process as long as it reaches either 0 or a macroscopic fraction of the population. Finally, in the case where the mutant population become macroscopic, the large population approximation changes and leads to a different equilibrium, see Figure 1.1. These approximations require first to have a slow mutation rate with respect to the scaling parameter of the population size so that only a single mutant type present and second to have a good knowledge of the limiting behaviour of systems of ODE that describe the dynamics in large population.

## 1.2 The effect of assortative mating on speciation

In this section, I present the work [CCLS18] in collaboration with Camille Coron, H el ene Leman and Charline Smadi that introduces and studies rigorously a stochastic model for speciation by assortative mating. Assortative mating, or mating preference, is a form of sexual

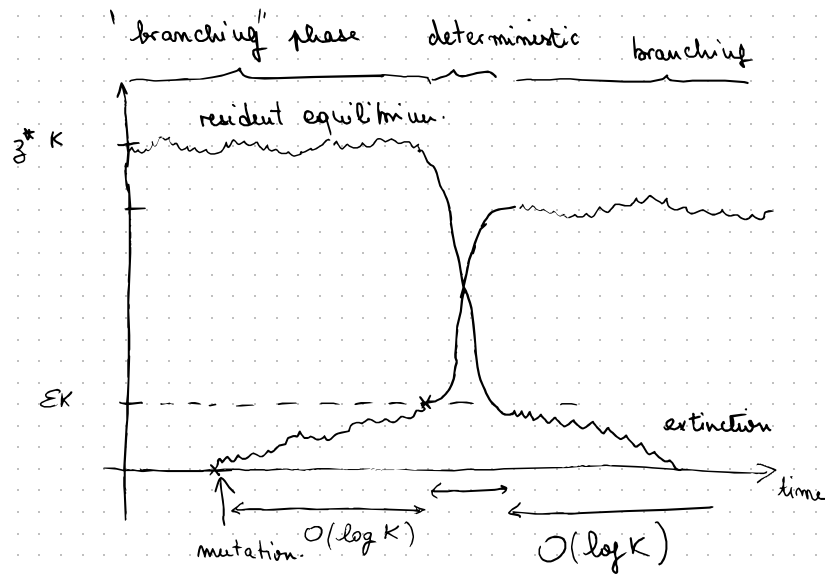


Figure 1.1: A typical example of the different phases of a mutant invasion

selection where individuals with similar genotypes have a higher reproductive success when they mate among themselves than with individuals with different genotypes. As presented in [Rit07, MMOD12], this form of selection plays a very important role in speciation.

### A simple model on two similar patches

We consider a population of haploid individuals characterized by their genotype at a given locus and by their position in space divided in two patches. More precisely, each individual carries an allele belonging to the genetic type space  $\mathcal{A} := \{A, a\}$ , and lives in a patch  $i$  in  $\mathcal{I} = \{1, 2\}$ . We denote by  $\mathcal{E} = \mathcal{A} \times \mathcal{I}$  the type space, by  $(\mathbf{e}_{\alpha,i}, (\alpha, i) \in \mathcal{E})$  the canonical basis of  $\mathbb{R}^{\mathcal{E}}$ , and by  $\bar{\alpha}$  the complement of  $\alpha$  in  $\mathcal{A}$ . The current state of the population is then given by  $\mathbf{n} = (n_{\alpha,i}, (\alpha, i) \in \mathcal{E})$  where  $n_{\alpha,i}$  the current number of  $\alpha$ -individuals in the patch  $i$ .

The dynamics in time is modeled by a multi-type birth-and-death process with competition taking values in  $\mathbb{N}^{\mathcal{E}}$ . We assume that the alleles are ecologically neutral in the sense that individuals with different genotypes are not characterized by different adaptations to the environment or by different resource preferences. However, individuals reproduce sexually according to mating preferences that depend on their genotype: two individuals with the same genotype have a higher probability of mating successfully. In addition, individuals can migrate from one patch to another, at a rate that depends on the number of individuals carrying the other genotype and living in the same patch.

The birth rate is computed as follows: at a rate  $B > 0$  any individual encounters another individual uniformly at random in its patch. Hence the probability of encountering an



individual of genotype  $\alpha$  in the patch  $i$  is

$$\frac{n_{\alpha,i}}{n_{\alpha,i} + n_{\bar{\alpha},i}},$$

Then, the probability that this encounter leads to a successful mating with a viable offspring is  $b\beta/B \leq 1$  if the two individuals carry the same genotype and  $b/B < 1$  otherwise. The parameter  $b > 0$  scales the individual birth rate while the parameter  $\beta > 1$  represents the "mating preference" and can be interpreted as follows: two mating individuals have a probability  $\beta$  times larger to give birth to a viable offspring if they carry the same allele  $\alpha$ . This modelling of mating preferences (that are directly determined by the genome of each individual) is biologically relevant, considering [HTPW97] or [HS05] for instance. As a consequence, the birth rate of an  $\alpha$ -individual in the patch  $i$  writes

$$\begin{aligned} \lambda_{\alpha,i}(\mathbf{n}) &= b \left( n_{\alpha,i} \beta \frac{n_{\alpha,i}}{n_{\alpha,i} + n_{\bar{\alpha},i}} + \frac{1}{2} n_{\alpha,i} \frac{n_{\bar{\alpha},i}}{n_{\alpha,i} + n_{\bar{\alpha},i}} + \frac{1}{2} n_{\bar{\alpha},i} \frac{n_{\alpha,i}}{n_{\alpha,i} + n_{\bar{\alpha},i}} \right) \\ &= b n_{\alpha,i} \frac{\beta n_{\alpha,i} + n_{\bar{\alpha},i}}{n_{\alpha,i} + n_{\bar{\alpha},i}}. \end{aligned} \quad (1.1)$$

In the same way, the death rate of  $\alpha$ -individuals in the patch  $i$  includes the effect of a natural death rate  $d$  and of logistic competition  $c$

$$d_{\alpha,i}^K(\mathbf{n}) = \left( d + \frac{c}{K} (n_{\alpha,i} + n_{\bar{\alpha},i}) \right) n_{\alpha,i}, \quad (1.2)$$

where  $K$  is an integer accounting for the quantity of available resources or space. This parameter is related to the concept of carrying capacity, which is the maximum population size that the environment can sustain indefinitely, and is consequently a scaling parameter for the size of the community. The individual intrinsic death rate  $d$  is assumed to be non-negative and less than  $b$ :

$$0 \leq d < b. \quad (1.3)$$

The death rate definition (1.2) implies that all the individuals are ecologically equivalent: the competition pressure  $c > 0$  does not depend on the alleles  $\alpha$  and  $\alpha'$  carried by the two individuals involved in an event of competition for food or space.

Last, the migration of  $\alpha$ -individuals from patch  $\bar{i} = \mathcal{I} \setminus \{i\}$  to patch  $i$  occurs at a rate

$$\rho_{\alpha,\bar{i} \rightarrow i}(\mathbf{n}) = p \left( 1 - \frac{n_{\alpha,\bar{i}}}{n_{\alpha,\bar{i}} + n_{\bar{\alpha},\bar{i}}} \right) n_{\alpha,\bar{i}} = p \frac{n_{\alpha,\bar{i}} n_{\bar{\alpha},\bar{i}}}{n_{\alpha,\bar{i}} + n_{\bar{\alpha},\bar{i}}}, \quad (1.4)$$

with  $p > 0$  (see Figure 1.2). The individual migration rate of  $\alpha$ -individuals is proportional to the frequency of  $\bar{\alpha}$ -individuals in the patch. It reflects the fact that individuals prefer to be in an environment with a majority of individuals of their own type. In particular, if all the individuals living in a patch are of the same type, there is no more migration outside this patch. This particular form of mating success dependent dispersal has also been studied by [PK97] for a continuous space. In natural populations, [CBGB<sup>+</sup>10] study the dispersal behaviour of the banded damselflies *Calopteryx splendens* that display a lek mating system. They observe that females move to find a suitable mate, and that they disperse less when the sex-ratio is male biased. This is in agreement with our hypothesis that individuals migration

rate is a decreasing function of the frequency of suitable mates. More generally, correlations between male dispersal and mating success have been empirically observed (see the articles by [Sch88] or [HWE<sup>+</sup>07] for instance).

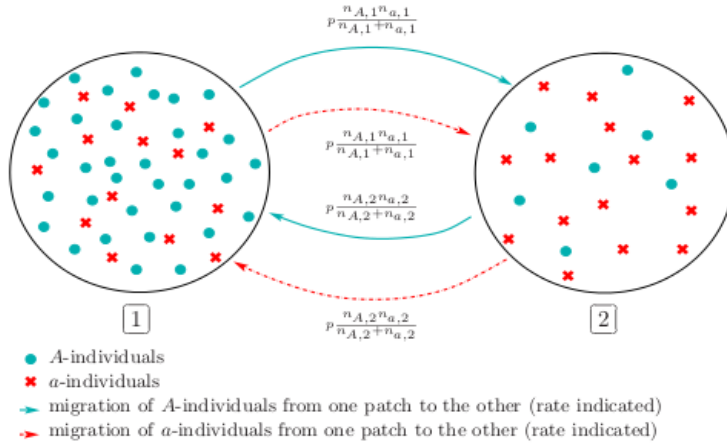


Figure 1.2: Migrations of A- and a-individuals between the patches.

**Remark 1.2.1.** *In contrast with our model, most papers about sexual preferences (see for instance the works by [GB98, MGG02, BS06, Ser10]) use generational models with infinite population size, and study the evolution through time of the frequency of each genotype. As a consequence, they express the population dynamics in terms of a table describing the frequencies of mating at each generation.*

The community is therefore represented at all time  $t \geq 0$  by a stochastic process with values in  $\mathbb{R}^{\mathcal{E}}$ :

$$(\mathbf{N}^K(t), t \geq 0) = (N_{\alpha,i}^K(t), (\alpha, i) \in \mathcal{E}, t \geq 0),$$

whose transitions are, for  $\mathbf{n} \in \mathbb{N}^{\mathcal{E}}$  and  $(\alpha, i) \in \mathcal{E}$ :

$$\begin{aligned} \mathbf{n} &\longrightarrow \mathbf{n} + \mathbf{e}_{\alpha,i} && \text{at rate } \lambda_{\alpha,i}(\mathbf{n}), \\ &\longrightarrow \mathbf{n} - \mathbf{e}_{\alpha,i} && \text{at rate } d_{\alpha,i}^K(\mathbf{n}), \\ &\longrightarrow \mathbf{n} + \mathbf{e}_{\alpha,i} - \mathbf{e}_{\alpha,\bar{i}} && \text{at rate } \rho_{\bar{i} \rightarrow i}(\mathbf{n}). \end{aligned}$$

From the standard results of Ethier and Kurtz (Chapter 11, [EK86]) we know that as soon as the initial population sizes  $(N_{\alpha,i}^K(0), (\alpha, i) \in \mathcal{E})$  are of order  $K$ , then the rescaled stochastic process

$$(\mathbf{Z}^K(t), t \geq 0) = (Z_{\alpha,i}^K(t), (\alpha, i) \in \mathcal{E}, t \geq 0) = \left( \frac{\mathbf{N}^K(t)}{K}, t \geq 0 \right),$$

converges in law as  $K \rightarrow \infty$  uniformly on bounded time intervals to the solution of the dynamical system

$$\begin{cases} \frac{d}{dt} z_{A,1}(t) = z_{A,1} \left[ b \frac{\beta z_{A,1} + z_{a,1}}{z_{A,1} + z_{a,1}} - d - c(z_{A,1} + z_{a,1}) - p \frac{z_{a,1}}{z_{A,1} + z_{a,1}} \right] + p \frac{z_{A,2} z_{a,2}}{z_{A,2} + z_{a,2}} \\ \frac{d}{dt} z_{a,1}(t) = z_{a,1} \left[ b \frac{\beta z_{a,1} + z_{A,1}}{z_{A,1} + z_{a,1}} - d - c(z_{A,1} + z_{a,1}) - p \frac{z_{A,1}}{z_{A,1} + z_{a,1}} \right] + p \frac{z_{A,2} z_{a,2}}{z_{A,2} + z_{a,2}} \\ \frac{d}{dt} z_{A,2}(t) = z_{A,2} \left[ b \frac{\beta z_{A,2} + z_{a,2}}{z_{A,2} + z_{a,2}} - d - c(z_{A,2} + z_{a,2}) - p \frac{z_{a,2}}{z_{A,2} + z_{a,2}} \right] + p \frac{z_{A,1} z_{a,1}}{z_{A,1} + z_{a,1}} \\ \frac{d}{dt} z_{a,2}(t) = z_{a,2} \left[ b \frac{\beta z_{a,2} + z_{A,2}}{z_{A,2} + z_{a,2}} - d - c(z_{A,2} + z_{a,2}) - p \frac{z_{A,2}}{z_{A,2} + z_{a,2}} \right] + p \frac{z_{A,1} z_{a,1}}{z_{A,1} + z_{a,1}}. \end{cases} \quad (1.5)$$

A detailed study of the long time behaviour of solutions of (1.5) is provided in [CCLS18]. We are particularly interested in conditions that ensure the convergence of the solution of (1.5) to a speciation equilibrium, i.e. an equilibrium where one patch is filled with individuals  $A$  and the other with  $a$ .

Let us introduce the parameter

$$\zeta := \frac{\beta b - d}{c}, \quad (1.6)$$

that corresponds to the equilibrium of the  $\alpha$ -population size for the dynamical system (1.5), in a patch with no  $\bar{\alpha}$ -individuals and no migration ( $p = 0$ ); and define the subset of  $\mathbb{R}_+^{\mathcal{E}}$

$$\mathcal{D} := \{\mathbf{z} \in \mathbb{R}_+^{\mathcal{E}}, z_{A,1} - z_{a,1} > 0, z_{a,2} - z_{A,2} > 0\}. \quad (1.7)$$

**Theorem 1.2.2.** *There exists an explicit  $p_0 > 0$  such that for all  $p < p_0$ , any solution to (1.5) starting from  $\mathcal{D}$  converges to the equilibrium  $(\zeta, 0, 0, \zeta)$ .*

*Symmetrical results hold for the equilibria  $(0, \zeta, \zeta, 0)$ ,  $(\zeta, 0, \zeta, 0)$  and  $(0, \zeta, 0, \zeta)$  when changing initial conditions.*

This result classifies the trajectories except for initial conditions on the lines  $z_{A,1} = z_{a,1}$  and  $z_{a,2} = z_{A,2}$ . Notice that the limit reached depends on which genotype is in the majority in each patch, since the subset  $\mathcal{D}$  is invariant under the dynamical system (1.5). Secondly, when  $p = 0$ , the results of Theorem 1.2.2 can be proven easily since the two patches are independent from each other. The difficulty is thus to prove the result when  $p > 0$ . Our strategy first relies on proving that trajectories enter an invariant set in which populations on each patch are lower and upper bounded. From this set, we find a Lyapunov function for the dynamics. Our argument allows us to deduce an explicit constant  $p_0$  under which we have convergence to an equilibrium with reproductive isolation between patches. However, we are not able to deduce a rigorous result for all  $p$ . Indeed, when  $p$  increases, there is more mixing between the two patches which is difficult to construct the invariant set. Nevertheless numerical simulations suggest that the result stays true.

Let us now introduce our main result on the probability and the time needed for the stochastic process  $\mathbf{N}^K$  to reach a neighborhood of the equilibria  $(\zeta, 0, 0, \zeta)$ .

**Theorem 1.2.3.** *Assume that  $\mathbf{Z}^K(0)$  converges in probability to a deterministic vector  $\mathbf{z}^0$  belonging to  $\mathcal{D}$ , with  $(z_{a,1}^0, z_{A,2}^0) \neq (0, 0)$ . Introduce the following bounded set depending on  $\varepsilon > 0$ :*

$$\mathcal{B}_\varepsilon := [(\zeta - \varepsilon)K, (\zeta + \varepsilon)K] \times \{0\} \times \{0\} \times [(\zeta - \varepsilon)K, (\zeta + \varepsilon)K],$$

and  $T_{\mathcal{B}_\varepsilon}^K$ , is the hitting time of the set  $\mathcal{B}_\varepsilon$  by the population process  $\mathbf{N}^K$ .

*Then there exist three positive constants  $\varepsilon_0$ ,  $C_0$  and  $m$ , and a positive constant  $V$  depending on  $(m, \varepsilon_0)$  such that if  $p < p_0$  and  $\varepsilon \leq \varepsilon_0$ ,*

$$\lim_{K \rightarrow \infty} \mathbb{P} \left( \left| \frac{T_{\mathcal{B}_\varepsilon}^K}{\log K} - \frac{1}{b(\beta - 1)} \right| \leq C_0 \varepsilon, \mathbf{N}^K \left( T_{\mathcal{B}_\varepsilon}^K + t \right) \in \mathcal{B}_{m\varepsilon} \forall 0 \leq t \leq e^{VK} \right) = 1. \quad (1.8)$$

$\mathbf{N}^K$ .

*Symmetrical results hold for the equilibria  $(0, \zeta, \zeta, 0)$ ,  $(\zeta, 0, \zeta, 0)$  and  $(0, \zeta, 0, \zeta)$ .*

This theorem gives the order of magnitude of the time to reproductive isolation between the two patches, as a function of the scaling factor  $K$ . This isolation time is infinite when considering the dynamical system (1.5) for which  $K$  is equal to infinity. Note that the time needed to reach the reproductive isolation is inversely proportional to  $\beta - 1$  suggesting, as studied previously, that the system behaves differently when  $\beta = 1$ . Moreover, the time does not depend on the parameter  $p$ , scaling migrations. Intuitively, this can be understood as follows: the time needed to reach a neighborhood of the state  $(\zeta, 0, 0, \zeta)$  is of order 1, and from this neighborhood the time needed for the complete extinction of the  $a$ -individuals in the patch 1 and the  $A$ -individuals in the patch 2 is much longer, it is of order  $\log K$ . During this second phase, the migrations between the two patches are already balanced, which entails the independence with respect to  $p$ . Furthermore, the constant does not depend on  $d$  and  $c$  since there is no ecological difference between the two types and the two patches: during the second phase, the natural birth rate of the  $a$ -individuals in the patch 1 is approximately  $b$  since the patch 1 is almost entirely filled with  $A$ -individuals, and their natural death rate can be approximated by  $d + c\zeta = b\beta$  where the term  $c\zeta$  comes from the competition exerted by the  $A$ -individuals. Thus, their natural growth rate is approximately  $b - b\beta$  which only depends on the birth parameters.

Note that Theorem 1.2.3 gives not only an estimation of the time to reach a neighborhood of the limit, but also it proves that the dynamics of the population process stays a long time in the neighborhood of equilibria  $(\zeta, 0, 0, \zeta)$  after this time.

## Generalizations

The methods used above are actually robust to different generalizations of the model.

**Ecological differences between patches** We assumed that the patches were ecologically equivalent in the sense that the birth, death and competition rates  $b$ ,  $d$  and  $c$ , respectively, did not depend on the label of the patch  $i \in \mathcal{I}$ . In fact we could make these parameters depend on the patch, and denote them  $b_i$ ,  $d_i$  and  $c_i$ ,  $i \in \mathcal{I}$ . In the same way, the sexual preference  $\beta_i$  and the migration rate  $p_i$  could depend on the label of the patch  $i \in \mathcal{I}$ . As a consequence, the dynamical system (1.5) becomes

$$\begin{cases} \frac{d}{dt} z_{A,1}(t) = z_{A,1} \left[ b_1 \frac{\beta_1 z_{A,1} + z_{a,1}}{z_{A,1} + z_{a,1}} - d_1 - c_1(z_{A,1} + z_{a,1}) - p_1 \frac{z_{a,1}}{z_{A,1} + z_{a,1}} \right] + p_2 \frac{z_{A,2} z_{a,2}}{z_{A,2} + z_{a,2}} \\ \frac{d}{dt} z_{a,1}(t) = z_{a,1} \left[ b_1 \frac{\beta_1 z_{a,1} + z_{A,1}}{z_{A,1} + z_{a,1}} - d_1 - c_1(z_{A,1} + z_{a,1}) - p_1 \frac{z_{A,1}}{z_{A,1} + z_{a,1}} \right] + p_2 \frac{z_{A,2} z_{a,2}}{z_{A,2} + z_{a,2}} \\ \frac{d}{dt} z_{A,2}(t) = z_{A,2} \left[ b_2 \frac{\beta_2 z_{A,2} + z_{a,2}}{z_{A,2} + z_{a,2}} - d_2 - c_2(z_{A,2} + z_{a,2}) - p_2 \frac{z_{a,2}}{z_{A,2} + z_{a,2}} \right] + p_1 \frac{z_{A,1} z_{a,1}}{z_{A,1} + z_{a,1}} \\ \frac{d}{dt} z_{a,2}(t) = z_{a,2} \left[ b_2 \frac{\beta_2 z_{a,2} + z_{A,2}}{z_{A,2} + z_{a,2}} - d_2 - c_2(z_{A,2} + z_{a,2}) - p_2 \frac{z_{A,2}}{z_{A,2} + z_{a,2}} \right] + p_1 \frac{z_{A,1} z_{a,1}}{z_{A,1} + z_{a,1}}. \end{cases} \quad (1.9)$$

We can prove similar results to those of Theorems 1.2.2 and 1.2.3 under the assumption that  $(p_1, p_2)$  satisfies

$$p_i c_i (2c_i \zeta_i + p_i) < c_i (b_i (\beta_i - 1) + p_i) (b_i (\beta_i + 1) - 2d_i - p_i), \text{ for } i \in \mathcal{I}.$$

**Migration** The migration rates under consideration increase when the genetic diversity increases. Indeed, let us consider

$$H_T^{(i)} := 1 - \left[ \left( \frac{n_{A,i}}{n_{A,i} + n_{a,i}} \right)^2 + \left( \frac{n_{a,i}}{n_{A,i} + n_{a,i}} \right)^2 \right],$$

as a measure of the genetic diversity in the patch  $i \in \mathcal{I}$ . Note that  $H_T^{(i)} \in [0, 1/2]$  is known as the "total gene diversity" in the patch  $i$  (see [Nei75] for instance) and is widely used as a measure of diversity. When we express the migration rates in terms of this measure, we get

$$\rho_{\alpha, \bar{i} \rightarrow i}(n) = p \frac{n_{A,i} n_{a,i}}{n_{A,i} + n_{a,i}} = \frac{p}{2} (n_{A,i} + n_{a,i}) H_T^{(i)}.$$

Hence we can consider that the migration helps the speciation. In our work, we prove that the convergence result hold if consider a symmetrical and bounded migration rate. This amounts to consider a migration written as

$$\rho_{\alpha, \bar{i} \rightarrow i}(n) = p(n_{A,\bar{i}}, n_{a,\bar{i}}),$$

and to assume

$$p(n_{A,\bar{i}}, n_{a,\bar{i}}) = p(n_{a,\bar{i}}, n_{A,\bar{i}}) \quad \text{and} \quad p(n_{A,\bar{i}}, n_{a,\bar{i}}) \frac{n_{A,\bar{i}} + n_{a,\bar{i}}}{n_{A,\bar{i}} n_{a,\bar{i}}} < p_0,$$

where  $p_0$  has been defined in Theorem 1.2.2. Note that the second condition on the function  $p$  imposes that as one of the population sizes goes to 0, then so does the migration rate. In particular, this condition ensures that the points given by  $(\zeta, 0, 0, \zeta)$  and  $(\zeta, 0, \zeta, 0)$  are still equilibria of the system. Theorems 1.2.2 and 1.2.3 still hold with this new definition for the migration rate.

**Number of patches** Finally, we restricted our attention to the case of two patches, but we can consider an arbitrary number  $N \in \mathbb{N}$  of patches. We assume that all the patches are ecologically equivalent but that the migrant individuals have a probability of migrating to another patch that depends on the geometry of the system. Moreover, we allow the individuals to migrate outside the  $N$  patches. In other words, for  $\alpha \in \mathcal{A}$ ,  $i \leq N$ ,  $j \leq N + 1$  and  $\mathbf{n} \in (\mathbb{N}^{\mathcal{A}})^N$ ,

$$\rho_{\alpha, i \rightarrow j}(\mathbf{n}) = p_{ij} \frac{n_{A,i} n_{a,i}}{n_{A,i} + n_{a,i}},$$

where the "patch"  $N + 1$  denotes the outside of the system.

We give an explicit condition on migration rates to obtain a similar behaviour of the population.

### 1.3 Emergence of assortative mating

We now address the question of emergence of assortative mating: if a mutant starts to mate preferentially with individuals of its own type, while the other individuals still choose their mate uniformly at random, can this mutant invade the population? In this section, I present

the work of [CCL<sup>+</sup>21] in collaboration with Camille Coron, Fabien Laroche, H el ene Leman and Charline Smadi which shows that a key feature to answering this question is how the assortative mating mutation affects the total reproduction rate of individuals.

Very similarly as in the previous work, we consider a population of individuals that reproduce sexually and compete with each other for a common resource. Individuals are haploid and are characterized by their genotype at two loci located on different chromosomes describing the genetic architecture for preference. Locus 1 presents two alleles, denoted by  $A$  and  $a$ , and codes for phenotypes. Locus 2 presents two alleles denoted by  $P$  and  $p$ , and codes for assortative mating relatively to the first locus (similar models were introduced in population genetics, see for example [Gre92, SB14]).

The genotype of each individual belongs to the set  $\mathcal{G} := \{AP, Ap, aP, ap\}$  and the state of the population is characterized at each time  $t$  by a vector in  $\mathbb{N}^4$  giving the respective numbers of individuals carrying each of these four genotypes. The dynamics of this population is modeled by a multi-type birth-and-death process

$$(N^K(t), t \geq 0) := (N_{AP}^K(t), N_{Ap}^K(t), N_{aP}^K(t), N_{ap}^K(t), t \geq 0),$$

with values in  $\mathbb{N}^4$ , integrating competition, Mendelian reproduction and assortative mating. As in the previous section,  $K$  is a scaling parameter of the total population size, quantifying the environment's carrying capacity. We will be interested in the behaviour of the system for large but finite  $K$ .

Let us now precise the dynamics. When the population is in a state  $\mathbf{n} = (n_{AP}, n_{Ap}, n_{aP}, n_{ap})$ , the death rate of an individual with genotype  $i \in \mathcal{G}$  is equal to

$$d_i(\mathbf{n}) = n_i \left( d + \frac{c}{K} n \right). \quad (1.10)$$

where  $n = n_{AP} + n_{Ap} + n_{aP} + n_{ap}$  is the total population size, and the parameters  $d \in \mathbb{R}^+$  and  $c > 0$  model the natural death rate and the logistic competition.

The reproduction rates are more complex to compute. We assume that all individuals try to reproduce at the same rate. To this aim, they choose a mate, uniformly at random among the other individuals of the population. Next, individuals carrying allele  $p$  reproduce indifferently with their chosen partner, while individuals carrying allele  $P$  reproduce with a higher probability with individuals carrying the same allele at locus 1. Note that reproduction is not completely symmetric: only the genotype of the individual initiating the reproduction determines the presence or not of assortative mating. When the population is in state  $\mathbf{n}$ , the rate  $b_i(n)$  at which an individual with genotype  $i \in \mathcal{G}$  is born, is equal to

$$\begin{aligned} b_{AP}(\mathbf{n}) &= b \left[ n_{AP} + \frac{1}{n} \left( \beta_1 n_{AP} \left( n_{AP} + \frac{n_{Ap}}{2} \right) \right. \right. \\ &\quad \left. \left. - \beta_2 \left( n_{AP} \left( n_{aP} + \frac{n_{ap}}{4} \right) + n_{Ap} \frac{n_{aP}}{4} \right) \right) + \frac{\Delta_{aP}}{2n} \right] \\ b_{Ap}(\mathbf{n}) &= b \left[ n_{Ap} + \frac{1}{n} \left( \beta_1 n_{Ap} \frac{n_{AP}}{2} - \beta_2 \left( n_{Ap} \frac{n_{aP}}{4} + n_{AP} \frac{n_{ap}}{4} \right) \right) - \frac{\Delta_{aP}}{2n} \right] \\ b_{aP}(\mathbf{n}) &= b \left[ n_{aP} + \frac{1}{n} \left( \beta_1 n_{aP} \left( n_{aP} + \frac{n_{ap}}{2} \right) \right. \right. \end{aligned} \quad (1.11)$$

$$b_{ap}(\mathbf{n}) = b \left[ n_{ap} + \frac{1}{n} \left( \beta_1 n_{ap} \frac{n_{aP}}{2} - \beta_2 \left( n_{ap} \frac{n_{AP}}{4} + n_{aP} \frac{n_{Ap}}{4} \right) \right) + \frac{\Delta_{aP}}{2n} \right],$$

where

$$\Delta_{aP} := n_{aP} n_{Ap} - n_{AP} n_{ap}.$$

The parameter  $b(1 + \beta_1)$  with  $b > 0$  and  $\beta_1 \geq 0$  is the rate at which any individual (called first parent) reproduces, the second parent being chosen uniformly in the population. Each reproduction leads to the birth of a new individual with probability  $1/(1 + \beta_1)$  when the first parent carries allele  $p$ , with probability 1 if the first parent carries allele  $P$  and both parents carry the same allele at locus 1, and with probability  $(1 - \beta_2)/(1 + \beta_1)$  if the first parent carries allele  $P$  and the two parents carry different alleles at locus 1. The parameters  $\beta_1$  and  $\beta_2$  respectively quantify benefits and penalties for homogamous individuals.

We will make the following assumptions on the parameters:

$$b > d, \quad \beta_1 \geq 0, \quad 0 \leq \beta_2 \leq 1. \quad (1.12)$$

The first assumption ensures that a population of individuals mating uniformly at random is not doomed to a rapid extinction because of a natural death rate larger than the birth rate under uniform random mating. The second (resp. third) assumption means that choosy individuals have a higher (resp. smaller) probability to give birth when mating with an individual with the same (resp. different) trait ( $A$  or  $a$ ).

The goal of our main theorem is to study a step in Darwinian evolution that consists in the progressive invasion of the new allele  $P$  and loss of initial allele  $p$  in the population. The proof of this theorem relies on the study of three phases in the population dynamics trajectories: mutant survival or extinction, mean-field phase, and resident allele extinction (see Figure 1.1).

We assume that before the first mutation, all individuals mate uniformly at random (no sexual preference, all individuals carry allele  $p$ ). In a large population approximation, the dynamics of  $Ap$  and  $ap$  individuals is in that case well approximated by the dynamical system:

$$\begin{cases} \dot{z}_{Ap} = z_{Ap}(b - d - c(z_{Ap} + z_{ap})) \\ \dot{z}_{ap} = z_{ap}(b - d - c(z_{Ap} + z_{ap})). \end{cases}$$

This system admits an infinity of equilibria:

- $(z_{Ap}, z_{ap}) = (0, 0)$ , which is unstable
- $(z_{Ap}, z_{ap}) = (\rho(b - d)/c, (1 - \rho)(b - d)/c)$  for all  $\rho \in [0, 1]$ , which are not hyperbolic.

We assume that at the time  $t = 0$  of the first mutation, the resident population is *at equilibrium*. This can be translated by

$$\left( \frac{N_{Ap}^K(0)}{K}, \frac{N_{ap}^K(0)}{K} \right) \xrightarrow{K \rightarrow \infty} \left( \rho_A \frac{b - d}{c}, (1 - \rho_A) \frac{b - d}{c} \right),$$

in probability with  $\rho_A > 1/2$ . The mutant carries an allele  $\alpha \in \mathcal{A}$  and

$$\left(N_{\alpha P}^K(0), N_{\alpha P}^K(0)\right) = (1, 0).$$

Our first goal is to determine the conditions under which the mutant population has a positive probability of surviving and invading this resident population. The classical approach developed by [Cha06] relies on a coupling of the mutant population with a branching process. This coupling requires to control the dynamics of the resident population during the invasion phase, that is as long as the size of the mutant population is negligible with respect to the carrying capacity  $K$ . Note that here the study of the resident population process is more involved than in the previous references on similar questions (see for instance [Cha06, CM11, BS17]) because the initial state of the population is not a globally asymptotically stable equilibrium, since alleles  $A$  and  $a$  are initially neutral, therefore, results based on large deviation for globally stable equilibria of dynamical systems [DE97] are not available. As a consequence, the fluctuations around the initial state may be substantial and may be strongly influenced by the presence of mutants, even in a small number.

In order to state rigorously our results, let us introduce different stopping times. The first one gives the first time when the genetic proportions in the  $p$ -population deviate considerably from their starting values: for any  $\varepsilon > 0$ ,

$$U_\varepsilon := \inf \left\{ t \geq 0, \left| \frac{N_{Ap}^K(t)}{N_p^K(t)} - \frac{N_{Ap}^K(0)}{N_p^K(0)} \right| > \varepsilon \right\}. \quad (1.13)$$

The second one concerns the total  $p$ -population size: for any  $\varepsilon > 0$ ,

$$R_\varepsilon := \inf \left\{ t \geq 0, \left| \frac{N_p^K(t)}{K} - \frac{b-d}{c} \right| > \varepsilon \right\}. \quad (1.14)$$

The last one handle the size of the mutant population: for any  $\varepsilon \geq 0$ ,

$$T_\varepsilon^P := \inf \left\{ t > 0, N_p^K(t) = \lfloor \varepsilon K \rfloor \right\}, \quad (1.15)$$

where  $\lfloor x \rfloor$  is the integer part of  $x$ . Note that these stopping times depend on the scaling parameter  $K$ , however, to avoid cumbersome notations, we drop the  $K$  dependency.

Using tools from stochastic calculus and Lyapunov type arguments, we obtain a strong control of the size and proportions in the resident population. This requires an assumption on the parameters, written as  $\lambda \neq 0$ , which translates the fact that the mutant population is not critical (further details will be given below).

**Proposition 1.3.1.** *Assume  $\lambda \neq 0$ . There exist two constants  $B_0$  and  $\varepsilon_0$ , independent from  $K$ , such that for any  $\varepsilon \leq \varepsilon_0$ ,*

$$\liminf_{K \rightarrow \infty} \mathbb{P} \left( T_\varepsilon^P \wedge T_0^P < R_{B_0 \varepsilon} \wedge U_{\varepsilon^{1/6}} \right) \geq 1 - C(\mathcal{A}_0) \varepsilon^{1/12},$$

where  $C(B_0)$  is a positive constant, independent from  $K$ .



Meanwhile, we compare the dynamics of the mutant population with a bi-type branching process  $\bar{\mathbf{N}} = (\bar{N}_A, \bar{N}_a)$  admitting the following transition rates:

$$\begin{aligned}
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A + 1, \bar{N}_a) & \text{at rate } & \bar{\beta}_{AA}\bar{N}_A + \bar{\beta}_{aA}\bar{N}_a \\
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A, \bar{N}_a + 1) & \text{at rate } & \bar{\beta}_{Aa}\bar{N}_A + \bar{\beta}_{aa}\bar{N}_a \\
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A - 1, \bar{N}_a) & \text{at rate } & b\bar{N}_A \\
(\bar{N}_A, \bar{N}_a) &\rightarrow (\bar{N}_A, \bar{N}_a - 1) & \text{at rate } & b\bar{N}_a,
\end{aligned} \tag{1.16}$$

where for  $\alpha \in \mathcal{A}$ ,  $\bar{\alpha} \in \mathcal{A} \setminus \{\alpha\}$ ,

$$\bar{\beta}_{\alpha\alpha} := \frac{b}{2} \left( 1 + (\beta_1 + 1)\rho_\alpha - \frac{\beta_2}{2}\rho_{\bar{\alpha}} \right), \quad \bar{\beta}_{\alpha\bar{\alpha}} := \frac{b}{2} \left( 1 - \frac{\beta_2}{2} \right) \rho_{\bar{\alpha}}, \tag{1.17}$$

and

$$\rho_A := \lim_{K \rightarrow \infty} \frac{N_{Ap}^K(0)}{N_p^K(0)} = 1 - \rho_a. \tag{1.18}$$

are the initial proportions in the resident population. The rates of this branching process have been obtained by considering the dynamics of  $(N_{Ap}^K, N_{aP}^K)$  described by (1.10) and (1.11) when  $(N_{Ap}^K, N_{aP}^K) = (K\rho_A \frac{b-d}{c}, K(1-\rho_A) \frac{b-d}{c})$ ,  $N^K = K \frac{b-d}{c}$  and the second order terms in  $N_{Ap}^K$  and  $N_{aP}^K$  are neglected. We denote the extinction probabilities of the process  $\bar{\mathbf{N}}$  by

$$q_\alpha := \mathbb{P}(\exists t < \infty, \bar{\mathbf{N}}(t) = 0 | \bar{\mathbf{N}}(0) = \mathbf{e}_\alpha), \tag{1.19}$$

$\alpha \in \mathcal{A}$ ,  $\mathbf{e}_A = (1, 0)$  and  $\mathbf{e}_a = (0, 1)$ , meaning that the process starts with only one individual of type  $A$  or  $a$ . Classical results of branching process theory (see [AN72]) ensure that these extinction probabilities correspond to the smallest solution to the system of equations

$$\begin{aligned}
u_A(s_A, s_a) &:= b(1 - s_A) + \bar{\beta}_{AA}(s_A^2 - s_A) + \bar{\beta}_{Aa}(s_A s_a - s_A) = 0 \\
u_a(s_A, s_a) &:= b(1 - s_a) + \bar{\beta}_{aa}(s_a^2 - s_a) + \bar{\beta}_{aA}(s_A s_a - s_a) = 0.
\end{aligned} \tag{1.20}$$

Moreover, the branching process  $\bar{\mathbf{N}}$  is supercritical (i.e.  $q_A$  and  $q_a$  are not equal to one) if and only if its mean matrix

$$J := \begin{pmatrix} \bar{\beta}_{AA} - b & \bar{\beta}_{Aa} \\ \bar{\beta}_{aA} & \bar{\beta}_{aa} - b \end{pmatrix}, \tag{1.21}$$

has a positive eigenvalue that can be written as

$$\beta_1 > \beta_2 \quad \text{or} \quad \rho_A(1 - \rho_A) < \frac{\beta_1(\beta_2 + 2)}{2(\beta_1 + \beta_2)(\beta_1 + 2)}. \tag{1.22}$$

We denote by  $\lambda$  the maximal eigenvalue of (1.21), which is thus positive when (1.22) holds and will be of interest to quantify the time before invasion. Notice that  $J$  can be written as  $b$  times a matrix only depending on  $(\rho_A, \beta_1, \beta_2)$ . As a consequence,  $\lambda$  can be written  $\lambda = b\tilde{\lambda}(\rho_A, \beta_1, \beta_2)$ .

If the mutant population invades and its size reaches order  $K$ , the population dynamics enters a second phase during which it is well approximated by a deterministic large population limit. More precisely, if we define the rescaled process

$$(\mathbf{Z}^K(t), t \geq 0) := \left( \frac{N_{AP}^K(t)}{K}, \frac{N_{Ap}^K(t)}{K}, \frac{N_{aP}^K(t)}{K}, \frac{N_{ap}^K(t)}{K}, t \geq 0 \right),$$

then it will be close to the solution of the dynamical system

$$\dot{z}_i = b_i(\mathbf{z}) - (d + cz)z_i, \quad i \in \mathcal{G}, \quad (1.23)$$

where  $z = z_{AP} + z_{Ap} + z_{aP} + z_{ap}$  is the total size of the population and the functions  $(b_i, i \in \mathcal{G})$  have been defined in Equation (1.11). This dynamical system has a unique solution starting from any point of  $\mathbb{R}_+^4$ , as the vector field is locally Lipschitz, and the solutions do not explode in finite time [Chi06]. We denote by

$$(\mathbf{z}^{(z^0)}(t), t \geq 0) = (z_{AP}(t), z_{Ap}(t), z_{aP}(t), z_{ap}(t), t \geq 0),$$

this unique solution starting from  $\mathbf{z}(0) = \mathbf{z}^0 \in \mathbb{R}_+^4$ .

Let us stress that again unlike in [Cha06, CM11, BS17] but similarly as in the previous section, the dynamical system arising as the limit of the rescaled population after the invasion phase admits many (stable and unstable) fixed points and we need to identify precisely the four dimensional zone reached by the rescaled population process after the invasion phase in order to determine the convergence point of the dynamical system. We prove that starting from on initial condition with a majority of  $A$  (or  $a$ ) in the  $P$  and in the  $p$  populations (which will actually be our case) the dynamical system converges to the equilibrium  $((1 + \beta_1)b_d/c, 0, 0, 0)$  (or  $(0, (1 + \beta_1)b_d/c, 0, 0)$ ). This corresponds to invasion and fixation of the mutant and extinction of the other phenotypes.

To conclude, it remains to control the phase where the  $aP$ ,  $Ap$  and  $ap$  populations are negligible with respect to  $K$  and the mean field approximation stops being a good approximation. We will again compare the dynamics of the small population sizes with these of branching processes (now subcritical). The birth and death rates of these branching processes will provide the time to extinction of these small populations.

Combining all these steps, we are able to describe the invasion/extinction dynamics of the mutant population, that is the subject of the main result of this paper. Let us introduce some last notations: a set of interest for the rescaled process  $\mathbf{Z}^K$ , for any  $\mu > 0$

$$S_\mu := \left[ \frac{b(1 + \beta_1) - d}{c} - \mu, \frac{b(1 + \beta_1) - d}{c} + \mu \right] \times \{0\} \times \{0\} \times \{0\}, \quad (1.24)$$

a stopping time describing the time at which  $\mathbf{Z}^K$  reaches this set,

$$T_{S_\mu} := \inf\{t \geq 0, \mathbf{Z}^K(t) \in S_\mu\}. \quad (1.25)$$

Recall that we defined in (1.15) the first time when the rescaled  $P$ -mutant population size reaches any threshold (from below or above): for any  $\varepsilon \geq 0$ ,

$$T_\varepsilon^P := \inf\left\{t > 0, N_P^K(t) = \lfloor \varepsilon K \rfloor\right\},$$

**Theorem 1.3.2.** Assume that  $\lambda \neq 0$ ,

$$\left( Z_{Ap}^K(0), Z_{ap}^K(0) \right) \xrightarrow{K \rightarrow \infty} \left( \rho_A \frac{b-d}{c}, (1-\rho_A) \frac{b-d}{c} \right),$$

in probability with  $\rho_A > 1/2$  and that for some  $\alpha \in \mathcal{A}$

$$\left( N_{\alpha P}^K(0), N_{\bar{\alpha} P}^K(0) \right) = (1, 0).$$

Then there exists a Bernoulli random variable  $B$  with parameter  $1 - q_\alpha$  such that for any  $0 < \mu < (b(1 + \beta_1) - d)/c$ :

$$\lim_{K \rightarrow \infty} \left( \frac{T_{S_\mu} \wedge T_0^P}{\log K}, \mathbf{1}_{\{T_{S_\mu} < T_0^P\}} \right) = B \times \left( \frac{1}{b\tilde{\lambda}(\rho_A, \beta_1, \beta_2)} + \frac{2}{b\beta_1}, 1 \right), \quad (1.26)$$

where the convergence holds in probability.

Moreover,

$$\mathbf{1}_{\{T_0^P < T_{S_\mu}\}} \left\| \frac{\mathbf{N}^K(T_0^P)}{K} - (0, \rho_A, 0, 1 - \rho_A) \frac{b-d}{c} \right\|_1 \xrightarrow{K \rightarrow \infty} 0 \quad \text{in probability}, \quad (1.27)$$

where  $\|\cdot\|_1$  stands for the  $L^1$ -norm.

Notice that if condition (1.22) does not hold,  $q_\alpha = 1$ , and the convergence in (1.26) is an almost sure convergence to  $(0, 0)$  meaning that the mutant population dies out in a time smaller than  $\log K$ . In this case, the allelic proportions in the resident population do not vary.

Condition (1.22) gives two possible sufficient conditions for the mutant population to invade with positive probability. The first one imposes that the trade-off between the advantage for homogamous reproduction ( $\beta_1$ ) and the loss for heterogamous reproduction ( $\beta_2$ ) has to be favourable enough. The second condition requires a low level of initial allelic diversity at locus 1 (alleles  $A$  and  $a$ ). In particular, even if the advantage for homogamy is very low, very asymmetrical initial conditions ( $\rho_A$  close to 0 or 1) will ensure the invasion of the mutation with positive probability. As expected, these conditions are the same as the conditions for the approximating branching process  $\bar{\mathbf{N}}$  defined in (1.16) to be supercritical. In fact, the random variable  $B$  will be the indicator of survival of a version of  $\bar{\mathbf{N}}$  coupled with the mutant process.

Let us emphasize that our result ensures that when the mutant population invades (whatever allele  $a$  or  $A$  the first mutant carries), then the final population is monomorphic, and all individuals carry the allele  $a$  or  $A$  that was in the majority in the resident  $p$ -population. Only the mutant invasion probability depends on the allele  $a$  or  $A$  carried by the first  $P$  individual.

We were not able to obtain an explicit formula in general for the extinction probability  $q_\alpha$  of the assortative mating mutation, solutions of (1.20). Results obtained with the help of the software Mathematica show a complex dependency with respect to the parameters. We performed numerical simulations of the extinction probabilities  $(q_A, q_a)$  using Newton

approximation scheme starting from  $(0,0)$ . We computed the values of  $q_A$  as a function of  $\rho_A$  for different values of  $\beta_1$  and  $\beta_2$ . Using the symmetry of our model, we have that  $q_a(\rho_A) = q_A(1 - \rho_A)$ . We observe on Figure 1.3 that  $q_A$  is a continuous function of  $\rho_A$  but that it is not differentiable near criticality.

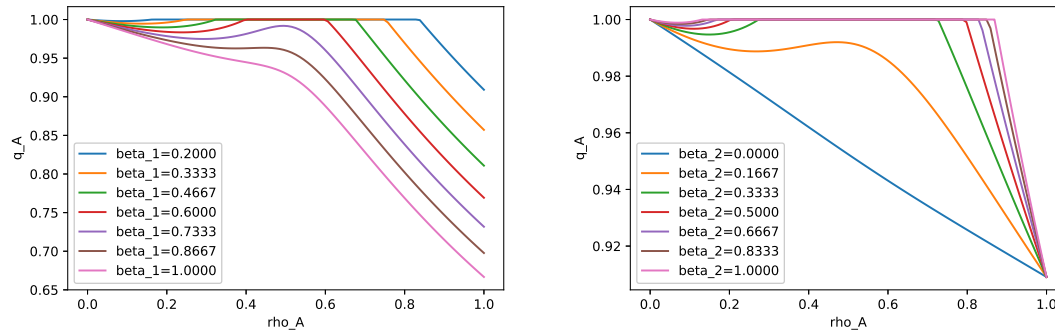


Figure 1.3: Values of  $q_A$  as a function of  $\rho_A$  for different values of  $\beta_1$  and  $\beta_2$ . On the left,  $\beta_2$  is fixed to 0.7 and  $\beta_1$  varies. On the right  $\beta_1$  is fixed to 0.2 and  $\beta_2$  varies. In both cases  $b = 1$ .

## 1.4 Dissassortative mating and structure of communities

In the two previous sections, we studied assortative mating as a driving force in selection mechanism. Here we are studying an opposite mechanism involved in maintaining diversity. Selective mechanisms favouring the emergence and persistence of polymorphism within populations are rare. Stochastic fluctuations in population densities generally limit the levels and duration of polymorphism in natural populations. Classical population genetics studies of the relative effects of genetic drift and selection regimes on the level of polymorphism. Kimura and Crow [KC64] have highlighted the fact that heterozygote advantage is a powerful balancing selection mechanism that allows high levels of polymorphism to persist within loci [LGT78]. This heterozygote advantage is often associated with mismatched pair preferences, with individuals tending to reproduce with partners with a phenotype different from their own. This peculiar mating behaviour is promoted when heterozygous offspring benefit from enhanced fitness, because disassortative pairs are then more likely to produce a fitter progeny [MBJ<sup>+</sup>21]. This mate preference generates powerful sexual selection promoting polymorphism within the populations by enhancing the reproductive success of rare phenotypes. In this section I present a work [CCL<sup>+</sup>23] in collaboration with Camille Coron, H el ene Leman, Violaine Llaurens and Charline Smadi that develops a unified theoretical framework to explore how polymorphism at targeted loci can be generated and maintained by either disassortative mating choice or balancing selection due to, for example, heterozygote advantage.

**The individual based model** We consider a population of haploid individuals characterized by a single locus  $A$ : although disassortative mating can occur in both haploid and

diploid species and may have complex genetic control, we rely on this simplified model to provide analytical predictions on the selective pressure acting on the level of polymorphism. Individuals reproduce sexually: they encounter mating partners uniformly at random and each mating event leads to the birth of a new offspring, with a probability that depends on the genotypes of both parents at locus  $A$ . We assume that crosses between individuals carrying different alleles at locus  $A$  (*disassortative matings*) have a greater reproductive success than crosses between individuals sharing the same  $A$ -allele (*assortative matings*). Because we use an haploid model, this difference in reproductive success among the different pairs of parents may account for disassortative sexual preferences, but is also akin to models where the survival probability of an offspring produced by parents with different alleles at the  $A$  locus is higher (*i.e.* heterozygote advantage benefiting the offspring).

We assume  $k$  possible alleles at locus  $A$  in the population (denoted  $1, 2, \dots, k$ ) and no mutation. We then consider Mendelian segregation of alleles during the crosses, so that the haploid offspring inherits one allele from one of its parents, chosen uniformly at random. All individuals have the same natural death rate  $d$ , and may also die from competition with other individuals, at a rate proportional to a competition parameter  $c > 0$  and to the population density. The population is characterized at each time  $t$  by the respective density of individuals carrying each allele. We use an infinite population size assumption (as in [SLL18]): we then model the dynamics of this population using a deterministic dynamical system, that can be obtained as the large-population limit of a stochastic birth-and-death process. Let us then denote by  $z_i(t)$  the density represented by allele  $i$  in the population at time  $t \geq 0$ . Then, the vector of functions  $(z_1(t), z_2(t), \dots, z_k(t))_{t \geq 0}$  is the unique solution of the following differential equations

$$\dot{z}_i(t) = z_i(t) \left( \sum_{j=1}^k \frac{\beta_i p_{ij} + \beta_j p_{ji}}{2} \frac{z_j(t)}{z(t)} - d - cz(t) \right), \quad i \in \{1, \dots, k\}, \quad t \geq 0 \quad (1.28)$$

starting from  $(z_1(0), \dots, z_k(0)) \in \mathbb{R}_+^k$ , where for each  $t > 0$ ,  $z(t) = \sum_{i=1}^k z_i(t)$  is the total population density at time  $t$ . The parameter  $\beta_i$  for  $i \in \{1, \dots, k\}$  is the rate at which an individual of type  $i$  (called first parent) mates, the second parent being chosen uniformly in the population. Each reproduction leads to the birth of a new individual with probability  $p_{ij}$ , where  $i$  is the allele carried by the first parent and  $j$  the allele carried by the second parent. The parameters  $p_{ij}$  therefore depend on both parents and model compatibility between them (with respect to either mate choice or offspring fitness), whereas the parameters  $\beta_i$  depend only on the genotype of the "first" parent and measures the rate at which a given genotype initiates reproduction. Note in particular that one can have one-sided incompatibilities, for which  $p_{ij} = 0$  but  $p_{ji} > 0$ .

We introduce the parameter

$$b := \inf_{1 \leq i, j \leq k} \frac{\beta_i p_{ij} + \beta_j p_{ji}}{2},$$

later called birth rate, and assume that  $b > 0$ , implying the impossibility of having strict genetic incompatibilities between some pair of alleles. Introducing incompatibilities may

however be possible and further studies could be performed using similar computations. For  $(i, j) \in \{1, \dots, k\}^2$ , we also introduce

$$s_{ij} := \frac{\beta_i p_{ij} + \beta_j p_{ji}}{2b} - 1,$$

For each  $i, j \in \{1, \dots, k\}$  the parameter  $s_{ij}$  may thus be interpreted as the selective advantage of a pair of parents with genotypes  $i$  and  $j$  respectively. Note that by construction  $s_{ij} = s_{ji}$  is positive or null. We highlight that the condition  $s_{ij} = 0$  does not correspond to a strict reproductive incompatibility of the pair  $(i, j)$  but to case where the pair  $(i, j)$  has the minimal birth rate in the population.

We may rewrite (1.28) as

$$\dot{z}_i(t) = z_i(t) \left( b \sum_{j=1}^k (1 + s_{ij}) \frac{z_j(t)}{z(t)} - d - cz(t) \right). \quad (1.29)$$

To maintain the population, we then always assume that

$$b \geq d > 0. \quad (1.30)$$

**Conditions for maintaining allelic polymorphism** We first give conditions on the selective advantage of each pair of genotypes under which allelic diversity is maintained. Mathematically speaking, this diversity is preserved when the system (1.29) admits a positive equilibrium, and when the population converges towards this equilibrium. Our conditions depend on the matrix  $M$  of selective advantages:

$$M := \begin{pmatrix} s_{11} & s_{12} & s_{13} & \dots & s_{1k} \\ s_{12} & s_{22} & s_{23} & \dots & s_{2k} \\ \dots & \dots & \dots & \dots & \dots \\ s_{1k} & s_{2k} & \dots & s_{k-1,k} & s_{kk} \end{pmatrix}, \quad (1.31)$$

and we say that a vector is positive ( $> 0$ ) if all its coordinates are positive.

**Proposition 1.4.1.** *Assume that  $\det(M) \neq 0$  and*

$$M^{-1} \mathbf{1} > 0, \quad \text{where } \mathbf{1} = \begin{pmatrix} 1 \\ \dots \\ 1 \end{pmatrix}. \quad (1.32)$$

*The System (1.29) admits a unique positive equilibrium*

$$Z^* := \frac{1}{c} \left( b + \frac{b}{\mathbf{1}^T M^{-1} \mathbf{1}} - d \right) \frac{M^{-1} \mathbf{1}}{\mathbf{1}^T M^{-1} \mathbf{1}}, \quad (1.33)$$

*where  $\mathbf{1}^T$  is the transpose of vector  $\mathbf{1}$ .*

*Furthermore, starting from any positive initial allelic distribution, the population will converge to this equilibrium if and only if the matrix  $M$  has exactly 1 positive eigenvalue and  $k - 1$  negative eigenvalues.*

This proposition gives a condition on the selective advantage parameters  $s_{ij}$  under which allelic diversity will be maintained. However, in the case where the matrix  $M$  has 2 or more positive eigenvalues, we cannot characterize the limiting allelic composition of the population. Note that the condition in Proposition 1.4.1 depends neither on the birth rate  $b$ , nor on the death rate  $d$ , nor on the competition term  $c$ , but only on the disassortative advantage parameters  $s_{ij}$ , ultimately modulating the reproductive success associated with the different allelic pairs (this is true because we assume that  $b > d$ ). Note that given a matrix  $M$  of selective advantages, Condition (1.32) can be easily verified numerically. Therefore, considering a specific model for the distribution of selective advantages, one might explore how many different alleles can be maintained in the long term (see Section 1.4.1 for an example).

From our general model, we indeed retrieve classical conditions for the maintenance of a large number of alleles at overdominant loci: we confirm that a large number of alleles can be maintained within a population when the advantages associated with the different disassortative pairs (akin to the different heterozygote advantages in diploid models exploring overdominance) are close to each other. Our disassortative advantage  $s_{ij}$  corresponds to the fitness of the genotype  $A_i A_j$  denoted by  $W_{ij}$  in [LGT78].

**Investigating the origin of polymorphism using successive introductions of mutants** To investigate the impact of the order of appearance of the mutations on the level of polymorphism, we assume that new alleles can arise in a population where one or several alleles already coexist. We thus refer to the new arising allele as a mutant and to the pre-existing alleles as resident alleles. We consider that mutations are rare enough so that the resident population reaches its equilibrium between two mutational events. We therefore aim at studying the fate of successive and non-simultaneous mutations in the population. We search for conditions on the mating success of the mutant allele when paired with the different resident alleles, such that the mutant allele can successfully invade and persist in the population.

We thus consider a population with  $k$  alleles whose disassortative advantage matrix is denoted by  $M$  as previously. We assume that  $M$  satisfies the conditions of Proposition 1.4.1, that ensure that the  $k$  alleles remain in the population for all times, as long as no mutation appears.

A mutant is characterized by new disassortative advantages  $S^T = (s_{k+1,1}, s_{k+1,2}, \dots, s_{k+1,k})$  and  $\sigma = s_{k+1,k+1}$ . We obtain that the condition for a mutant to invade is

$$S^T M^{-1} \mathbf{1} > 1. \quad (1.34)$$

Once the mutant invades the population, it can modify the evolutionary fate of the resident alleles. We thus investigate the effect of the mutant invasion on the number of alternative alleles maintained. We are able to give a necessary and sufficient condition under which a mutant invasion leads to a population with  $k + 1$  alleles maintained, *i.e.* to an increased allelic diversity. Let us consider the new selective matrix

$$\bar{M} = \begin{pmatrix} M & S \\ S^T & \sigma \end{pmatrix},$$

where  $S$  is the transpose of the line vector  $S^T$ , characterizing the population with  $k+1$  alleles. Proposition 1.4.1 tells us that there exists a  $(k+1)$ -alleles equilibrium if  $\bar{M}^{-1}\mathbf{1} > 0$ . If the mutant satisfies the invasion condition (1.34), the associated  $(k+1)$ -type equilibrium is also globally asymptotically stable whenever it exists (see [CCL<sup>+</sup>23] for details).

In the general case, we do not know the long time behaviour of the population when the equilibrium with all  $k+1$  alleles does not exist; however it can be detailed in the simple case of two resident alleles or it can be studied using numerical simulations.

### 1.4.1 Investigating the levels of allelic differentiation maintained within a population

In natural populations, disassortative advantage is likely to be shaped by the level of genetic differentiation associated with the different allelic pairs. In this second part of our study, we thus explicitly consider the effect of genetic differentiation on parameters describing the selective advantages, by assuming that the advantage associated with a disassortative pair is defined as an increasing function of the genetic distance between the two alleles. We specifically test different shapes of this function, and investigate their impact on the level of polymorphism maintained at locus  $A$ .

**Modelling the link between genetic distance among alleles and their disassortative mating success** To investigate the levels of differentiation among alleles that can be maintained within a population, we now consider an extension of the previous model: we assume that allelic dissimilarity has a positive effect on the selective advantages of the different disassortative mating pairs.

In this framework, the set of possible alleles is  $\{0, 1\}^L$ , where  $L$  is the number of sites where mutations can occur in a locus  $A$  (fig. 1.4). This hypothesis is relevant for modelling actual loci targeted by disassortative mate choice, such as the *MHC* in vertebrates [SMP<sup>+</sup>19].

We assume that the genetic distance between alleles carried by the parents modifies the reproductive success of disassortative pairs (fig. 1.4). We assume that the distance  $d(x, y)$  between two alleles  $x = (x_1, \dots, x_L) \in \{0, 1\}^L$  and  $y = (y_1, \dots, y_L) \in \{0, 1\}^L$  is defined by  $d(x, y) = \sum_{i=1}^L \mathbf{1}_{x_i \neq y_i}$ . The higher the distance between  $x$  and  $y$ , the higher the reproductive success of pairs of individuals with respective alleles  $x$  and  $y$ . We introduce an increasing non-negative function  $f$  on  $\mathbb{R}_+$ , such that  $s_{xy} = f(d(x, y))$ , where we recall that  $s_{xy} \geq 0$  is the selective advantage associated to pairs of parents with alleles  $x$  and  $y$  respectively and that it quantifies the probability for this pair of individuals to mate and produce a viable progeny. Here, we assume that the function  $f$  is a power function ( $f(x) = x^\alpha$  with  $\alpha > 0$ ). In particular the selection coefficient increases with the genetic distance, and the power  $\alpha$  modulates this relationship.

**Role of  $\alpha$**  We first examine the stability of two specific final population states, chosen as the two most extreme levels of polymorphism: (1) a final population with all possible alleles maintained, and (2) a final population with only the two most differentiated alleles maintained.

Whatever the number of possible sites  $L$  at locus  $A$ , we find that the equilibrium with all



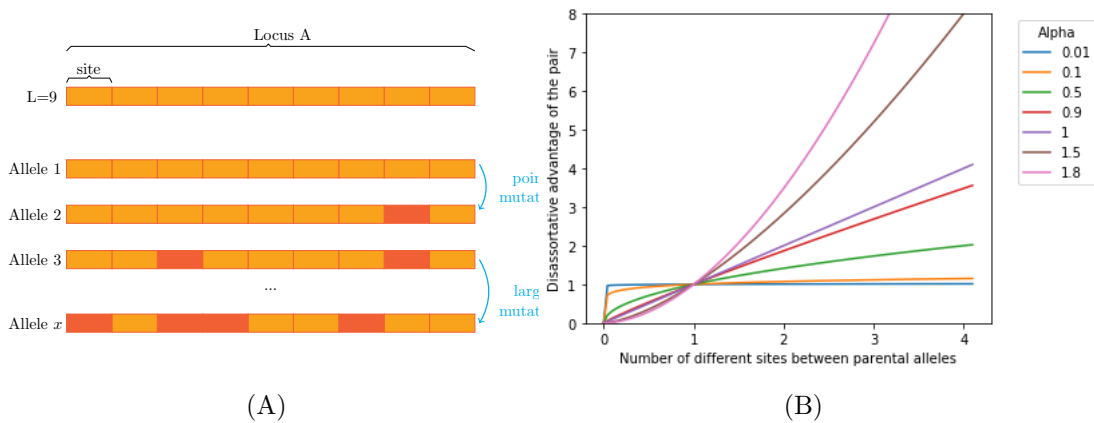


Figure 1.4: **Mutation sizes and their effects on the disassortative advantage.** Panel (a): The locus A contains  $L$  sites where mutations can occur. We model either point mutations, whereby one mutation leads to a change at a single site or other mutation kernels, where a mutation event can simultaneously affect several sites within the locus A. Panel (b): the number of sites differing between alleles in a parental pair will influence the disassortative advantage in reproduction for this pair. The parameter  $\alpha$  then determines the shape of the function, *i.e.* how much the distance between alleles enhances the reproductive success of the pair. Note that in our model, we thus distinguished the size of the mutation (*i.e.* number of differing sites) from the effect of the mutations on fitness (*i.e.* the effect of genetic distance between alleles on the reproductive success).

possible alleles maintained in the population always exists. Indeed, each allele has the same number of alleles at distances  $1, 2, \dots, L$  and thus the condition  $M^{-1}\mathbf{1} > 0$  reduces to a single vector of conditions that is always satisfied. However, using numerical simulations, we show that this equilibrium is only locally and globally stable when  $\alpha < 1$ . In fact, according to the Proposition 1.4.1, the global stability depends on the sign of the second largest eigenvalues of the matrix  $M$ , which is easy to compute numerically, even if it cannot be studied by theoretical arguments.

We then explore the conditions leading to a final population composed of two alleles at maximal genetic distance  $L$ , *i.e.*  $A_1 = (0, \dots, 0)$  and  $A_2 = (1, \dots, 1)$ . The selective advantage enjoyed in crosses between parents carrying these two alleles is  $s = f(L)$ . We then investigate whether a third allele might invade this population and modify the distribution of alleles. We introduce a third allele  $A_3 \in \{0, 1\}^L \setminus \{A_1, A_2\}$ . This allele is at distance  $x$  from  $A_1$  and  $L - x$  from  $A_2$  for some  $x \in \{1, \dots, L - 1\}$ . Using the monotonicity of  $f$ , we deduce that the invasion condition of the mutant  $A_3$  and the condition for existence and stability of a population with the three alleles can be reduced to

$$f(L) < f(x) + f(L - x).$$

As a consequence, for  $\alpha \geq 1$  the population with the two most differentiated alleles cannot be invaded by any new allele.

These highly-contrasted case-studies illustrate a phase transition in stability that occurs when  $\alpha = 1$  enlightening the role of the form of the *genotype-to-reproductive advantage*

function  $f$  in stability patterns. When  $\alpha < 1$ , all alleles can be maintained simultaneously, while for  $\alpha \geq 1$  a population with the two most differentiated alleles corresponds to an evolutionary stable equilibrium.

**Emergence of allelic diversity** We aim at exploring how allelic diversity might emerge from multiple rounds of mutant invasions. We first assume that each mutation affects only one site within the locus  $A$ , which is then shifted towards the opposite value. More precisely, when an offspring is born, it is either identical to one of its parents, or it differs from one of its parental alleles at exactly one site. We observe two different evolutionary outcomes depending on the shape of the *genotype-to-reproductive advantage* function  $f$ , determined by the value of  $\alpha$ , as detailed below.

**Case  $\alpha < 1$ .** From the results above, we already know that for  $\alpha < 1$ :

1. Any resident population with two alleles  $A_1$  and  $A_2$  can be invaded by any new mutant  $B$  at distance 1 of either  $A_1$  or  $A_2$  and will lead to a population with 3 alleles.
2. The population with all possible alleles maintained exists and is stable.

However, we have no theoretical evidence that successive introductions of mutants will actually lead to the population maintaining a large number of alleles each having another allele at distance 1. We thus numerically explore the successive invasions of mutations.

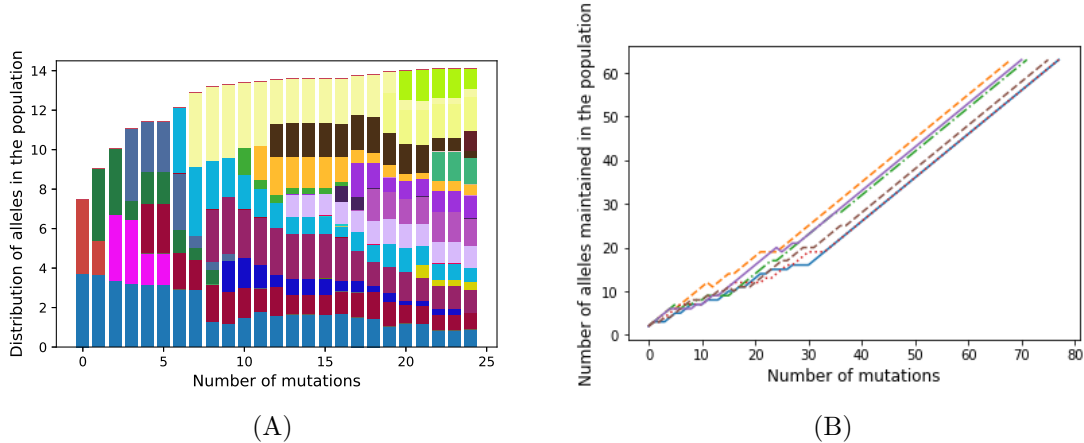


Figure 1.5: **Evolution of the number of alleles maintained in the population, assuming point mutations and convex shape of the function determining the fitness of allelic pairs ( $\alpha \leq 1$ ).** From an initial population with two alleles, we numerically induce successive mutations and track their invasion success through time. Panel (a) shows the distribution of alleles in the population through time. Each color corresponds to a given allele and the height of the bar is the number of individuals carrying each allele within the population at a each time. Panel (b) gives number of alleles maintained at equilibrium after each mutation until the total number of alleles is reached. Each line corresponds to a different numerical simulation ( $n = 6$ ). Here  $L = 6$  and  $\alpha = 0.6$  such that there are  $2^6 = 64$  possible alleles.

Figure 1.5 illustrates the contrasted evolutionary fates followed by the introduced mutants: the mutant can either go extinct rapidly if its fitness is negative, or it can invade the population. If it invades, it either co-exists with all resident alleles or triggers the loss of one or several resident alleles. The right panel (Fig. 1.5A) explicitly shows the allelic turn-over through time, suggesting that the number of alleles can be high but may also strongly vary through time in natural populations. Note that the different alleles do not have the same frequency in the population, even when the number of alleles maintained is large.

Figure 1.5B then highlights that, after a sufficiently large number of mutations, all possible alleles are maintained in the population. We furthermore observe that when the number of coexisting alleles reaches a sufficiently high level, any new mutant invades and increases the allelic diversity.

**Genotype-to-reproductive advantage function where disassortative advantage is always enhanced when differentiation between allele increases ( $\alpha \geq 1$ )**. For  $\alpha \geq 1$  we have shown that only two alleles can coexist in the population through time, meaning that any mutant invasion leads to the extinction of one out of the two resident alleles. We assume that the initial population is composed of two alleles  $A_1$  and  $A_2$  at distance  $x$ . When a mutant  $A_3$  arises, it is at distance 1 of its parental allele (say  $A_1$  by symmetry) and at distance  $x + 1$  or  $x - 1$  of the other parental allele  $A_2$ . This particular property arises from the choice of the mutation kernel (see the section below on the influence of the mutation kernel).

When assuming that the mutant is at distance  $x - 1$  from  $A_2$ , the invasion condition reads  $1 + (x - 1)^\alpha - x^\alpha > 0$ , and is thus never true for  $\alpha \geq 1$ . Therefore, a mutant allele closer to the resident allele  $A_2$  than to the resident allele  $A_1$  can never invade the population. In contrast, when assuming that the mutant is at distance  $x + 1$  from  $A_2$ , then the invasion condition reads  $1 + (x + 1)^\alpha - x^\alpha > 0$  and is always true, since  $x \mapsto x^\alpha$  is increasing for  $\alpha > 0$ . We can deduce that, when the mutant allele invades, the allele  $A_1$  is eliminated from the population. The resulting population is then composed of the two most differentiated alleles  $A_2$  and  $A_3$  at distance  $x + 1$ . In case of successive emergence of new alleles by point mutations, we thus observe increasing distances between the pairs of alleles maintained in the evolving population. This result goes further than the global stability of the population formed of the two most differentiated alleles: it proves that starting from any initial couple of alleles, the successive mutations always increase the genetic distances between the two surviving alleles. As a consequence, after a sufficient number of mutations, the population will be composed of two alleles at distance  $L$ .

Note that this general result may be modified if mutations can affect several sites at once. In [CCL<sup>+</sup>21] we develop an example showing that, with a mutation kernel allowing large mutations, coexistence of more than two alleles can be observed, even if  $\alpha \geq 1$ .

## Perspectives

In this work we have developed a unified framework for modelling the evolution of mating patterns in populations and the consequences for the community structure. Many questions remain open from our work. A first question would be to consider that mating preference

---

acts on a cue locus rather than in a self-referencing manner, and to study the conditions for invasion of such a preference. A second question would be to consider that mating preference acts on a locus targeted by selection. There is much work in the population genetics literature on this second question, but the model used considers the evolution of genotype frequencies in an infinite population (see among others [SB14][MBJ<sup>+</sup>21][OSN08]) where we consider the demographic dynamics. Incorporating stochastic population size dynamics should alter evolutionary trajectories by introducing drift, which can have a large effect on small populations. Barton and Otto [BO05] showed that genetic drift causes a delay in the fixation of beneficial mutations. The question of its effect in the context of mating preferences is an interesting line of research that I am starting to work on with Thomas Aubier (EDB, CNRS, Toulouse). This question is also closely related to the work presented in the next chapter, section 2.3.



## Chapter 2

---

# Stochastic models for Ecology II

## Recent perspectives on different models

---

The goal of this section is to present questions motivated by ecology or biology that I would like to study in the next years. It is based both on published or submitted works and on on going discussions with colleagues.

### 2.1 Matching types prey-predators models

In a wide range continuity of my PhD work , I am studying with Raphaël Forien and Pete Czuppon a matching type prey-predator system with polynomial mutation. The main motivation for this work is to study the co-evolutionary dynamics of predator-prey or host-parasite systems that are often described as Red Queen dynamics. The Red Queen dynamics, or arm race describe a persisting co-adaptation of prey and predator species to survive [MLC92, DML95]. This phenomenon is usually associated with oscillations in genotype or phenotype due to selection that favors rare types. These co-evolutionary oscillations are predicted theoretically, e.g. [SST20], but are difficult to validate empirically, though some evidence in host-parasite and host-pathogen systems exists [LFDE13, PGY<sup>+</sup>18]. To be more precise, we want to explore under what conditions, the prey-predator interaction could mediate the existence of changes in the types of most present prey and predator phenotypes. In order to have the most simple model as possible, we chose to focus on a matching type interaction [DJN14], that is to consider two types of preys and two types of predators denoted by 0 or 1, and suppose that predator 1 feeds only on prey 1 and similarly for predator 0. We also assume that the types of preys and predators are neutral with respect to the environment that means that natural birth and death rate as well as logistic competition are the same independently of the phenotype.

**Individual based model and scaling** We model the community as a multi-type birth and death process with interaction and mutations. For  $t \geq 0$  and  $i \in \{0, 1\}$ , let  $N_i^K(t)$

(resp.  $H_i^K(t)$ ) denote the number of preys (resp. predators) of type  $i$  alive at time  $t$ . The parameter  $K$  scales the total size of the population as in the previous chapter. In order to handle different scalings that might exist between the prey and the predator populations (think for example to trees insect interactions), we assume that there exists a parameter  $m > 0$  such that the prey population scales as  $K$  and the predator population as  $K^m$ .

Each prey of type  $i \in \{0, 1\}$  produces a new offspring at rate  $b > 0$  and dies at rate

$$d + \frac{c}{K}[N_0^K(t) + N_1^K(t)] + \frac{p}{K^m}H_i^K(t),$$

where  $d > 0$  denotes the natural death rate,  $c > 0$  the competition and  $p > 0$  the predation rate. Each new prey is either of the same type as its parent (with probability  $1 - v_K$ ), or of the other type (with probability  $v_K$ ), for some  $v_K \in (0, 1)$ .

Each predator of type  $j \in \{0, 1\}$  produces a new offspring at rate

$$\beta + \frac{\rho}{K}N_j^K(t),$$

that includes natural birth at rate  $\beta > 0$  and the positive effect from predation at rate  $\rho > 0$ . Each predator dies at rate  $\delta + \frac{\gamma}{K^m}H^K(t)$ , composed of a natural death rate  $\delta > 0$  and a logistic competition rate  $\gamma > 0$ . Moreover, each new predator inherits the type of its parent with probability  $1 - \vartheta_K$ , or mutates to the other type with probability  $\vartheta_K$ , for some  $\vartheta_K \in (0, 1)$ .

We denote by  $(Z^K(t), t \geq 0)$  the pure jump Markov process taking values in  $\mathbb{N}^4$  that describes the behaviour of the total population:

$$Z^K(t) = (N_0^K(t), N_1^K(t), H_0^K(t), H_1^K(t)).$$

These assumptions imply that the typical size of the prey (resp. predator) population is  $K$  (resp.  $K^m$ ). We will therefore say that a prey (resp. predator) population is at a macroscopic level when it is of order  $K$  (resp.  $K^m$ ), and is at a microscopic level otherwise.

A first limit of interest corresponds to the large population limit with  $K \rightarrow \infty$  and vanishing mutations  $\mu_K, \vartheta_K \rightarrow 0$ . The rescaled stochastic process

$$(\mathbf{Z}_{sc}^K(t), t \geq 0) = \left( \frac{N_0^K(t)}{K}, \frac{N_1^K(t)}{K}, \frac{H_0^K(t)}{K^m}, \frac{H_1^K(t)}{K^m} \right), \quad (2.1)$$

then converges in law to the solution of the dynamical system

$$\begin{cases} \frac{dn_0(t)}{dt} = n_0(t)(b - d - c(n_0(t) + n_1(t)) - ph_0(t)) \\ \frac{dn_1(t)}{dt} = n_1(t)(b - d - c(n_0(t) + n_1(t)) - ph_1(t)) \\ \frac{dh_0(t)}{dt} = h_0(t)(\beta - \delta - \gamma(h_0(t) + h_1(t)) + \rho n_0(t)) \\ \frac{dh_1(t)}{dt} = h_1(t)(\beta - \delta - \gamma(h_0(t) + h_1(t)) + \rho n_1(t)) \end{cases}. \quad (2.2)$$

One difficulty in our analysis arises from the perfect symmetry of the ODE system that consequently contains an infinite number of equilibria:

- the null equilibria

- non matching equilibria  $(\bar{n}, 0, 0, \bar{h})$  and  $(0, \bar{n}, \bar{h}, 0)$ , where  $\bar{n} = (b - d)/c, \bar{h} = (\beta - \delta)/\gamma$
- an infinite line of equilibria with two preys or two predators  $(\alpha\bar{n}, (1 - \alpha)\bar{n}, 0, 0)$ , or two predators  $(0, 0, \alpha\bar{h}, (1 - \alpha)\bar{h})$  for  $\alpha \in [0, 1]$ ,
- matching-type equilibria  $(\hat{n}, 0, \hat{h}, 0)$  and  $(0, \hat{n}, 0, \hat{h})$  with

$$\hat{n} = \frac{\bar{n} - \frac{\rho}{c}\bar{h}}{1 + \frac{\rho\rho}{c\gamma}}, \quad \hat{h} = \frac{\bar{h} + \frac{\rho}{\gamma}\bar{n}}{1 + \frac{\rho\rho}{c\gamma}}. \quad (2.3)$$

- an unique four type equilibrium  $\mathbf{z}^* = (n^*, n^*, h^*, h^*)$  where

$$n^* = \frac{2\bar{n} - \frac{\rho}{c}\bar{h}}{4 + \frac{\rho\rho}{c\gamma}}, \quad h^* = \frac{2\bar{h} + \frac{\rho}{\gamma}\bar{n}}{4 + \frac{\rho\rho}{c\gamma}}. \quad (2.4)$$

Note that depending on the parameters, these equilibria might not lie in  $(\mathbb{R}_+)^4$ .

In order to observe long time changes in the phenotypic composition of the community, we consider a polynomial mutation scale very different from the adaptive dynamics setting:

$$v_K = \frac{1}{K^v}, \quad \vartheta_K = \frac{1}{K^{m\vartheta}},$$

with  $v$  and  $\vartheta$  are in  $(0, 1)$ . This scaling corresponds to rapid mutations since when a population is macroscopic, the number of mutations stemming from this population will be of the order of  $K^{1-v}$  or  $K^{m(v-1)}$  growing to  $\infty$  as  $K \rightarrow \infty$ . Such a scaling has been introduced by Durrett and Mayberry [DM11] and further developed by [BCS19, CMT21, CKS21] notably and allows to follow the dynamics of populations with different sizes: macroscopic populations (of size of order  $K$  for preys or  $K^m$  for predators) will be compared to solutions of differential equations, while microscopic population (of size of order smaller than  $K$  for preys or  $K^m$  for predators) will be studied using comparisons with branching processes. In order to give some intuition on the dynamics during invasion, let consider a branching birth and death process  $U_t$  with growth rate  $r = b - d$  and initial condition  $K^a$ . It is well known that  $U_t e^{-rt}$  is a martingale, such that

$$\mathbb{E}(U_t) = e^{rt} \mathbb{E}(U_0) = e^{rt} K^a.$$

As a consequence, on the  $\log(K)$  time scale

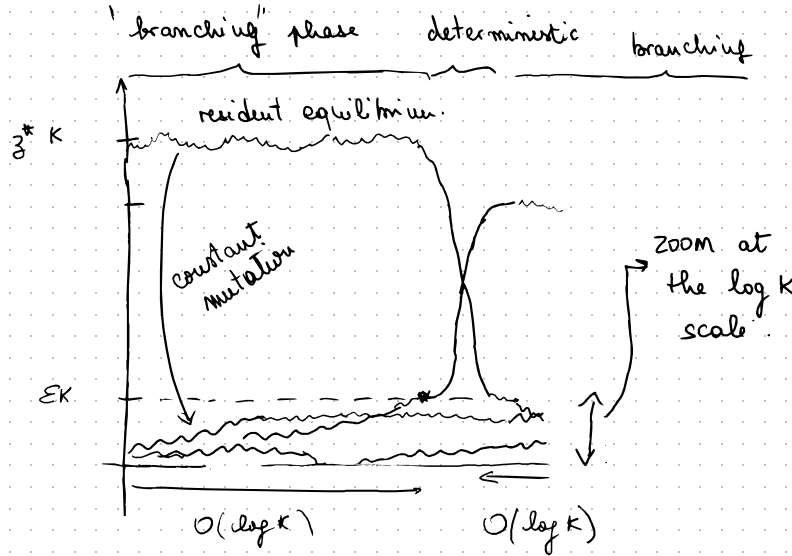
$$\mathbb{E}(U_{t \log(K)}) = e^{rt \log(K)} K^a = K^{a+rt}.$$

We see here that in expectation,  $\log(\mathbb{E}(U_{t \log(K)}))$  evolves linearly with times, and actually a similar convergence holds in probability for the stochastic process  $\log(U_{t \log(K)} + 1)/\log(K)$  [CMT21].

The main idea is thus to study the dynamics of the population through the logarithmic exponents defined as

$$X_i^K(t) := \frac{\log(1 + N_i^K(t))}{\log(K)}, \quad Y_i^K(t) := \frac{\log(1 + H_i^K(t))}{m \log(K)}, \quad i \in \{0, 1\}. \quad (2.5)$$





This means that  $X_i^K(t) \geq 0$  and  $Y_i^K(t) \geq 0$  are such that

$$N_i^K(t) = K^{X_i^K(t)} - 1, \quad H_i^K(t) = K^{mY_i^K(t)} - 1.$$

We prove, in our case that the logarithmic exponents  $((X_i^K(t \log(K)), Y_i^K(t \log(K)))_{i \in \{0,1\}}, t \in [0, T])$  converge in probability toward a deterministic piecewise linear limit  $((x_i(t), y_i(t))_{i \in \{0,1\}}, t \in [0, T])$  taking values in  $(0, 1]^4$ . Population with exponent 1 corresponds to the macroscopic population and will still be compared with solutions of ODE systems, while smaller exponents corresponds to microscopic population and will grow or decay linearly with time depending on their fitness (or growth rate) in the resident equilibrium. The limiting process can be constructed recursively on time intervals corresponding to successive invasions of microscopic populations. An invasion lasts the time for a microscopic growing population to reach a macroscopic size or in the limit, for its logarithmic exponent to reach 1.

**Rapid overview of the different cases** I will not present in detail all the patterns that can be observed nor the proofs that are (not always direct) adaptations of previous works [CM11][BCS19], but I'll focus on two cases for which we can prove that the invasion time accumulate in finite time.

**Case A** - In the first case, all equilibria presented above are positive. This corresponds to choosing  $\beta - \delta > 0$  and  $\bar{h} < c\bar{n}/p$ . Let us denote by  $\bar{z}_{0,1} = (\bar{n}, 0, 0, \bar{h})$  and  $\bar{z}_{1,0} = (0, \bar{n}, \bar{h}, 0)$  the two equilibria featuring a prey population and a non-matching predator population. Similarly, let  $\hat{z}_0 = (\hat{n}, 0, \hat{h}, 0)$  and  $\hat{z}_1 = (0, \hat{n}, 0, \hat{h})$  denote the two equilibria consisting of a prey population and the matching predator population.

We observe that in this case, the successive invasions will lead to cyclic dynamics of the

equilibrium with the successive states

$$\begin{array}{ccc}
 \hat{z}_0 & \longrightarrow & \bar{z}_{1,0} \\
 \uparrow & & \downarrow \\
 \bar{z}_{0,1} & \longleftarrow & \hat{z}_1
 \end{array}$$

Indeed, when the non matching type coexists at the macroscopic level, the matching preda-

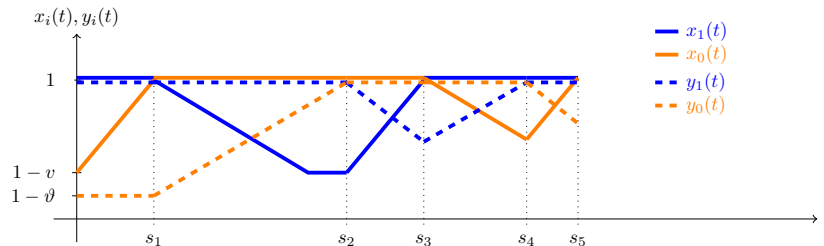


Figure 2.1: A trajectory of the limiting exponents in case A. Solid lines corresponds to the exponents of the prey populations while dashed lines are for predators. The colors are associated with types.

tor invades because its prey is abundant while the second prey decays since its predator is abundant. On the contrary, when the matching prey-predator types have macroscopic size, then the second prey can invade since it does not suffer predation and the associated predator decays due to competition with the resident predator.

The situation is illustrated in figure 2.1. The associated slopes depend on whether the resident equilibrium is composed of matching or non-matching predator-prey types. These cycles look like red queen dynamics, but we can show that the limiting process accumulates to 1 in finite time, suggesting that logistic competition prevents these cycles from continuing ad infinitum.

**Case B** - In a second case, we assume  $\beta - \delta > 0$  and  $c\bar{n}/p < \bar{h} < 2c\bar{n}/p$  which corresponds to the case where the matching type equilibrium is negative but not the other ones. Here, when two predators and a prey coexists at a macroscopic level, the associated deterministic dynamics drives prey to extinction and the two predator types coexists at some proportion. We then observe successive invasion of preys whose type corresponds to the less abundant predators (see figure 2.2).

Note here, that after each prey invasion, the number of prey remains microscopic and the proportions of predators change. We can also prove that these invasion accumulate in finite time on the deterministic model.

**Perspectives** A main unresolved question here is what happens to the stochastic system after accumulation time, that is when all the exponents are close to 1. In both cases A and B presented above the four type equilibrium  $\mathbf{z}^*$  exists and is globally asymptotically stable from positive initial condition, therefore we might expect that the four exponents remain equal to 1 and the associated population sizes remain close to  $\mathbf{z}^*$ . However the mathematical proof requires to control the evolution of the four types dynamical system (2.2) from an initial condition of the form  $(K\bar{n}, K^{1-\eta}, K^{1-\eta}, K\bar{h})$  and to prove that they reach a neighbourhood

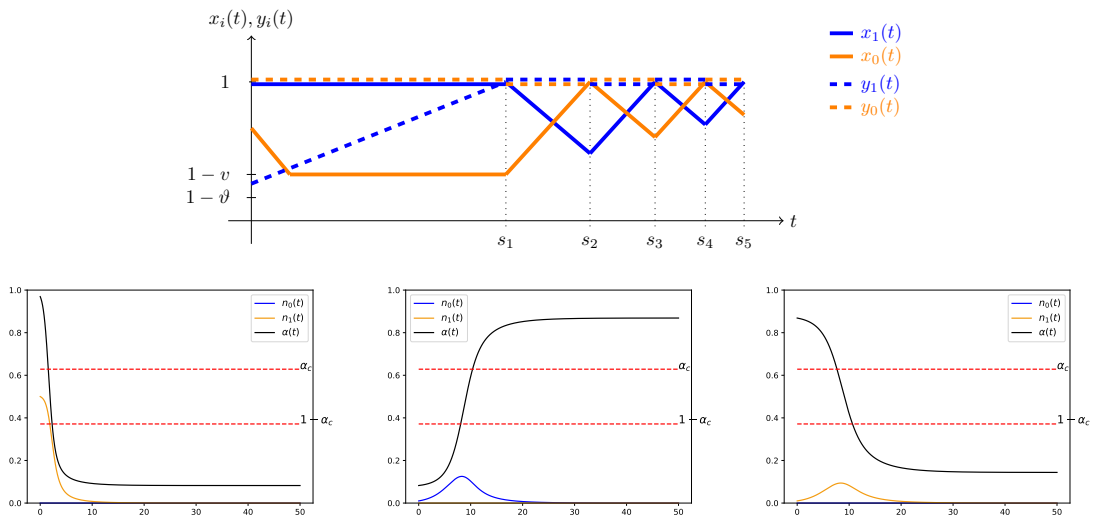
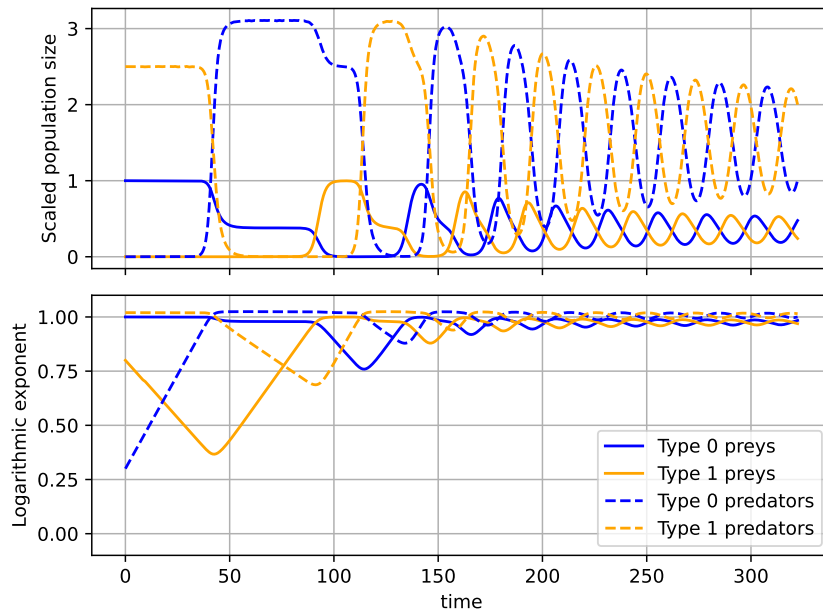


Figure 2.2: Case B. The above Figure represents a typical trajectory of the limiting process  $(x_0, x_1, y_0, y_1)$  in case B. The three graphs on the second line are the solutions of the dynamical system associated with the macroscopic populations at time  $s_1, s_2, s_3$ . A prey can only invade if the proportion of its matching predator is less than  $\alpha_c$  that is greater than  $1/2$  in this case.

of  $\mathbf{z}^*$  in a time of order  $\eta \log(K)$  for small values of  $\eta$ . This question is difficult, in particular because standard Lyapunov functions associated to Lotka-Volterra system are degenerated due to the symmetry of (2.2). We conclude this section with a numerical simulation of the deterministic system with mutations for a large  $K$ .



## 2.2 Oncological viruses

In this section I will present a joint work with Jérôme Fehrenbach (IMT) and Pierre Cordelier [CCF22] who is an oncologist at the CRCT Toulouse.

An oncolytic virus (OV) is a virus that has the potential to selectively kill tumour cells. The key phenomenon is that this virus replicates inside the tumour cell and upon lysis (cell destruction), a large number of virus copies are released to infect other cells. The action of OVs was reported as early as the mid-twentieth century [KR07], and the first OV for the treatment of head and neck cancer was approved in China in 2005 and then by the FDA and EMA in 2015 for herpes simplex virus (HSV)-based virotherapy for melanoma. More than 80 clinical trials were included in the review [SHK<sup>+</sup>17]. However, many questions remain about the optimal parameters for an OV to achieve its goal, and the design and testing of OVs is an active area of research. With Jérôme Fehrenbach, we consider an ODE system for the population of uninfected and infected cancer cells and of free viruses, close to previous models [BCJ<sup>+</sup>08, BKL<sup>+</sup>10, Wod03], and explore optimal strategies for virus injection that allow to control the proliferation of the cancer. We focus on the case where the virus dynamics is faster than the cell dynamics and model the treatment as a large injection of viruses at different times.

Let us specify the model. Let  $z(t)$  be the amount of free virus. We differentiate the cancer cells according to whether they are infected by viruses or not, and denote by  $x(t)$  the number of cancer cells not infected by viruses, and  $y(t)$  the number of cancer cells infected by viruses.

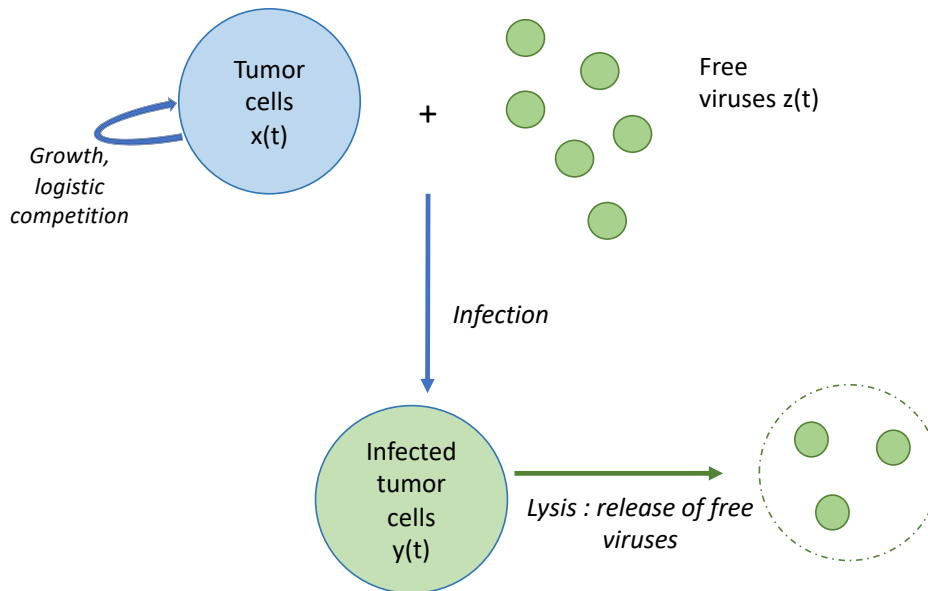


Figure 2.3: Schematic representation of the dynamics of the OV system.

We assume that uninfected cancer cells grow according to a logistic law, with a growth rate of  $\mu$  and a carrying capacity of 1. The unit for tumour cells is thus equal to the carrying capacity. These cells then become infected tumour cells at a rate proportional to the amount

of free virus, and we denote the infection rate by  $\lambda$ . Note that in our model the logistic growth is only limited by the number  $x$  of non-infected cancer cells and not by the total number  $x + y$  of (infected and non-infected) cancer cells. This choice amounts to assuming that the growth limitation is mainly due to resource consumption, since in our model the infected cells  $y$  do not grow and cannot consume resources.

We assume that infected tumour cells  $y(t)$  are lysed at a rate  $\gamma$ . Finally, free virus is used to infect cancer cells at a rate proportional to the number of cells. The unit for measuring free virus has been chosen so that the amount of virus needed to infect one cell is 1, so a proportion  $\lambda x(t)$  of virus disappears due to new infections. Free virus is released into the system when an infected cell is lysed, and the parameter  $N$  scales the amount released by each lysed infected cell. We also assume that viruses are cleared by the immune system and spread at a rate  $\delta$ . In the following we will assume that the dynamics of the viruses is fast compared to the dynamics of the tumour cells, which is modeled by the parameter  $\epsilon$ .

$$\begin{cases} \frac{d}{dt}x_\epsilon(t) = -\lambda x_\epsilon(t)z_\epsilon(t) + \mu x_\epsilon(t)(1 - x_\epsilon(t)), \\ \frac{d}{dt}y_\epsilon(t) = \lambda x_\epsilon(t)z_\epsilon(t) - \gamma y_\epsilon(t), \\ \epsilon \frac{d}{dt}z_\epsilon(t) = -\lambda x_\epsilon(t)z_\epsilon(t) + N\gamma y_\epsilon(t) - \delta z_\epsilon(t). \end{cases} \quad (2.6)$$

We expect that as  $\epsilon \rightarrow 0$ , the virus' dynamics would be infinitely faster and the system can be described by a two species system where the virus population is always at equilibrium

$$z_0(t) = \frac{N\gamma y_0(t)}{\lambda x_0(t) + \delta}. \quad (2.7)$$

By replacing in the other equations one obtains the following two species model:

$$\begin{cases} \frac{d}{dt}x_0(t) = -\frac{\lambda N\gamma x_0(t)y_0(t)}{\lambda x_0(t) + \delta} + \mu x_0(t)(1 - x_0(t)), \\ \frac{d}{dt}y_0(t) = \frac{\lambda N\gamma x_0(t)y_0(t)}{\lambda x_0(t) + \delta} - \gamma y_0(t). \end{cases} \quad (2.8)$$

**Treatment modelling** The treatment amounts to add to the system a quantity  $V$  of viruses at the time  $t_0$ . A protocol composed of successive administrations of virus is described by the administration of different quantities  $V_0, V_1, \dots, V_k$  at the instants  $t_0, t_1, \dots, t_k$ .

For the slow-fast system (2.6), the treatment is modeled by adding a source term in the equation for  $z_\epsilon$  that is a Dirac in time at  $t = t_0$ .

$$\epsilon \frac{d}{dt}z_\epsilon(t) = -\lambda x_\epsilon(t)z_\epsilon(t) + N\gamma y_\epsilon(t) - \delta z_\epsilon(t) + D_{t=t_0}V,$$

The main issue in this part is that the amount of virus at initial time explodes as  $\epsilon \rightarrow 0$  such that standard results do not apply. We have to deal with a singular perturbation (i.e. of order  $1/\epsilon$ ) of a singular system. The strategy here consists in proving that this burst induces in the limit system an instantaneous change in the amount of cancer cells, while the quantity of viruses rapidly decreases towards its equilibrium value. This phenomenon is called a boundary layer [LL54].

**Theorem 2.2.1.** *Let  $(\bar{x}_\epsilon, \bar{y}_\epsilon, \bar{z}_\epsilon)$  be the solution of the slow-fast system (2.6) starting from  $(x(t_0), y(t_0), z(t_0) + \frac{V}{\epsilon})$  and  $(x_0^Q, y_0^Q, z_0^Q)$  the solutions of (2.8)-(2.7) with initial condition  $(x(t_0) - Q, y(t_0) + Q)$ . Assume that the positive constant  $Q$  is defined as*

$$Q = V - \delta \int_0^\infty Z_0(s) ds, \quad (2.9)$$

and solves the implicit equation

$$Q + x_0(t_0) \exp(\beta(Q - V)) = x_0(t_0), \quad (2.10)$$

then, the difference

$$|(\bar{x}_\epsilon, \bar{y}_\epsilon, \bar{z}_\epsilon) - (x_0^Q, y_0^Q, z_0^Q)|$$

vanishes as  $\epsilon \rightarrow 0$  uniformly on any time interval  $[t, T]$  with  $T > t > t_0$ .

We then explored numerically different treatment strategies

- The first possible objective is to cure completely the cancer, which amounts to reach the equilibrium  $(0, 0)$ . A more realistic objective [BKL<sup>+</sup>10] would be to reach the region  $\{x \leq \eta\}$  where  $\eta$  is a small constant (say  $\eta = 10^{-8}$ ) so that  $x \leq \eta$  in our continuous modelling amounts to saying that the number of tumour cells is below 1, or the number of tumour cells is so small that an alternative therapy such as tumour resection should have a great probability of success.
- The second possible objective is to reach the endemic equilibrium (when it is stable), or the limit cycle that loops around the endemic equilibrium (when it is unstable). This amounts to control the volume of the tumour, so that it stays at a reasonable load. This strategy is adapted when the tumour load at the endemic equilibrium  $(x^*, y^*)$  is bearable, which means that  $x^*$  is small.

**Perspectives** From a mathematical point of view the model as coupled ODE is not completely convincing since we are interested in the case where both viruses or cells can be scarce, therefore a question is the construction of a stochastic process. A natural way to model the dynamics would be to consider that the cancer cells divide according to a linear branching process and can be infected by viruses that change their reproduction behaviour. Moreover, the viruses can proliferate inside a cancer cell driving it to burst. Stochastic models for growth fragmentation of infected cells have been developed by [Kim97, Ban08, Ban09] and studied in the recent years by [MS24, MS21] including complex pattern in the dependence of the division rate with the infection level or death of infected cells. A main question in extending these models for oncolytic virus is to take interactions into account as the death of infected cells increases the chance of new infections which would correspond to the appearance of "migrant" cells with low levels of infection.

## 2.3 Perspectives on impact of drift in some population genetics models

Recently, I started a project with Alphonse Emakoua (IMAG Montpellier), Sylvain Gandon (CEFE Montpellier) and Sepideh Mirrahimi (IMAG, IMT) on multi-drug resistance, which

is a challenging issue for public health. Our work is motivated by the question of how to develop new ways of using our arsenal of antibiotics to slow the spread of resistance and maintain the efficacy of treatments. One promising strategy is the cyclical use of different types of antibiotics. Mathematical models have been proposed to explore the effectiveness of this strategy [BPM10, BPMGI17, BLL04, BLL97], but most of these papers exclude the possibility of multiple resistance and neglect the effect of demographic stochasticity. These two factors, however, have critical evolutionary consequences: (1) the emergence of multiple resistance can undermine the efficiency of antibiotics, (2) in certain treatment strategies, stochasticity can lead to the extinction of resistant genotypes and extend the efficacy of drugs. We aim to develop a model that takes these factors into account in order to better characterize the consequences of different treatment strategies and to identify those that will allow sustainable control of infections.

We first built an individual-based epidemiological model of the form SIS (Susceptible-Infected-Susceptible) including Mendelian reproduction. We assume that resistance can be observed at two different loci, each with two alleles  $A/a$  and  $B/b$ . Both alleles  $A$  and  $B$  encode resistance and have a selective advantage. The numerical simulation shows that for a fixed but large population size, the invasion of the advantageous allele is slower in the stochastic model than in its deterministic large population limit. This is due to randomness interfering with advantageous alleles and creating negative linkage disequilibrium. Such a phenomenon has been described in a much simpler model by Hill Robertson and Barton and Otto [BO05], where the population size is fixed over time. To understand this lag, [MG21] recently emphasized the crucial role of linkage disequilibrium in the dynamics of multiple resistance.

A first step in this work is to specify the approximations of [BO05] that give expression to the "quasi linkage equilibrium"[Nag77, Nag93] in a two-locus diallelic model including genetic drift. The authors consider a deterministic and discrete-time genetic population model in which they include genetic drift by sampling  $N$  individuals in each generation. Their model is a discrete-time Markov process describing the evolution over time of the different allele frequencies in a fixed-size population model. They then consider an approximation under weak selection and large population size, extending the "quasilinkage equilibrium" to a stochastic setting.

Our aim is to go one step further in the extension to describe and express the delay with the deterministic case ( $N \rightarrow \infty$ ), and to detail on what time scale the quasilinkage equilibrium approximation holds. The computations are complex and require expanding the dynamics both in terms of the small selection parameter and the population size, and in terms of the deviation of the sampling from the mean behaviour. In a second step, we want to specify the role of epistasis, i.e. the additive effect of several beneficial mutations, and then adapt their method to cases where  $N$  varies in time, as might be the case during an infection phase with periodic treatment.

## Chapter 3

---

# Hawkes processes with inhibition

---

### 3.1 Introduction

Hawkes processes, originally introduced by [Haw71], model the successive arrival of events in time and the cascading influence of each event on the randomized upcoming events. They have been used in a wide variety of contexts: to model the occurrence of earthquakes and their aftershocks, to describe the spiking times of neurones [RBRGTM14, CDRB15, L17], to model sell actions in finance [BM14, Haw18] or successive calls to insurance policies [BRBH21] or more recently for epidemiology or ecology.

A Hawkes process, denoted below as  $(N_t)_{t \geq 0}$ , is defined by its conditional intensity, which describes the instantaneous rate of occurrence of an event conditionally on the past. This conditional intensity at time  $t$ ,  $\Lambda_t$ , is defined in terms of a stochastic integral of the past trajectory of the process, which can be written as

$$\Lambda_t = \Phi \left( \int_0^{t-} h(t-s) dN_s \right). \quad (3.1)$$

In the equation (3.1), the non-negative (and possibly non-linear) function  $\Phi : \mathbb{R} \rightarrow \mathbb{R}_+$  accounts for the integration of the process with respect to its past trajectory, while  $h : [0, \infty) \rightarrow \mathbb{R}$  models the temporal self-interaction between events and is often called the *reproduction function*. We can include an initial condition to the definition which accounts for the event occurring on  $(-\infty, 0)$ , but I will restrict my presentation here to the case of an empty initial condition and consider only positive times.

The original framework corresponds to the linear Hawkes process with a purely excitatory kernel  $h \geq 0$  and  $\Phi(x) = \lambda + x$ ,  $\lambda > 0$ . Here,  $\lambda$  can be seen as the baseline arrival rate of ‘ancestral’ events, which generate clusters of ‘descendant’ events through a time-continuous Galton-Watson process with reproduction function  $h$  [HO74]. Many generalizations of the linear case have been introduced recently to account for the variety of applications, e.g. possibly multivariate Hawkes processes with complex time interaction patterns or inhomogeneous connectivity within a very large population of individuals [LV14, LTMB<sup>+</sup>18, CSSBW17] or non-poissonian arrival of ancestors [WFS16, RY23].



The main question I have been interested in is understanding the effects of inhibition that arise when considering a reproduction function  $h$  that takes negative values. The first two articles [CGMT20] and [CCC22] introduce and study a renewal structure for Hawkes processes in order to obtain information about the long time behaviour of these Hawkes processes, while the third and fourth [CMM24a, CMM24b] study a discrete version of the inhibited Hawkes process.

### 3.1.1 A brief recap on linear Hawkes processes

Let me start by recalling well known properties of the linear Hawkes process (see [LLT21] for an easy introduction or [DVJ06] for a more thorough reading).

**Construction from a Poisson point measure** Let  $Q$  be a two-dimensional Poisson point process on  $(0, +\infty) \times (0, +\infty)$  with unit intensity. The Hawkes process  $N$  is the solution of the stochastic differential equation

$$\begin{cases} N_t = \int_{(0,t] \times (0,+\infty)} \delta_u \mathbf{1}_{\theta \leq \Lambda(u)} Q(du, d\theta) \\ \Lambda(u) = \lambda + \int_{(0,u)} h(u-s) N(ds), \quad u > 0, \end{cases} \quad (3.2)$$

where  $\lambda > 0$  is an immigration rate,  $h : (0, +\infty) \rightarrow \mathbb{R}$  is a non negative measurable function. This construction holds in a very general framework including the nonlinear case. The classical proof based on a Picard fixed point argument [DFH16] requires that  $h \in \mathcal{L}^1$  (see [BM96] for the non empty past case).

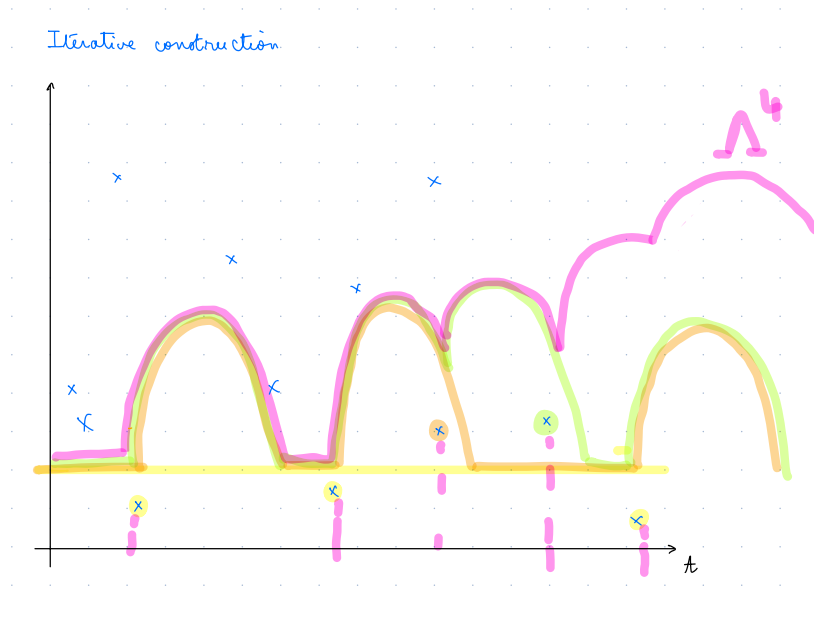


Figure 3.1: Construction of the intensity by thinning of a Poisson a point measure.

The above construction can be interpreted as a thinning of the Poisson point measure in the case where the intensity remains bounded. This construction is illustrated on Figure 3.1,

and gave rise to the thinning algorithm proposed by Ogata [Oga78].

**Cluster interpretation** A quite famous property of linear Hawkes processes is their interpretation as Poissonian clusters. The general theory of cluster point processes (see [DVJ06]) allows to construct point processes as superposition of independent copies of a given point process. The cluster point process is then fully described by a point process of "centres" corresponding to the arrival of the independent copies of "satellites" processes. In the case of the linear Hawkes process, Hawkes and Oakes [HO74] provided a construction based on centres distributed as a Poisson point process with rate  $\lambda$  and satellites being all the points generated by these ancestors that can be seen as a time inhomogeneous linear birth and death process.

A way to fully understand this construction, is to come back to the thinning construction (3.2). Let us denote by  $(T_n)_{n \geq 1}$  the sequence of times in a linear Hawkes process, and associate  $(\theta_n)_{n \geq 1}$  the sequence of "weights" corresponding to these times through the Poisson measure  $Q$  and distributed as independent uniform on  $[0, \Lambda(T_n)]$ . Each time  $T_n$  represents the birth time (or arrival time) of the  $n$ th individual. We define recursively the cumulative impact of each birth on the intensity as

$$\begin{cases} \lambda_t^{(0)} = \lambda, & t \geq 0 \\ \lambda_t^{(n)} = \lambda_t^{(n-1)} + h(t - T_n), & t \geq T_n. \end{cases}$$

We then have

$$\Lambda^h(u) = \sum_{n \geq 1} \lambda_u^{(n-1)} \mathbf{1}_{u \in [T_{n-1}, T_n)},$$

with the convention  $T_0 = 0$ .

To construct the genealogy, we say that  $n$ -th individual

- is an ancestor if  $\theta_n \leq \lambda_{T_n-}^{(0)} = \lambda$ ,
- is a descendant of the individual born at  $T_i$  (with  $1 \leq i < n$ ) if  $\theta_n \in (\lambda_{T_n-}^{(i-1)}, \lambda_{T_n-}^{(i)})$ .

This construction is illustrated in Figure 3.2. We see here well that ancestors arise from a Poisson point process with intensity  $\lambda$ , and that the number of descendants of a given individual is a Poisson random variable with parameter  $\|h\|_1$ . The birth dates of descendants of an individual born at  $T_i$  are then given by  $T_i + T$  where  $T$  is drawn according to  $h(t - T_i) / \|h\|_1$ . Considering only the genealogy, each ancestor gives then birth to a Galton Watson tree of individuals with reproduction law  $\mathcal{P}(\|h\|_1)$  independently from the other. We see from this construction that if  $\|h\|_1 < 1$ , the Galton Watson trees generated by ancestors are sub-critical and thus almost surely finite. This assumption ensures the existence of a stationary version of the Hawkes process (see [BM96], [HO74]).

A main interest of this construction is that it allows to split the contribution of the different time arrival into independent clusters with same law, whose impact may of course superpose

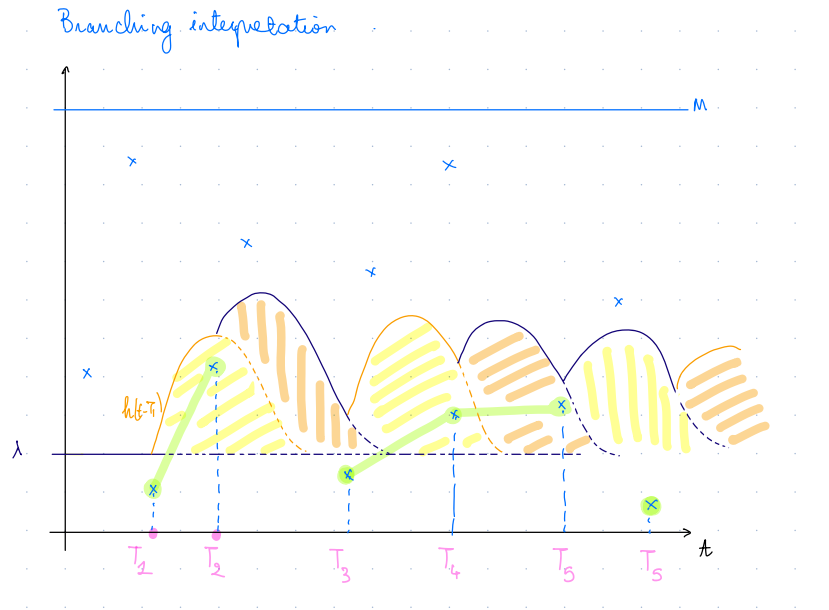


Figure 3.2: Branching interpretation of a linear Hawkes process. Genealogical trees appear in green.

through time. This construction has been used notably by [BT07] to obtain large deviation results but also by [RBR07] to derive deviation inequalities. These deviation inequalities are important to build and study Bayesian estimators for Hawkes processes [DRR20, HRBR15, SRR24]. Note that an important issue is that the cluster superposition only holds in the linear case, and finding alternative tools to derive deviation inequalities in non linear setting was the motivation of our work [CGMT20] on Hawkes processes.

**Limit theorem for the linear Hawkes process** Let us now focus on the subcritical case  $\|h\|_1 < 1$ , which ensures that a stationary Hawkes process is well defined [BM96]. The limit behaviour of the number of events in  $[0, t]$  is well understood. Let us first remark that thanks to the linearity,

$$\mathbb{E}(N_t) = \mathbb{E} \left( \int_0^t \Lambda_s ds \right) = \lambda t + \int_0^t h(t-s) \mathbb{E}(N_s) ds.$$

As a consequence,  $\mathbb{E}(N_t)$  satisfies a Volterra renewal equation [Fel91, Bru17], which admits an explicit solution

$$\mathbb{E}(N_t) = \lambda t + \int_0^t \lambda s H(t-s) ds,$$

where  $H(t) = \sum_{n=1}^{\infty} h^{*n}(t)$  and  $h^{*n}$  is the  $n$ th convolution of  $h$

$$h^{*n}(t) = \int h(t-s) h^{*(n-1)}(s) ds.$$

In the limit  $t \rightarrow \infty$ , we deduce that the expected number of events on  $[0, t]$  satisfies

$$\frac{\mathbb{E}(N_t)}{t} \xrightarrow[t \rightarrow \infty]{} \frac{\lambda}{1 - \|h\|_1}.$$

From this result, we can derive a law of large number

$$\frac{N_t}{t} \xrightarrow[t \rightarrow \infty]{} \frac{\lambda}{1 - \|h\|_1} := \mu \quad a.s. , \quad (3.3)$$

and a Central limit theorem

$$\frac{N_t - \mu t}{\sqrt{t}} \xrightarrow[t \rightarrow \infty]{} \mathcal{N}(0, \sigma^2) \quad \text{with } \sigma^2 = \frac{\lambda}{(1 - \|h\|_1)^3} ,$$

where the convergence hold in law (see e.g. [DVJ06]). Bacry *et al.* [BDHM13] have obtained additionally a functional Central limit theorem in the multivariate case and a large deviation principle was proved in [BT07] (see also [Zhu13, GZ21]).

Outside the linear case, according to the general seminal paper by Brémaud and Massoulié [BM96], if  $\Phi$  is  $L$ -Lipschitz and  $L \|h\|_{L^1(du)} < 1$ , there exists a unique stationary version of the Hawkes process and

$$\frac{N_t}{t} \xrightarrow[t \rightarrow \infty]{} \mu = \mathbb{E}_s[N^h([0, 1])] \quad a.s. , \quad (3.4)$$

where  $\mathbb{E}_s$  denotes the expectation with respect to the stationary ergodic distribution.

In [Zhu13], Zhu proved a functional CLT at equilibrium but obtaining an explicit expression for  $\mu$  and the limiting variance can rapidly become a difficult task since few is known on the stationary distribution.

## 3.2 A renewal structure for inhibited Hawkes processes

As presented above, the linear Hawkes process only models an excitatory behaviour as occurrences increases the probability of future occurrences. A main interest for multiple applications, and notably in neuroscience is to add a possible inhibition, reducing the probability of future events. To model such inhibitory behaviour, a possibility is to assume that the function  $h$  is signed. The negative values then reduce the intensity and correspond to the inhibition. In this case, the activation function  $\Phi$  cannot be chosen as a linear function since the intensity should remain nonnegative and we will focus on the case where  $\Phi(x) = (\lambda + x)_+$  with  $(y)_+ = \max(y, 0)$  which corresponds to a simple extension of the linear case.

In order to underline the dependence in the function  $h$ , we will denote by  $N^h$  the Hawkes process solution of the stochastic differential equation

$$\begin{cases} N_t^h = \int_{(0,t) \times (0,+\infty)} \delta_u \mathbf{1}_{\theta \leq \Lambda^h(u)} Q(du, d\theta) \\ \Lambda^h(u) = \left( \lambda + \int_{(0,u)} h(u-s) N^h(ds) \right)^+ , \quad u > 0, \end{cases} \quad (3.5)$$

where  $\lambda > 0$  is an immigration rate,  $h : (0, +\infty) \rightarrow \mathbb{R}$  is a signed measurable function. Note that when  $h$  is positive, this definition is the same as (3.2).

### 3.2.1 Couplings with linear Hawkes processes

A first question is to couple the nonlinear Hawkes process with a linear one in order to obtain upper and lower bounds. The intuition being that since inhibition reduces the intensity, an

inhibited Hawkes process should have less points than a non inhibited one. We obtained in [CGMT20] the following upper bound.

**Proposition 3.2.1.** *Let  $Q$  be a  $(\mathcal{F}_t)_{t \geq 0}$  - two-dimensional Poisson point process on  $(0, +\infty) \times (0, +\infty)$  with unit intensity. If  $N^h$  and  $N^{h^+}$  are two solutions of (3.5) associated respectively with the signed function  $h$  and the positive function  $h^+ = \max(h(\cdot), 0)$ , then, in the sense of measures,  $N^h \leq N^{h^+}$ , meaning that for all  $0 \leq s \leq t < +\infty$ ,  $N^h([s, t]) \leq N^{h^+}([s, t])$  almost surely.*

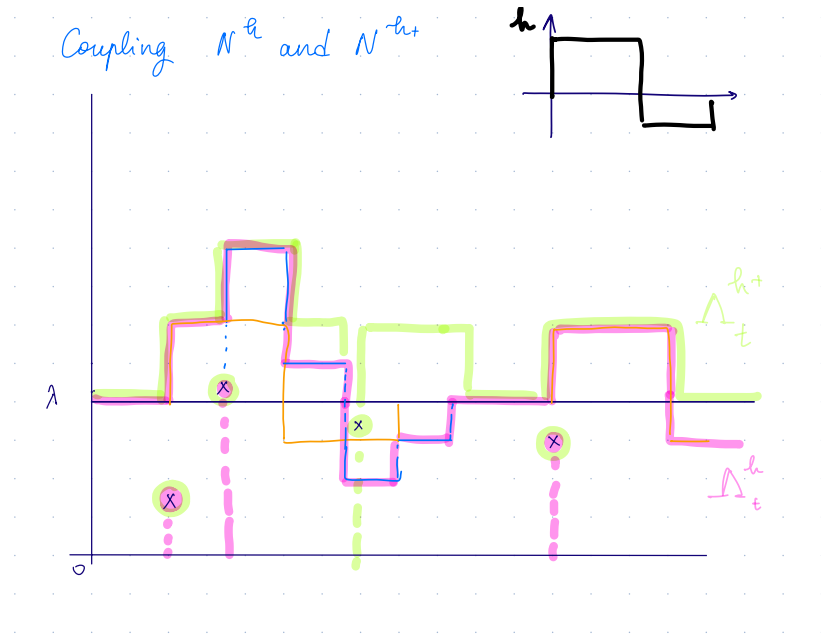


Figure 3.3: Coupling of two trajectories of  $N^h$  and  $N^{h^+}$  constructed with the same Poisson point measure  $Q$ .

The proof of the upper bound consists in showing that the intensity  $\Lambda^h$  and  $\Lambda^{h^+}$  of the Hawkes process associated with  $h$  and  $h_+$  respectively satisfy

$$\Lambda_t^h \leq \Lambda_t^{h^+}, \quad a.s.$$

This order relationship can actually be generalized to any positive function larger than  $h$ , but here the choice of  $h_+$  corresponds to the smallest upper bound. Note however that the result does not hold with a signed function  $g$  such that  $h \leq g$ .

The question of the lower bound is more delicate and we only obtain a partial result in [CCC22].

**Proposition 3.2.2.**

*Let  $h$  be a function with finite support included in  $[0, L]$ . Let  $\lambda > 0$  and define  $g = -\lambda \mathbf{1}_{[0, L]}$ . One can find a coupling of two Hawkes processes  $N^h$  and  $N^g$ , respectively associated with the reproduction functions  $h$  and  $g$ , such that for any  $t \geq 0$ :*

$$N_t^h \geq N_t^g \quad a.s.$$

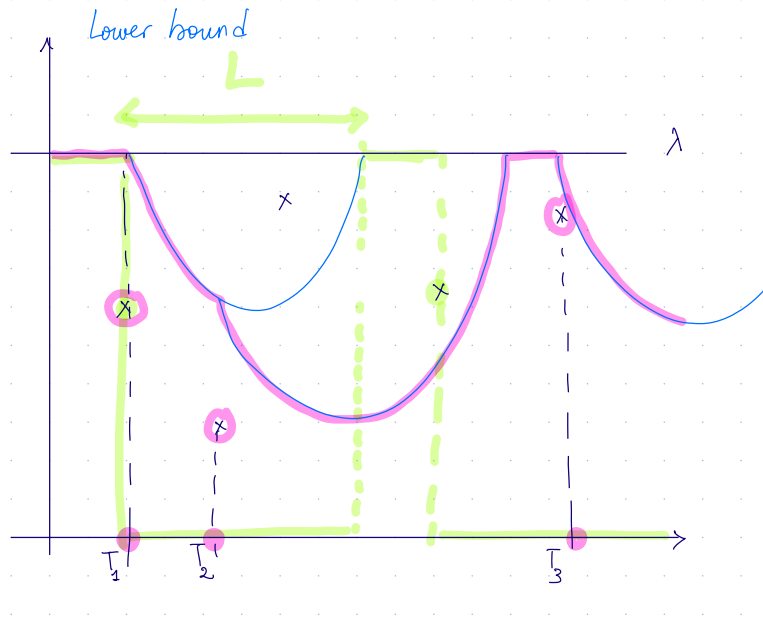


Figure 3.4: Coupling of two trajectories of  $N^h$  and  $N^{h+}$  constructed with the same Poisson point measure  $Q$ .

Note that this comparison result is weaker than the upper bound via  $h^+$ , since we do not have  $N^h([s, t]) \geq N^g([s, t])$  for all  $s \leq t$ , but only for  $s = 0$ . On Figure 3.4, we can observe that the intensities are not ordered any more which explains why a strong coupling is not possible.

### 3.2.2 A renewal structure for finite memory cases

Our main contribution in [CGMT20] is to introduce a renewal structure for nonlinear Hawkes processes whose construction relies both on the coupling with a linear process and the cluster construction of the latter. This renewal structure, introduced for compactly supported reproduction functions, has been extended to other settings by [Gra21, SRR24]. Alternative constructions have also been proposed in [CSSBW17, Raa19].

We focus on the case where  $h$  is a signed function with compact support included in  $[0, L]$  (i.e.  $h(t) = 0, \forall t > L$ ) and such that  $\|h_+\|_1 < 1$ . We can rewrite

$$\begin{aligned} \Lambda^h(t) &= \left( \lambda + \int_0^t h(t-s) N^h(ds) \right)_+ \\ &= \left( \lambda + \int_{t-L}^t h(t-s) N^h(ds) \right)_+ . \end{aligned}$$

This translates that  $N^h|_{(t, +\infty)}$  depends on  $N^h|_{(0, t]}$  only through the events occurring in  $(t-L, t]$ . To formalize this idea, let us define the shift operator  $S_t$  for counting measures on  $\mathbb{R}$

$$S_t : N \mapsto S_t N \triangleq N(\cdot + t) \in \mathcal{N}(\mathbb{R}),$$

and built an auxiliary process  $Y = (Y_t)_{t \geq 0}$  defined by

$$Y_t = (S_t N^h)|_{(-L, 0]} = N^h|_{(t-L, t]}(\cdot + t).$$

The measure  $Y_t$  corresponds to the point process  $N^h$  on the time window  $(t - L, t]$ , shifted back to  $(-L, 0]$ .

We prove that  $(Y_t)_{t \geq 0}$  is a strong  $(\mathcal{F}_t)_{t \geq 0}$ -Markov process with initial condition  $Y_0 = N^0|_{(-L, 0]}$  and sample paths in the Skorohod space  $\mathbb{D}(\mathbb{R}_+, \mathcal{N}((-L, 0]))$ . Furthermore, we obtain that if  $T$  is a stopping time such that  $N^h|_{(T-L, T]} = \emptyset$ , then  $N^h|_{(T, +\infty)}$  is independent of  $N^h|_{(-\infty, T]}$  and behaves as  $N^h$  started from  $\emptyset$  and translated by  $T$ .

We therefore define the first hitting time of  $\emptyset \in \mathcal{N}((-L, 0])$  for  $Y$ , given by

$$\tau = \inf\{t > 0 : Y_{t-} \neq \emptyset, Y_t = \emptyset\} = \inf\{t > 0 : N^h[t - L, t] \neq \emptyset, N^h(t - L, t] = \emptyset\}. \quad (3.6)$$

In order to prove the existence of a renewal structure, we need to justify that  $\tau$  is almost surely finite. Using results from the queuing theory [Asm03, Tak62], we obtain that  $\tau$  admits exponential moments.

**Proposition 3.2.3.** *Under our assumptions, the stopping time  $\tau$  given by (3.6) satisfies*

$$\forall \alpha < \min\left(\lambda, \frac{\|h^+\|_1 - \log(\|h^+\|_1) - 1}{L(h_+)}\right), \quad \mathbb{E}_m(e^{\alpha\tau}) < +\infty.$$

*In particular  $\tau$  is finite,  $\mathbb{P}_m$ -a.s., and  $\mathbb{E}_m(\tau) < +\infty$ .*

The proof relies on the coupling described in Proposition 3.2.1 ensuring that the stopping time  $\tau^+$  associated with  $N^{h^+}$  is always larger than  $\tau$ . Then, from the cluster construction, we can interpret  $\tau^+$  as the time where a serving queue empties. This specific queue measures the length of survival of the ancestral line of each ancestors in the cluster decomposition of the linear Hawkes  $N^{h^+}$ . More precisely the customers are the ancestors and are arrive according to a Poisson point process with parameter  $\lambda$  while their service length is the length of the Galton Watson tree of descendants they generate. As a consequence, when the queue empties, all the  $h(t - T_i)$  are null, where  $T_i$  are the time of clients arrivals.

The strong Markov property of  $Y$  yields a sequence of regeneration times  $(\tau_k)_{k \geq 0}$  that are the successive visits of  $Y$  to the positive recurrent state  $\emptyset$ , defined as follows

$$\begin{aligned} \tau_0 &= 0, \\ \tau_k &= \inf\{t > \tau_{k-1} : Y_{t-} \neq \emptyset, Y_t = \emptyset\}, \quad k \geq 1. \end{aligned} \quad (\text{Successive return times at } \emptyset)$$

We finally provide an ergodic theorem for  $Y$ .

- a) The  $\tau_k$  for  $k \geq 0$  are finite stopping times, a.s.
- b) The cycles  $(Y_{\tau_{k-1}+t})_{t \in [0, \tau_k - \tau_{k-1}]}$  for  $k \geq 0$  are i.i.d. and distributed as  $(Y_t)_{t \in [0, \tau]}$  under  $\mathbb{P}_\emptyset$ . In particular their durations  $(\tau_k - \tau_{k-1})_{k \geq 1}$  are distributed as  $\tau$  under  $\mathbb{P}_\emptyset$ , and  $\lim_{k \rightarrow +\infty} \tau_k = +\infty$  a.s.

In particular, this proves the existence of an invariant measure for  $Y_t$ .

**Theorem 3.2.4.** *With our assumptions, the strong Markov process  $(Y_t)_{t \geq 0}$  admits a unique invariant law  $\pi_A$  defined for every Borel nonnegative function  $f$  on  $\mathcal{N}((-A, 0])$  by*

$$\pi_A f = \frac{1}{\mathbb{E}_\emptyset(\tau)} \mathbb{E}_\emptyset \left( \int_0^\tau f(Y_t) dt \right).$$

Moreover  $\pi_A\{\emptyset\} = 1/(\lambda \mathbb{E}_\emptyset(\tau))$ .

We exploited this renewal structure in order to obtain different results on the long time behaviour of functionals of the nonlinear Hawkes process. In the next section, I will focus on the PhD work of Laetitia Colombani which uses this renewal structure to study the limiting behaviour of the counting process associated with the Hawkes process (3.2)  $t \mapsto N_t^h = N^h([0, t])$  (see [CCC22]).

**Interpretation as a cumulative process** Let me precise the renewal structure introduced above in the case where we are only interested to the counting process  $N_t$ . Recall that the jump time of the Hawkes process are denoted by  $(T_n)_{n \geq 1}$ . The renewal process introduced above allows to define recursively

$$\tau_1 = \inf\{t > T_1, N^h((t - L), t) = 0\},$$

that is the first time after  $T_1$  such that there has been no jump during a time  $L$ . We also set

$$S_0 = 0 \quad \text{and} \quad S_1 = \tau_1.$$

Let us now define

$$W_1 = N^h([T_1, S_1]) = N^h([0, S_1]),$$

the number of jumps of the process in this first time window.

Recursively let  $i \in \mathbb{N}^*$  such that  $(\tau_1, W_1), \dots, (\tau_i, W_i)$  are well defined (and a.s. finite). Let  $S_i = \sum_{k=1}^i \tau_k$  and define the first time in the  $i + 1$ th window  $T_1^{i+1} = T_{W_1 + \dots + W_{i+1}}$ , then

$$\tau_{i+1} = \inf\{t > T_1^{i+1}, N^h((t - L), t) = 0\} - S_i, \quad (3.7)$$

Notice that there is at least one jump in  $[S_i, S_i + \tau_{i+1}]$ . We finally introduce the number of jumps in the  $(i + 1)$ 'th window as

$$W_{i+1} = N^h([U_1^{i+1}, S_i + \tau_{i+1}]) = N^h([S_i, S_i + \tau_{i+1}]), \quad (3.8)$$

and rename the associated jump times as:

$$T_j^{i+1} = T_{W_1 + \dots + W_{i+j}}, \quad \forall j \in \{1, \dots, W_{i+1}\}.$$

By construction and using the renewal property shown above, the  $(\tau_i, W_i)_i$  are i.i.d. random variables. Moreover, the time between the beginning of a window and the first point of this window  $(T_1^i - S_{i-1})$  follows an exponential law with parameter  $\lambda$ .



### 3.2.3 Limiting results

Using the above decomposition, we can now rewrite

$$N_t^h = \sum_{i=1}^{\infty} \mathbf{1}_{T_i \leq t} = \sum_{i=1}^{\infty} \sum_{j=1}^{W_i} \mathbf{1}_{T_i^j \leq t}.$$

We introduce the renewal process associated to the  $S_i$ 's

$$M_t^h := \sum_{i=1}^{\infty} \mathbf{1}_{S_i \leq t}.$$

and now obtain that

$$N_t^h = \sum_{i=1}^{M_t^h} W_i + R_t^h, \quad (3.9)$$

where the remaining term  $R_t^h \leq W_{M_t^h+1}$ , the  $W_i$ 's being i.i.d.. The first term in (3.9)  $\hat{N}_t^h = \sum_{i=1}^{M_t^h} W_i$  is an example of a *cumulative processes* as it can be called in the literature.

A Law of large numbers and a central limit theorem for general cumulative processes can be found in [Asm03]. In order to use these results and extend them to  $N_t^h/t$  the key is to control moments for the  $(\tau_i, W_i)$  and the rest term. We obtain in [CCC22]:

**Theorem 3.2.5.** *Consider the Hawkes process  $N^h$  given by (3.2). Then we have the following:*

$$\frac{N_t^h}{t} \xrightarrow[t \rightarrow \infty]{a.s.} \frac{\mathbb{E}[W_1]}{\mathbb{E}(\tau_1)},$$

and furthermore

$$\sqrt{t} \left( \frac{N_t^h}{t} - \frac{\mathbb{E}[W_1]}{\mathbb{E}(\tau_1)} \right) \xrightarrow[t \rightarrow \infty]{} \mathcal{N}(0, \sigma^2),$$

with

$$\sigma^2 = \frac{\text{Var}\left(W_1 - \tau_1 \frac{\mathbb{E}[W_1]}{\mathbb{E}(\tau_1)}\right)}{\mathbb{E}(\tau_1)}.$$

Thanks to our comparison results and to (3.3) we have

$$\frac{\lambda}{1 + \lambda L(h)} \leq \frac{\mathbb{E}[W_1]}{\mathbb{E}(\tau_1)} \leq \frac{\lambda}{1 - \|h^+\|_1}.$$

Exact computations for moments of  $\tau$  and  $W$  are difficult and we only achieved to get information when looking at functions  $h$  of the form  $h = -\lambda \mathbf{1}_{[r, r+A]}$ .

**Further results.** To go further than the central limit theorem, large deviation principle for cumulative processes is studied in [LMZ11] in the special case

$$W_i = F(\tau_i),$$

for some non-negative, bounded and continuous function  $F$  (see the references in [LMZ11] for some previous results in still more specific cases). These results do not directly apply

for Hawkes processes, and we had to establish a more general large deviation principle in [CCC23] which I will not detail here.

In [CGMT20] we also provide limit theorem for functionals of the form

$$\frac{1}{T} \int_0^T f(S_t N^h)|_{(-L,0]} dt,$$

where  $f$  is a real measurable locally bounded function of the trajectory. These results generalize the deviation inequalities of [RBR07] and are important for statistical purposes (see notably recent results of [SRR24]).

### 3.3 Stability of a discrete time Hawkes process with inhibition and finite memory

In this section, I detail results that have been obtained during the PhD thesis of Anthony Muraro, co-supervised with Pascal Maillard.

Our main question was to enhance sufficient conditions on  $h$  providing the existence of a stationary version of Hawkes process. For signed  $h$ , Brémaud and Massoulié [BM96] proved that a stable version of the process exists if  $\|h\|_1 < 1$ . Using the coupling argument in Proposition 3.2.1, we obtained in [CGMT20], that it is sufficient to have  $\|h^+\|_1 < 1$ . Unfortunately, this sufficient criterion does not take into account the effect of inhibition, captured by the negative part of  $h$ .

#### 3.3.1 The discrete time Hawkes process

We choose to consider a simplified, discrete analogue of the Hawkes process with finite memory. Namely we study an auto-regressive process  $(\tilde{X}_n)_{n \geq 1}$  with initial condition  $(\tilde{X}_0, \dots, \tilde{X}_{-p+1})$  such that conditionally on the past  $\mathcal{F}_{n-1} = \sigma(X_{-p+1}, \dots, X_{n-1})$

$$\tilde{X}_n \sim \mathcal{P} \left( \phi \left( a_1 \tilde{X}_{n-1} + \dots + a_p \tilde{X}_{n-p} + \lambda \right) \right), \tag{3.10}$$

where  $\mathcal{P}(\rho)$  denotes the Poisson distribution with parameter  $\rho$ , and  $a_1, \dots, a_p$  are real numbers.

#### Existing results for the linear case

The case where all parameters  $a_1, \dots, a_p$  are nonnegative is well understood. The process  $(\tilde{X}_n)$  is an *Integer-Valued GARCH process* (INGARCH(0, p) process, see [FLO06] and [LZLS23]), also known in the literature as an *Auto-regressive Conditional Poisson process* (ACP(p) see [Hei03]) Its long time behaviour depends on  $\sum_{i=1}^p a_i$ . Namely,

- If  $\sum_{i=1}^p a_i < 1$ , it has already been established that this process admits a stationary version. This result can be understood as a subcritical condition for the branching interpretation of the INGARCH processes [Kir16], in which  $\sum_{i=1}^p a_i$  is the mean number of offspring generated by a living individual.
- If  $\sum_{i=1}^p a_i > 1$ , then  $\tilde{X}_n$  grows exponentially in  $n$  almost surely.

### A first result for the non-linear version

Similarly as for continuous time Hawkes process, the use of a finite memory  $p$  allows to construct an auxiliary Markov chain

$$X_n := (\tilde{X}_n, \tilde{X}_{n-1}, \dots, \tilde{X}_{n-p+1}) \in \mathbb{N}^p. \quad (3.11)$$

In order to study the recurrence of this Markov chain, we use drift criteria, initially introduced by Foster [Fos53], and extensively studied and popularized by Meyn and Tweedie [MT09]. The main challenge of this approach is to construct a Lyapunov function, that is a function  $V$  taking values in  $[1, \infty]$

$$\forall x \in \mathbb{N}^p, \quad \Delta V(x) := \mathbb{E}_x[V(X_1) - V(X_0)] \leq -\varepsilon V(x) + K \mathbf{1}_C(x). \quad (3.12)$$

where  $K$  is a finite constant, and  $C$  a subset of  $\mathbb{N}^p$ . In our case where  $X$  is aperiodic and irreducible in a weak sense, proving regodicity also requires that the set  $C$  is a *small set*, i.e. that there exists  $x_0 \in \mathbb{N}^p$  and  $n \in \mathbb{N}$  satisfying :

$$\inf_{x \in C} P^n(x, x_0) > 0.$$

We refer notably to [DMPS18] for further details and results well adapted to the case of discrete state spaces.

Using a linear function  $V$ , we obtain the following sufficient condition

**Theorem 3.3.1.** *If  $(a_1)_+ + \dots + (a_p)_+ < 1$ , then  $(X_n)$  is a geometrically ergodic Markov chain.*

The condition in Theorem 3.3.1 is an analogue of the sufficient condition for the existence of a continuous-time Hawkes process given in [CGMT20] and the previous section. However, this sufficient condition is quite restrictive because it is limited to parameters strictly less than 1. It is natural to inquire whether, in the case where one of the parameters is greater than 1, one can offset the others parameters sufficiently negatively to obtain stability. This question appears to be quite challenging to resolve, therefore, we confine ourselves in the following to the specific case of  $p = 2$  or 3.

### 3.3.2 Two parameter discrete time Hawkes process

In the case of  $p = 2$ , our model of interest can be written as :

$$\text{conditioned on } \tilde{X}_{n-2}, \tilde{X}_{n-1}: \tilde{X}_n \sim \mathcal{P} \left( \left( a\tilde{X}_{n-1} + b\tilde{X}_{n-2} + \lambda \right)_+ \right), \quad (3.13)$$

with  $a, b \in \mathbb{R}$  and  $\tilde{X}_0, \tilde{X}_1 \in \mathbb{N}$ .

In [CMM24a], we study the behaviour of the associated Markov Chain  $X_n = (\tilde{X}_n, \tilde{X}_{n-1})$  and provide a complete classification. Let us define the function

$$b_c(a) = \begin{cases} 1 & a \leq 0 \\ 1 - a & a \in (0, 2) \\ -\frac{a^2}{4} & a \geq 2 \end{cases},$$

and define the following sets (see Figure 3.5):

$$\mathcal{R} = \{(a, b) \in \mathbb{R}^2 : b < b_c(a)\}, \tag{3.14}$$

$$\mathcal{T} = \{(a, b) \in \mathbb{R}^2 : b > b_c(a)\}. \tag{3.15}$$

Our main result is the following

**Theorem 3.3.2.**    • *If  $(a, b) \in \mathcal{R}$ , then the sequence  $(X_n)_{n \geq 0}$  is geometrically ergodic*  
 • *if  $(a, b) \in \mathcal{T}$ , then  $(X_n)_{n \geq 0}$  is transient.*

We stress out that the condition for ergodicity is not symmetrical in  $a$  and  $b$ . More precisely, for any  $a \in \mathbb{R}$ , the Markov chain  $(X_n)$  can be ergodic provided that  $b$  is chosen small enough, but the converse is not true as soon as  $b > 1$ . This induces that inhibition has a stronger regulating effect when it occurs after an excitation, rather than before.

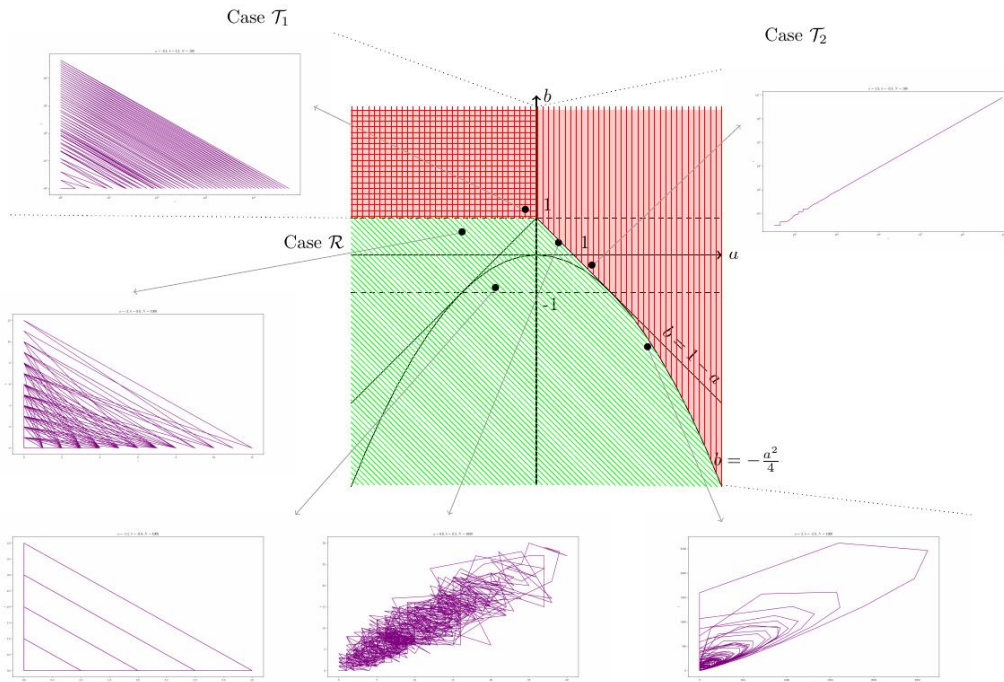


Figure 3.5: The partition of the parameter space described in Theorem 3.3.2. The green region corresponds to  $\mathcal{R}$ , while the red region corresponds to  $\mathcal{T}$ . The smaller figures are typical trajectories of the Markov chain  $(X_n)_{n \geq 0}$  for each region of the parameter space. In the all the simulations, we chose  $\lambda = 1$ .

The proof of recurrence relies on finding adequate Lyapunov functions. Depending on the parameters  $(a, b)$ , we use, a linear function  $V(i, j) = \alpha i + \beta j + 1$ , a quadratic function  $V(i, j) = \alpha i^2 + \beta j^2 + \gamma ij + 1$  or  $V(i, j) = \frac{i}{j+1} + 1$  a function of the angle.

For transience, we exhibit trajectories that grow to infinity. In the case  $b > 1$  and  $a < 0$ , let us remark than starting from a state  $(i, 0)$ , the parameter of the Poisson random variable for the next step is  $ai + \lambda < 0$  if  $i$  is large enough since  $a < 0$ . As a consequence, the chain

goes to  $(0, i)$  and then to  $(j, 0)$  where  $j \sim \mathcal{P}(bi + \lambda)$ . Using the concentration of the Poisson random variable,  $j \geq bi > i$  with large probability since  $b > 1$ .

For the other cases, we use a comparison with the naturally associated linear and deterministic recursive sequence  $u_{n+1} = au_n + bu_{n-1} + \lambda$ . For these choice of parameters, its characteristic polynomial  $X^2 - aX - b$  admits an eigenvalue larger than 1, which induces that the solution  $u_n$  diverges.

### 3.3.3 Three-parameter discrete-time Hawkes process

In the case where  $p = 3$ , we relabel similarly the parameters as  $a_1 = a$ ,  $a_2 = b$  and  $a_3 = c$ . The discrete time process  $(\tilde{X}_n)_{n \geq 1}$  is now defined with initial condition  $(\tilde{X}_0, \tilde{X}_{-1}, \tilde{X}_{-2})$  as  $n \geq 1$ :

$$\text{conditioned on } \tilde{X}_{n-3}, \tilde{X}_{n-2}, \tilde{X}_{n-1}: \tilde{X}_n \sim \mathcal{P}\left(\left(a\tilde{X}_{n-1} + b\tilde{X}_{n-2} + c\tilde{X}_{n-3} + \lambda\right)_+\right). \quad (3.16)$$

In that case, we did not obtain a complete characterization but we complete in [CMM24b] the results of Theorem 3.3.1 by considering cases where the three parameters have different signs and one of them is larger than 1 .

**Link with the linear recurrent equation** As in the case of two parameters, due to the concentration property of the Poisson distribution, we expect that as long as the parameter of the Poisson random variable in the definition of  $(\tilde{X}_n)_{n \geq 0}$  remains positive, the process behaves similarly to the deterministic sequence  $(\tilde{u}_n)_{n \geq 0}$ , which is defined as follows:

$$\tilde{u}_n = a\tilde{u}_{n-1} + b\tilde{u}_{n-2} + c\tilde{u}_{n-3} + \lambda. \quad (3.17)$$

It is a well-known fact that the roots of the polynomial

$$P(X) := X^3 - aX^2 - bX - c, \quad (3.18)$$

plays a significant role in the identification of sequences that satisfy the recurrence (3.17). More precisely, the stability of the linear recurrence  $(\tilde{u}_n)_{n \geq 0}$  (in the sense that sequences satisfying recurrence (3.17) remain bounded for all  $n \in \mathbb{N}$ ) is characterized by the condition :

$$\max\{|\zeta|, P(\zeta) = 0\} < 1. \quad (3.19)$$

In our study, we used the *discriminant* of the polynomial  $P$  which can be easily computed from the coefficients as:

$$\text{Disc}(P) := a^2b^2 + 4b^3 - 4a^3c - 18abc - 27c^2. \quad (3.20)$$

The sign of  $\text{Disc}(P)$  determines the number of roots of  $P$  that are complex. For a degree 3 polynomial, on the one hand, if  $\text{Disc}(P) < 0$ , the polynomial  $P$  has an unique simple real root, and two complex conjugate roots. On the other hand, if  $\text{Disc}(P) \geq 0$ , all the roots of  $P$  are real numbers, and they are simple when  $\text{Disc}(P) > 0$ . For more details about the discriminant, we refer the reader to chapter 12 of [GKZ94].

Returning to our comparison involving deterministic sequences  $(\tilde{u}_n)_{n \geq 0}$  satisfying (3.17), when  $\text{Disc}(P) < 0$ , since the polynomial  $P$  has 2 complex roots, we know that the solutions  $(\tilde{u}_n)_{n \geq 0}$  will exhibit oscillatory behaviour, potentially diverging under stronger conditions. Consequently,  $(\tilde{u}_n)_{n \geq 0}$  will eventually become negative. This is where the positive part in the definition of  $(\tilde{X}_n)_{n \geq 0}$  comes into play : by truncating the parameter value of the Poisson random variable to 0, we can expect a stabilizing effect on the asymptotic behaviour of this process.

**Transience results** We start by considering cases where  $b$  or  $c$  are larger than 1. We derive sufficient conditions that guarantee that the process  $(\tilde{X}_n)_{n \geq 0}$  will diverge in the sense that almost surely,

$$\tilde{X}_n + \tilde{X}_{n+1} + \tilde{X}_{n+2} \xrightarrow[n \rightarrow +\infty]{} +\infty.$$

In terms of the Markov chain  $X_n = (\tilde{X}_n, \tilde{X}_{n+1}, \tilde{X}_{n+2})$  we prove:

**Theorem 3.3.3.** *If  $a, b, c \in \mathbb{R}$  satisfy one of the following condition*

1.  $a, b < 0$  and  $c > 1$ , or
2.  $b > 1$  and  $ab + c < 0$

*then the Markov chain  $(X_n)$  is transient.*

Theorem 3.3.3 states that when  $b$  or  $c$  is greater than 1, and the other parameters are negative,  $(X_n)_{n \geq 0}$  is transient. To rephrase, there is no inhibition, no matter how strong, that yields a stable process as long as  $b$  or  $c$  is greater than 1.

**Recurrence results** We prove that for  $a > 1$ , it is possible to choose,  $b$  and  $c$  sufficiently negative, for the process  $X_n$  to be ergodic.

**Theorem 3.3.4.** *Let  $a > 1$ . If  $b < 0$  and  $c < 0$  are such that  $\text{Disc}(P) < 0$ , then  $(X_n)_{n \geq 0}$  is a geometrically ergodic Markov chain.*

Theorem 3.3.4 states that for any given excitation  $a > 1$ , there exist sufficiently strong inhibitions  $b, c < 0$  such that the parameters of the process  $(\tilde{X}_n)_{n \geq 0}$  satisfy the assumptions of Theorem 3.3.4. This completes the results of Theorem 3.3.3 and implies that the order of parameters is important, introducing an asymmetry in their roles, which is a non trivial property of the process  $(\tilde{X}_n)_{n \geq 0}$  given its initial definition.

**Conjecture and numerical illustrations** The proof of Theorem 3.3.4 is based on a Lyapunov argument. Under assumptions  $a > 1, b < 0, c < 0$  such that  $\text{Disc}(P) < 0$ , we proved that there exists a Lyapunov function for the Markov chain  $(X_n)$ , and in a second time, we identified a *small* set. The construction of the Lyapunov function requires  $c < 0$  while  $b < 0$  is used to construct an appropriate small set. Based on the numerical experiments, we believe that it is possible to slightly relax the assumptions of Theorem 3.3.4 allowing  $b \leq 1$ . Note that furthermore if  $b > 1$  (but  $\text{Disc}(P) < 0$ ) then using properties of the discriminant,

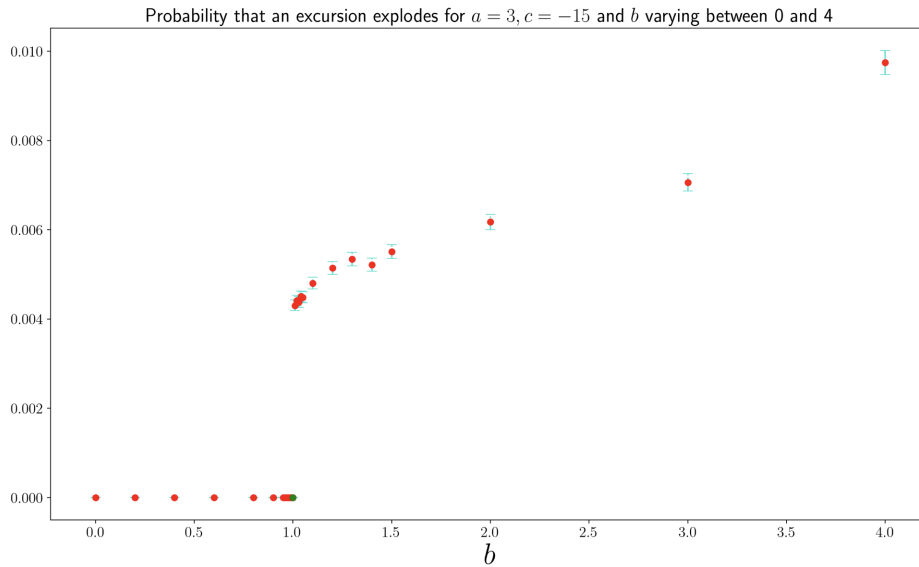


Figure 3.6: Plot (in red) of the proportion of excursions that have exploded relative to the total number of excursions, for some  $b \in [0, 4]$ . For clarity, we put a green dot when  $b = 1$ . Clopper-Pearson type confidence intervals (with  $\alpha = 0.01$ ) around the calculated values have also been plotted in blue. It is noteworthy that the positive proportions exhibit remarkably low values : even within the context where  $b > 1$ , the number of excursions that have exploded remains very small compared to the overall number of excursions.

it is possible to prove that  $ab + c < 0$ . Theorem 3.3.3 *ii*) then ensures that the Markov chain  $(X_n)$  is transient.

In all the following simulations we chose the initial condition  $(X_{-2}, X_{-1}, X_0) = (0, 0, 0)$  and  $\lambda = 1$ . The state  $(0, 0, 0)$  is of notable interest since it is an accessible state: for any values of  $a, b, c$ , it is always possible to reach this state from any other state in 3 steps. We illustrate the fact that the stopping time  $\tau_0 := \inf\{n \geq 1 \mid X_n = (0, 0, 0)\}$ , satisfies

$$\begin{cases} \mathbb{P}_{(0,0,0)}(\tau_0 = +\infty) = 0 & \text{if } b \leq 1, \\ \mathbb{P}_{(0,0,0)}(\tau_0 = +\infty) > 0 & \text{if } b > 1. \end{cases}$$

We estimate the probability that an excursion starting from  $(0, 0, 0)$  explodes from Monte Carlo estimator based on  $10^6$  repetitions and provide confidence intervals using Clopper-Pearson intervals which are *exact* confidence intervals for binomial proportion  $p$  [Thu14].

### 3.4 Perspectives

The research prospects arising from this part of my work are numerous and are for the most part the subject of the ANR JCJC HAPPY that I coordinate. In particular, there are still many interesting questions concerning the study of the renewal renewal and, in particular, the connection between the moments of the variables  $\tau$  and  $W$  and the parameters  $\lambda$  and  $h$  of the Hawkes process. Similar questions arise in a multivariate setting, since inhibition plays an important role in the regulation of activity in complex networks. Another difficulty

(and a long-standing issue) is to understand the interplay between inhibition and the size and geometry of the network.

From a statistical point of view, while the question of estimating the parameters of nonlinear Hawkes processes has been thoroughly investigated in univariate and in multivariate settings from a Bayesian perspective [SRR22] as well as in parametric cases [BHS21], little theoretical work exists to test for the presence or absence of inhibition or for specific non-linearity [DLP22]. This is crucial for understanding the structure of interacting networks and the associated patterns of regulation.

I'm going to detail here two directions that seem to me to be the most advanced.

### 3.4.1 On the discrete model

Let me recall the discrete Hawkes process  $(\tilde{X}_n)_{n \geq 1}$  defined as

$$\tilde{X}_n \sim \mathcal{P} \left( \phi \left( a_1 \tilde{X}_{n-1} + \cdots + a_p \tilde{X}_{n-p} + \lambda \right) \right),$$

conditionally on the past  $\{X_{n-p}, \dots, X_{n-1}\}$ .

**Larger memories** The generalization of the Lyapunov techniques used for the discrete model (3.16) for  $p = 2$  and  $p = 3$  remains an open and difficult question. For example, a better understanding of the eigenvalues of the polynomial  $X^p + a_1 X^{p-1} + \cdots + a_p$  seems necessary, even if we have not yet succeeded in finding necessary and sufficient conditions on these eigenvalues to ensure the recurrence of the Markov chain transience.

**Critical cases** A second interesting question concerns the critical cases. As we have seen for example when  $p = 2$ , the critical boundary is composed of different zones depending on the signs of the coefficients. This should be a general observation in higher dimensions. The simplest case of interest is the case where all coefficients are positive and  $\sum_{i=1}^p a_i = 1$ . In this setting, we observed in dimension  $p = 2$  that

$$Y_n(t) = \frac{X_{[nt]} + bX_{[nt]-1}}{n},$$

converges as  $n \rightarrow \infty$  in distribution to the solution  $Y_t$  of

$$dY_t = \lambda + \frac{1}{\sqrt{1+b}} \sqrt{Y_t} dW_t,$$

where  $W$  is a standard Brownian motion. We expect to generalize this result to larger memories. Note also that a similar convergence has recently been proved by [HX24] for the continuous-time Hawkes process.

The question of what happens for critical cases with a negative parameter remains open and may lead to different limiting objects.

**Link with truncated Galton Watson tree** A third question that has long intrigued me is how to couple a nonlinear Hawkes process with a linear one without breaking the underlying branching structure. More specifically, I expect that an appropriate coupling



would prune the branching structure of the linear process. This would allow to obtain more precise information about the long-term stability conditions of the Hawkes process. The discrete case seems to be easier to deal with, although it leads to consider Galton-Watson trees in a random environment, where the environment depends on the past of the process. Connections with Polya's urns or the elephant walk are worth looking at.

### 3.4.2 More generally on non linear Hawkes

In my previous work I have mainly considered the case where inhibition is modeled by using a negative function  $h$  and the non-linear function  $\Phi = (\cdot)_+$ . An alternative model for inhibition has been proposed in [DLP22] adding a multiplicative weight in  $(0, 1)$  to the basal intensity. In their article, the authors consider a mean-field case, motivated by applications in neuroscience, with two types of neurons  $A$  and  $B$ . In the large population limit, they obtain two coupled inhomogeneous Poisson processes  $N_A$  and  $N_B$ , for which the intensity of the type  $A$  neurons is given by

$$\lambda_A(t) = \left( \mu + \int_0^t h_A(t-s) dN_A(s) \right) \Phi_{BA} \left( \int_0^t g(t-s) dN_B(s) \right).$$

Here  $\Phi_{BA} : \mathbb{R}_+ \rightarrow [0, 1]$  encodes the inhibition of the population  $B$  with respect to  $A$ . The interest of this model is to preserve positivity and order relations.

Together with Céline Duval, Eric Luçon and Eva Löcherbach, we want to use this framework to study the refractory behaviour of neurons, i.e. the fact that spikes of the same neuron cannot follow each other too quickly. To this end, we consider a mean-field model in which each particle exerts excitation on other particles and additionally an auto-inhibition in a multiplicative way. This amounts to looking at a family of  $K$  Hawkes processes  $(N_i(t))_{1 \leq i \leq K}$  whose intensities write

$$\Lambda_i(t) = \left( \mu + \sum_{j=1}^K \frac{1}{K} \int_0^t g(t-s) dN_j(s) \right) \Phi \left( \int_0^t f(t-s) dN_i(s) \right).$$

In the limit  $K \rightarrow \infty$ , some stochasticity remains and we have to consider a nonlinear stochastic differential equation for the intensity, as is the case for the closely related models [FL16, LM22]. Our main questions are the existence of the limiting object as well as its long time behaviour.

---

## Stochastic algorithms for quantile and superquantile estimation. Application to financial data.

---

In this final section, I present a body of work on the recursive estimation of quantiles and superquantiles of random variables that summarizes the results obtained with Bernard Bercu, Sébastien Gadat and Lorick Huang in three articles [BCG21, CG21, CGH24]. At the end of this section I also present research perspectives related to these works.

### 4.1 Introduction

Estimating quantiles has a longstanding history in statistics and probability. Except in parametric models where explicit formula are available, the estimation of quantiles is a real issue. The most common way to estimate quantiles is to use order statistics, but this method has a significant numerical cost (see among others [Bah66, Gho71]). Another strategy is to use stochastic approximation algorithms, following the pioneering work of Robbins and Monro [RM51].

Let us describe precisely the framework we are studying. Let  $X$  be an integrable continuous real valued random variable with strictly increasing cumulative distribution function  $F$  and probability density function  $f$ . For any  $\alpha \in ]0, 1[$ , the quantile  $\theta_\alpha$  of order  $\alpha$  of  $F$  satisfies

$$F(\theta_\alpha) = \mathbb{P}(X \leq \theta_\alpha) = \alpha,$$

whereas the superquantile  $\vartheta_\alpha$  of order  $\alpha$  is defined by

$$\vartheta_\alpha = \mathbb{E}[X \mid X \geq \theta_\alpha] = \frac{\mathbb{E}[X \mathbf{1}_{\{X \geq \theta_\alpha\}}]}{\mathbb{P}(X \geq \theta_\alpha)} = \frac{\mathbb{E}[X \mathbf{1}_{\{X \geq \theta_\alpha\}}]}{1 - \alpha}.$$

Both the quantile and the superquantile provide information on the tail of the distribution of the random variable  $X$ . In a financial context, quantiles and superquantiles are called *value*

at risk (V@R) and conditional value at risk (CV@R) and have become increasingly popular as measures of risk [RU00, RU02]. I will come back to this interpretation in Section 4.3.

Our first aim is to estimate quantiles and superquantiles simultaneously using the Robbins-Monro algorithm and to provide results on the convergence rate of these algorithms.

**A brief reminder on Robbins Monro algorithm** We recall that the Robbins Monro algorithm [RM51] was developed in order to find roots of a function that can be written as an expectation. More precisely, we are looking for  $\lambda^*$  that solves

$$0 = h(\lambda) = \mathbb{E}[H(X, \lambda)].$$

Given a sequence of i.i.d random variables  $(X_n)_{n \geq 1}$  with same law as  $X$ , the Robbins Monro algorithm constructs a sequence  $(\lambda_n)_{n \geq 0}$  by

$$\lambda_{n+1} = \lambda_n - \gamma_{n+1} H(\lambda_n, X_{n+1}),$$

where  $(\gamma_n)_{n \geq 1}$  is a sequence of step-sizes satisfying  $\sum \gamma_n = +\infty$  and  $\sum \gamma_n^2 < \infty$  in the standard framework [Duf97]. These assumptions on the step sizes sequence are most of the time required to ensure almost sure convergence of the algorithm.

In many cases, it is useful to write the recursive equation as a drift-martingale decomposition. Let  $(\mathcal{F}_n)_{n \geq 0}$  be the associated filtration, we can write

$$\begin{aligned} \lambda_{n+1} &= \lambda_n - \gamma_{n+1} \mathbb{E}[H(\lambda_n, X_{n+1}) | \mathcal{F}_n] - \gamma_{n+1} (H(\lambda_n, X_{n+1}) - \mathbb{E}[H(\lambda_n, X_{n+1}) | \mathcal{F}_n]) \\ &= \lambda_n - \gamma_{n+1} h(\lambda_n) + \gamma_{n+1} \Delta M_{n+1}, \end{aligned}$$

where  $\Delta M_n$  is a sequence of martingale increments

$$\Delta M_{n+1} = h(\lambda_n) - H(\lambda_n, X_{n+1}).$$

When  $h$  is a gradient, this expression corresponds to a stochastic gradient descent.

Note furthermore that since the objective point is  $\lambda^*$  such that  $h(\lambda^*) = 0$  we can use a Taylor expansion to obtain a linearized algorithm

$$\begin{aligned} \lambda_{n+1} - \lambda^* &= \lambda_n - \lambda^* - \gamma_{n+1} [h(\lambda_n) - h(\lambda^*)] + \gamma_{n+1} \Delta M_{n+1}, \\ &= (\lambda_n - \lambda^*) (1 - \gamma_{n+1} h'(\lambda_n)) + \gamma_{n+1} \Delta M_{n+1} + \text{rests terms}, \end{aligned}$$

Such recursion will be the main ingredient to obtain asymptotic information on the algorithm. An interesting case corresponds to  $h'(\lambda^*) > 0$  that creates a contraction. In the context of a gradient descent this amounts to consider a convex function having nice properties for optimization.

**Application to the quantile and superquantile estimation** A standard application of this recursive method is the estimation of a quantile  $\theta_\alpha$  (or of more general geometric medians). Indeed, the quantile  $\theta_\alpha$  solves the following equation in  $\theta$

$$\mathbb{E}[\mathbf{1}_{X \leq \theta}] - \alpha = \mathbb{E}[\mathbf{1}_{X \leq \theta} - \alpha] = 0.$$

The Robbins Monro algorithm associated to the function  $\theta \mapsto \mathbb{E}[\mathbf{1}_{X \leq \theta} - \alpha]$  then constructs a sequence  $(\theta_n)_{n \geq 0}$  as

$$\theta_{n+1} = \theta_n - a_n (\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha).$$

For superquantile approximation, different approaches have been proposed. The more naive one consists in considering that  $\vartheta_\alpha$  solves the following equation in  $\vartheta$

$$\frac{\mathbb{E}[X \mathbf{1}_{\{X \geq \theta_\alpha\}}]}{1 - \alpha} - \vartheta = 0,$$

and built the associated Robbins Monro algorithm. Note however that this strategy requires to know the quantile  $\theta_\alpha$ .

As a consequence, an alternative is to estimate simultaneously  $(\theta_\alpha, \vartheta_\alpha)$  which leads to construct a sequence  $(\theta_n, \vartheta_n)_{n \geq 0}$  as a bi-dimensional Robbins Monro algorithm:

$$\begin{cases} \theta_{n+1} = \theta_n - a_n (\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha) \\ \vartheta_{n+1} = \vartheta_n + b_n \left( \frac{X_{n+1}}{1 - \alpha} \mathbf{1}_{X_{n+1} > \theta_n} - \vartheta_n \right) \end{cases}$$

Alternatively, [BFP09] used a different approach proposed by [RU00, BTT86] with a different update function. Let us define the function

$$L(\theta, x) = \theta + \left( \frac{x - \theta}{1 - \alpha} \right) \mathbf{1}_{x > \theta},$$

and remark that  $\mathbb{E}[L(\theta_\alpha, X)] = \frac{\mathbb{E}[X \mathbf{1}_{\{X \geq \theta_\alpha\}}]}{1 - \alpha}$ . As a consequence the couple  $(\theta_\alpha, \vartheta_\alpha)$  solves

$$\mathbb{E}[L(\theta_\alpha, X)] - \vartheta_\alpha = 0.$$

A main interest for using the function  $L$  in the Robbins Monro algorithm, is that  $\theta \mapsto \mathbb{E}(L(\theta, X))$  is convex. The algorithm proposed in [BFP09] is then given by  $(\theta_n, \tilde{\vartheta}_n)_{n \geq 1}$  where

$$\begin{aligned} \tilde{\vartheta}_{n+1} &= \tilde{\vartheta}_n + b_n (L(\theta_n, X_{n+1}) - \tilde{\vartheta}_n) \\ &= \tilde{\vartheta}_n + b_n \left( \theta_n + \frac{(X_{n+1} - \theta_n)}{(1 - \alpha)} \mathbf{1}_{X_{n+1} > \theta_n} - \tilde{\vartheta}_n \right), \end{aligned}$$

The authors then use the Ruppert-Polyak averaging procedure, introduced in the seminal contributions [Rup88, PJ92], to obtain a central limit theorem for their stochastic algorithm. Note that in their work, the author used a single step sequence  $a_n = b_n$  where we aim at considering a two time scale algorithm which offers more flexibility than the one-time-scale algorithm [Bor97, GPS18, KT04, MP06].

**Existing results** The almost sure convergence  $\lim_{n \rightarrow \infty} \theta_n = \theta_\alpha$  was established by Robbins and Monro [RM51], Robbins and Siegmund [RS71]. Moreover, the asymptotic normality is due to Sacks, see Theorem 1 in [Sac58] and requires the additional assumption that the probability density function  $f$  is differentiable with bounded derivative in every neighbourhood of the quantile  $\theta_\alpha$ . More precisely, if the step-size is given by  $a_n = a_1/n$  where  $a_1 > 0$  and  $2a_1 f(\theta_\alpha) > 1$ , we have the asymptotic normality

$$\sqrt{n}(\theta_n - \theta_\alpha) \implies \mathcal{N}\left(0, \frac{a_1^2 \alpha (1 - \alpha)}{2a_1 f(\theta_\alpha) - 1}\right).$$

One can observe that in the special case where the value  $f(\theta_\alpha) > 0$  is known, it is possible to minimize the previous limiting variance by choosing  $a_1 = 1/f(\theta_\alpha)$  which leads to an asymptotic variance  $\frac{\alpha(1-\alpha)}{f^2(\theta_\alpha)}$ .

Besides the classical choice  $a_n = a_1/n$  where  $a_1 > 0$ , slower step-size  $a_n = a_1/n^a$  where  $a_1 > 0$  and  $1/2 < a < 1$  have been studied in depth. Chung [Chu54] and Fabian [Fab68] obtained that the asymptotic normality still holds, and more precisely, if  $f(\theta_\alpha) > 0$ , they showed that

$$\sqrt{n^a}(\theta_n - \theta_\alpha) \implies \mathcal{N}\left(0, \frac{a_1\alpha(1-\alpha)}{2f(\theta_\alpha)}\right).$$

Further results such as law of iterated logarithm (LIL) or quadratic strong law (QSL) are available for the quantile recursive estimation. These results refine the convergence in the Central Limit Theorem by ensuring almost sure existence of  $\limsup \left(\frac{n}{2\log\log n}\right)^{1/2}(\theta_n - \theta_\alpha)$  for the LIL and of  $\lim \frac{1}{\log n} \sum_{k=1}^n (\theta_k - \theta_\alpha)^2$  for the QSL [GK75, Ker77, LR79, Pel98]. Consistently with a CLT, the limiting value obtained in the above limits equals the limiting variance. Note that the scaling given here correspond to the case of a step sequence  $a_1/n$  and can be adapted for slower step sizes.

Concerning the superquantile estimation, [BFP09] proved, in the special case of the one-time-scale stochastic algorithm where  $a_n = b_n$ , the almost sure convergences

$$\lim_{n \rightarrow \infty} \theta_n = \theta_\alpha \quad \text{and} \quad \lim_{n \rightarrow \infty} \tilde{\vartheta}_n = \vartheta_\alpha \quad \text{a.s.}$$

and established the joint asymptotic normality of the averaged version [PJ92, Rup88] of their one-time-scale stochastic algorithm

$$\bar{\theta}_n = \frac{1}{n} \sum_{k=1}^n \theta_k \quad \text{and} \quad \bar{\vartheta}_n = \frac{1}{n} \sum_{k=1}^n \tilde{\vartheta}_k.$$

as

$$\sqrt{n} \begin{pmatrix} \bar{\theta}_n - \theta_\alpha \\ \bar{\vartheta}_n - \vartheta_\alpha \end{pmatrix} \implies \mathcal{N}(0, \Sigma)$$

where the asymptotic covariance matrix  $\Sigma$  is explicitly calculated.

## 4.2 Different coupled algorithms

In [BCG21] and [CG21], we study different algorithms for the superquantile recursive approximation. Namely we focus on the classical algorithm

$$\begin{cases} \theta_{n+1} = \theta_n - a_n (\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha) \\ \vartheta_{n+1} = \vartheta_n + b_n \left( \frac{X_{n+1}}{1-\alpha} \mathbf{1}_{X_{n+1} > \theta_n} - \vartheta_n \right), \end{cases} \quad (4.1)$$

and on its convexified counterpart

$$\tilde{\vartheta}_{n+1} = \tilde{\vartheta}_n + b_n \left( \theta_n + \frac{(X_{n+1} - \theta_n)}{(1-\alpha)} \mathbf{1}_{X_{n+1} > \theta_n} - \tilde{\vartheta}_n \right). \quad (4.2)$$

We also introduce a third version for which we expect to benefit from the acceleration of the Cesaro averaging procedure  $\bar{\theta}_n$ , to obtain a better recursion on the superquantile estimation.

We thus introduce:

$$\begin{cases} \theta_{n+1} = \theta_n - a_n (\mathbf{1}_{X_{n+1} \leq \theta_n} - \alpha) \\ \bar{\theta}_{n+1} = \frac{1}{n} \sum_{k=1}^{n+1} \theta_k \\ \hat{\vartheta}_{n+1} = \hat{\vartheta}_n + b_n \left( \frac{X_{n+1}}{1-\alpha} \mathbf{1}_{X_{n+1} > \bar{\theta}_n} - \hat{\vartheta}_n \right) \end{cases}. \quad (4.3)$$

In our work, we study these three algorithms with different step-size sequences of the form

$$a_n = a_1 n^{-a}, \quad b_n = b_1 n^{-b}, \quad \text{with} \quad \frac{1}{2} < a, b \leq 1. \quad (4.4)$$

This corresponds to the standard framework of Robbins-Monro algorithms. We emphasize that the case where  $a_n = b_n$  was studied in [BFP09] and that we specifically consider the case of two time scale algorithms, similarly to [MP06], or even three time scales in (4.3).

Our goal is to compare the performance of these different algorithms. Our first approach uses asymptotic variances as a criterion, while our second considers a finite horizon measure given by the mean squared error.

We will assume throughout this section that the density  $f$  of interest is bounded, with  $f(\theta_\alpha) > 0$  and that  $\theta \mapsto (1 + |\theta|)|f'(\theta)|$  is a bounded function. We further assume that  $X$  has a moment of order strictly larger than 2:

$$\exists p > 2 : \quad \int x^p f(x) dx < \infty.$$

### 4.2.1 Asymptotic results

We introduce two key functions in order to better understand the algorithms introduced in Equations (4.1) and (4.2). Let  $h_\alpha(\theta)$  and  $l_\alpha(\theta)$  be the functions defined, for all  $\theta \in \mathbb{R}$ , by:

$$h_\alpha(\theta) = \frac{1}{1-\alpha} \mathbb{E}[X \mathbf{1}_{\{X > \theta\}}] \quad \text{and} \quad l_\alpha(\theta) = \mathbb{E}[L(\theta, X)] = \theta + \frac{1}{1-\alpha} \mathbb{E}[(X - \theta) \mathbf{1}_{\{X > \theta\}}].$$

We shall observe that

$$h_\alpha(\theta_\alpha) = \vartheta_\alpha \quad \text{and} \quad l_\alpha(\theta_\alpha) = \vartheta_\alpha = \min_{\theta \in \mathbb{R}} l_\alpha(\theta). \quad (4.5)$$

We can write the drift-martingale decomposition

$$\begin{aligned} \vartheta_{n+1} &= \vartheta_n + b_n (h_\alpha(\theta_n) - \vartheta_n) + b_n \varepsilon_{n+1}, \\ &= \vartheta_n (1 - b_n) + b_n h_\alpha(\theta_n) + b_n \varepsilon_{n+1}, \end{aligned} \quad (4.6)$$

and

$$\begin{aligned} \hat{\vartheta}_{n+1} &= \hat{\vartheta}_n + b_n (l_\alpha(\theta_n) - \hat{\vartheta}_n) + b_n \xi_{n+1}, \\ &= \hat{\vartheta}_n (1 - b_n) + b_n l_\alpha(\theta_n) + b_n \xi_{n+1}, \end{aligned} \quad (4.7)$$

where  $\varepsilon_n$  and  $\xi_n$  are sequence of martingales increments such that  $\mathbb{E}[\varepsilon_{n+1}^2|\mathcal{F}_n] = \sigma_\alpha^2(\theta_n)$  and  $\mathbb{E}[\xi_{n+1}^2|\mathcal{F}_n] = \tau_\alpha^2(\theta_n)$  where for all  $\theta \in \mathbb{R}$ :

$$\sigma_\alpha^2(\theta) = \frac{1}{(1-\alpha)^2} \text{Var}(X\mathbf{1}_{\{X>\theta\}}) \quad \text{and} \quad \tau_\alpha^2(\theta) = \frac{1}{(1-\alpha)^2} \text{Var}((X-\theta)\mathbf{1}_{\{X>\theta\}}).$$

Using the convergence of  $\theta_n$  to  $\theta_\alpha$ ,  $\sigma_\alpha^2(\theta_\alpha)$  and  $\tau_\alpha^2(\theta_\alpha)$  are good candidates for the limiting variances of our recursive sequences. Note that from their definition

$$\tau_\alpha^2(\theta_\alpha) = \sigma_\alpha^2(\theta_\alpha) - \frac{\alpha\theta_\alpha}{(1-\alpha)}(2\vartheta_\alpha - \theta_\alpha),$$

such that for all  $\theta_\alpha \geq 0$ ,  $\tau_\alpha^2(\theta_\alpha) \leq \sigma_\alpha^2(\theta_\alpha)$ .

**Central limit theorem** There are two classical ways to obtain a central limit theorem, either use a martingale approach [Sac58, MP06], or through a diffusion approximation [BMP90, GPS18]. The limiting behaviour of  $(\theta_n, \vartheta_n)$ , and  $(\theta_n, \tilde{\vartheta}_n)$  derives from Theorem 1 in [MP06] and relies on a CLT for martingales.

**Theorem 4.2.1.** *Assume that the step-size sequences  $(a_n)_{n \geq 0}$  and  $(b_n)_{n \geq 0}$  defined in (4.4) satisfy  $1/2 < a < b \leq 1$  with  $b_1 > 1/2$  if  $b = 1$ . Then, the recursive sequences  $(\vartheta_n)_{n \geq 0}$  and  $(\tilde{\vartheta}_n)_{n \geq 0}$  share the same joint asymptotic normality*

$$\sqrt{n^b}(\vartheta_n - \vartheta_\alpha) \Longrightarrow \mathcal{N}(0, \Gamma_{\vartheta_\alpha}),$$

where the asymptotic variance is given by

$$\Gamma_{\vartheta_\alpha} = \begin{cases} \frac{b_1^2 \tau_\alpha^2(\theta_\alpha)}{2b_1 - 1} & \text{if } b = 1, \\ \frac{b_1 \tau_\alpha^2(\theta_\alpha)}{2} & \text{if } b < 1. \end{cases}$$

Note that in the case where  $1/2 < b < a < 1$ , the CLT still holds but the variance for  $(\vartheta_n)$  increases to  $\frac{b_1^2 \sigma_\alpha^2(\theta_\alpha)}{2b_1 - 1}$  while it does not change for  $(\tilde{\vartheta}_n)$ .

The limiting behaviour of (4.3) is obtained in [CG21] using a diffusion approximation as developed by [BMP90, GPS18]. The main idea behind the proof is to consider the evolution of the rescaled algorithm  $(\sqrt{n}(\bar{\theta}_n - \theta_\alpha), \sqrt{n}(\hat{\vartheta}_n - \vartheta_\alpha))$  and prove that it is closed to a discretization of an Ornstein-Uhlenbeck process. The limiting variance in the CLT then appears as the variance of the invariant law of the diffusion.

**Theorem 4.2.2.** *Under our assumptions*

i) *If  $b \in (1/2, 1)$ , then:*

$$\sqrt{n^b}(\hat{\vartheta}_n - \vartheta_\alpha) \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}\left(0, \frac{b_1 \sigma_\alpha^2(\theta_\alpha)}{2}\right).$$

ii) If  $b = 1$ , and  $b_1 > 1/2$  then:

$$\sqrt{n} \begin{pmatrix} \bar{\theta}_n - \theta_\alpha \\ \hat{\vartheta}_n - \vartheta_\alpha \end{pmatrix} \xrightarrow[n \rightarrow +\infty]{\mathcal{L}} \mathcal{N}(0, S^2),$$

where:

$$S^2 = \begin{pmatrix} \frac{\alpha(1-\alpha)}{f(\theta_\alpha)^2} & \frac{\alpha}{f(\theta_\alpha)}(\vartheta_\alpha - \theta_\alpha) \\ \frac{\alpha}{f(\theta_\alpha)}(\vartheta_\alpha - \theta_\alpha) & \frac{b_1^2}{(2b_1-1)}\sigma_\alpha^2(\theta_\alpha) - \frac{2b_1}{2b_1-1}\frac{\alpha\theta_\alpha(\vartheta_\alpha-\theta_\alpha)}{(1-\alpha)} \end{pmatrix}.$$

In the case where  $b < 1$  and  $\vartheta_\alpha$  and  $\theta_\alpha$  are positive, the asymptotic variance obtained in Theorem 4.2.2 ii) is larger than  $\Gamma_{\vartheta_\alpha}$ .

The discussion in the case where  $b = 1$  is more interesting. For a fixed value of  $b_1$  we observe that

$$S_{22}^2 < \Gamma_{\vartheta_\alpha} \iff b_1 < \frac{2\vartheta_\alpha - 2\theta_\alpha}{2\vartheta_\alpha - \theta_\alpha} = 1 - \frac{\theta_\alpha}{2\vartheta_\alpha - \theta_\alpha}.$$

Therefore if the distribution  $f$  satisfies that  $\frac{\vartheta_\alpha}{\theta_\alpha} > \frac{3}{2}$ , there exists  $b_1$  such that the asymptotic variance of  $\hat{\vartheta}_n$  is smaller than the one of  $\vartheta_n$  and  $\tilde{\vartheta}_n$ .

**LIL and QSL** Further results can be obtained using the recursive drift martingale decomposition and the standard proof scheme of [GK75]. Let me say a few words on the strategy in the case of  $(\tilde{\vartheta}_n)$ . We deduce from (4.7) and the fact that  $l_\alpha(\theta_\alpha) = \vartheta_\alpha$

$$\hat{\vartheta}_{n+1} - \vartheta_\alpha = (\hat{\vartheta}_n - \vartheta_\alpha)(1 - b_n) + b_n(l_\alpha(\theta_n) - l_\alpha(\theta_\alpha)) + b_n\xi_{n+1}.$$

and then by recursion that for all  $n \geq 2$ :

$$\tilde{\vartheta}_{n+1} = \frac{1}{P_n}(\tilde{\vartheta}_1 + N_{n+1} + L_{n+1}), \quad (4.8)$$

where  $P_n = \prod_{k=1}^n (1 - b_k)^{-1}$ , and

$$N_{n+1} = \sum_{k=1}^n P_k b_k \xi_{k+1}, \quad L_{n+1} = \sum_{k=1}^n P_k b_k (l_\alpha(\theta_k) - l_\alpha(\theta_\alpha)). \quad (4.9)$$

Our results derive from a sharp control of  $P_n$  as well as LIL and QSL for martingales [Sto70, Ber04]. The drift terms are then controlled with a Taylor expansion using the convexity of  $l_\alpha$ . A different strategy is needed for  $\vartheta_n$  since the associated function  $h_\alpha$  is not convex and we introduce a modified sequence  $\vartheta_n + \delta_n \theta_n$  (with an appropriate  $\delta_n$  stemming from spectral analysis) in order to come back to a convex case. For precise results and proofs, I refer to the main article [BCG21].

## 4.2.2 Non asymptotic results

In this part we aim at studying the finite horizon performance of the three scales algorithm (4.3). We obtain non asymptotic controls of the mean squared error  $\mathbb{E}[(\theta_n - \theta_\alpha)^2]$ ,  $\mathbb{E}[(\bar{\theta}_n - \theta_\alpha)^2]$  and  $\mathbb{E}[(\hat{\vartheta}_n - \vartheta_\alpha)^2]$  based on adequate linearizations of the algorithm.



**For the quantile recursion** Our results for the quantile estimation sequence  $(\theta_n)$  are based on the article [GP23] in which the Robbins Monro algorithm for quantile is seen as a gradient descent. More precisely we shall remark that:

$$\begin{aligned}\theta_{n+1} &= \theta_n - a_n(F(\theta_n) - F(\theta_\alpha)) + a_n\Delta M_{n+1} \\ &= \theta_n - a_n\Phi'(\theta_n) + a_n\Delta M_{n+1},\end{aligned}$$

where the function  $\Phi$  equals

$$\Phi(\theta) := \int_{\theta_\alpha}^{\theta} \int_{\theta_\alpha}^u f(s) ds du.$$

and the martingale increment is defined by

$$\Delta M_{n+1} := F(\theta_n) - \mathbf{1}_{X_{n+1} \leq \theta_n}.$$

Following the roadmap of [GP23] we define the Lyapunov function

$$V_q(\theta) := \Phi(\theta)^q \exp(\Phi(\theta)), \quad \forall q \in \mathbb{N}. \quad (4.10)$$

and obtain a recursion on  $V_q(\theta_n)$  using a Taylor expansion. We obtain that

**Theorem 4.2.3.** *If  $a_n = a_1 n^{-a}$  with  $a \in (0, 1)$ , then for any integer  $q \geq 1$ , we have*

$$\exists K_q \geq 0 \quad \forall n \in \mathbb{N} \quad \mathbb{E}|\theta_n - \theta_\alpha|^{2q} \leq K_q a_n^q.$$

We then consider the averaged estimator  $\bar{\theta}_n = \frac{1}{n} \sum_{k=1}^n \theta_k$  and write a drift-martingale decomposition

$$\begin{aligned}\bar{\theta}_{n+1} - \theta_\alpha &= \frac{n}{n+1}(\bar{\theta}_n - \theta_\alpha) + \frac{1}{n+1}(\theta_{n+1} - \theta_\alpha) \\ &= \left(1 - \frac{1}{n+1}\right)(\bar{\theta}_n - \theta_\alpha) + \frac{1}{n+1}(\theta_n - \theta_\alpha - a_n(F(\theta_n) - F(\theta_\alpha)) + a_n\Delta M_{n+1}) \\ &= \left(1 - \frac{1}{n+1}\right)(\bar{\theta}_n - \theta_\alpha) + \frac{1}{n+1}((\theta_n - \theta_\alpha)(1 - a_n f(\theta_\alpha)) + a_n\Delta M_{n+1} + a_n R_n).\end{aligned}$$

where the last line derives from the linear approximation of  $F$  around  $\theta_\alpha$

$$F(\theta_n) - F(\theta_\alpha) = f(\theta_\alpha)(\theta_n - \theta_\alpha) + \int_{\theta_\alpha}^{\theta_n} [f(u) - f(\theta_\alpha)] du.$$

As a consequence we see that we have to study the two dimensional vector  $Z_n = (\theta_n - \theta_\alpha, \bar{\theta}_n - \theta_\alpha)^T$

$$Z_{n+1} = A_n Z_n + a_n \begin{pmatrix} \Delta M_{n+1} \\ \frac{1}{n+1} \Delta M_{n+1} \end{pmatrix} + a_n \begin{pmatrix} R_n \\ \frac{1}{n+1} R_n \end{pmatrix}, \quad (4.11)$$

where  $A_n$  translates the linearization of the algorithm around  $(\theta_\alpha, \theta_\alpha)$  at step  $n$ :

$$A_n = \begin{pmatrix} 1 - a_n f(\theta_\alpha) & 0 \\ \frac{1 - a_n f(\theta_\alpha)}{n+1} & 1 - \frac{1}{n+1} \end{pmatrix}.$$

and  $R_n$  is a rest term. We then consider a change of variable to diagonalize matrix  $A_n$  and use the contraction produced by the eigenvalue  $(1 - \frac{1}{n+1})$  to obtain the following result.

**Theorem 4.2.4.** Consider a step sequence  $a_n = a_1 n^{-a}$  with  $a \in (1/2, 1)$ , then, there exists a positive constant  $\Gamma_a$  such that

$$\mathbb{E}[(\bar{\theta}_n - \theta_\alpha)^2] \leq \frac{\alpha(1-\alpha)}{f(\theta_\alpha)^2 n} + \kappa' \frac{\alpha(1-\alpha)}{f(\theta_\alpha)^3} n^{-(2-a)\wedge(\frac{1}{2}+a)} + \Gamma_a n^{-(3-a)\wedge(\frac{3}{2}+a)} \quad \forall n \geq 1,$$

where  $\kappa'$  is an explicit constant.

Note that the optimal choice of  $a$  corresponds to  $a = 3/4$  and in this case

$$\mathbb{E}[(\bar{\theta}_n - \theta_\alpha)^2] \leq \frac{\alpha(1-\alpha)}{f(\theta_\alpha)^2 n} + \frac{8\kappa\alpha(1-\alpha)}{7f(\theta_\alpha)^3} n^{-5/4} + \Gamma_{3/4} n^{-9/4}, \quad \forall n \geq 1.$$

Our computations precise the computations of the second order term with respect to Corollary 6 in [GP23]. We observe that in our case, if the second order term involved the local curvature given by the Cramer-Rao lower bound  $f(\theta_\alpha)^{-2}$ , this curvature is also involved in the second order term with a larger power (3 instead of 2), that is of course compensated by  $n^{-5/4}$ .

**For the superquantile** A similar strategy for the superquantile recursion (4.6) leads to

$$\vartheta_{n+1} = \vartheta_n(1 - b_n) + \frac{b_n}{1 - \alpha} \theta_\alpha f(\theta_\alpha) (\bar{\theta}_n - \theta_\alpha) + b_n \varepsilon_{n+1} + b_n \hat{R}_n$$

Therefore, we can write a three dimensional linearization

$$\begin{pmatrix} \theta_{n+1} - \theta_\alpha \\ \bar{\theta}_{n+1} - \theta_\alpha \\ \hat{\vartheta}_{n+1} - \vartheta_\alpha \end{pmatrix} = B_n \begin{pmatrix} \theta_n - \theta_\alpha \\ \bar{\theta}_n - \theta_\alpha \\ \hat{\vartheta}_n - \vartheta_\alpha \end{pmatrix} + \begin{pmatrix} a_n \Delta M_{n+1} \\ \frac{a_n}{n+1} \Delta M_{n+1} \\ b_n \varepsilon_{n+1} \end{pmatrix} + \begin{pmatrix} a_n R_n \\ \frac{a_n}{n+1} R_n \\ b_n \hat{R}_n \end{pmatrix}$$

where

$$B_n := \begin{pmatrix} 1 - a_n f(\theta_\alpha) & 0 & 0 \\ \frac{1 - a_n f(\theta_\alpha)}{n+1} & 1 - \frac{1}{n+1} & 0 \\ 0 & -\frac{\theta_\alpha f(\theta_\alpha)}{1-\alpha} b_n & 1 - b_n \end{pmatrix}. \quad (4.12)$$

Here an additional difficulty arise if we want to consider step sequences  $b_n$  close to  $1/(n+1)$ , since in that case, the matrix  $B_n$  cannot be diagonalized. However performing an approximated change of basis, we obtain the following non asymptotic bounds.

**Theorem 4.2.5.** Assume that the step sequence satisfy (4.4), then

i) If  $b \in (1/2, 1)$ , then a large enough constant  $\Gamma$  exists such that

$$\forall n \geq 1, \quad \mathbb{E}[(\hat{\vartheta}_n - \vartheta_\alpha)^2] \leq \frac{\sigma_\alpha^2(\theta_\alpha)}{2} b_n + \Gamma n^{-\frac{b+1}{2}}.$$

ii) If  $b = 1$  and  $b_1 > (\frac{1+a}{2}) \wedge (\frac{5}{2} - a)$ , then a large enough constant  $\Gamma$  exists such that:

$$\forall n \geq 1, \quad \mathbb{E}[(\hat{\vartheta}_n - \vartheta_\alpha)^2] \leq \frac{C_{\alpha, b_1}}{n} + \Gamma n^{-(1+\frac{a}{2})\wedge(2-a)},$$

where

$$C_{\alpha, b_1} = \frac{4b_1^2 \alpha(1-\alpha)}{(2b_1 - 1)^2 f(\theta_\alpha)^2} \left[ 1 + \sqrt{1 + \frac{\sigma_\alpha^2(\theta_\alpha) f(\theta_\alpha)^2 (2b_1 - 1)}{4\alpha(1-\alpha)}} \right]^2.$$

Note that our non-asymptotic results achieve the optimal rate of convergence, and that except for *ii*) in Theorem 4.2.5 the first order constant matches the asymptotic variance obtained in Theorem 4.2.2. The degraded constant in that last case is due a technical issue in diagonalizing the matrix  $B_n$ , it might be improved through another method.

### 4.3 CV@R penalized portfolio optimization with biased stochastic mirror descent

I present here a work in collaboration with Lorick Huang and Sébastien Gadat in which we consider an application of these algorithms in a financial setting. The main question is: given a portfolio of financial assets, what is the best way to allocate resources in order to maximize profits while complying with risk management constraint? Without any constraint, it is likely that the asset with the highest expected return would be favored, leading to a trivial optimization problem. However, ethical issues arise when this latter return has a high variability, since even though some gains are expected in the long run, large losses may occur in between the gains. Consequently, banking regulations have been introduced, involving different metrics to measure risk, and policies have been developed that use one metric or another, leading to different investment optimization problems and strategies.

We focus on portfolio optimization under a conditional value-at-risk (CV@R) constraint (which corresponds to the superquantile studied above). Such a question has already been studied in previous works, notably by Rockafellar and Uryasev [RU00], who develop a Monte Carlo strategy to approximate the CV@R of the portfolio (see also [KPU01, RU02, RR14]). In these works, the authors use the approximation of V@R and CV@R with an empirical batch sum to feed an optimization algorithm via classical tools of convex analysis. In our work we aim to infer the CV@R and the optimization strategy simultaneously. Furthermore, we consider a setting where the observations used to feed the algorithm can be corrupted either from observation or from a numerical strategy, inducing some bias that will impact the estimation procedure.

#### 4.3.1 The optimization problem

We consider the relative return of a portfolio with  $m$  assets  $(A_1(t), \dots, A_m(t))_{t \geq 0}$ , defined by the random vector  $Z = (Z_1, \dots, Z_m)$ , where  $Z_i = \frac{A_i(T)}{A_i(0)} - 1$  is a relative return at fixed horizon  $T$  of the asset  $A_i$ . Note that here  $T$  is supposed to be fixed and will disappear from notations.

An investment strategy corresponds to an allocation of the initial capital  $u = (u_1, \dots, u_m)$  that belongs to the  $m - 1$  dimensional simplex denoted by  $\Delta_m$  and defined as:

$$\Delta_m := \left\{ u \in \mathbb{R}_+^m : \sum_{i=1}^m u_i = 1 \right\}.$$

We are interested in a constrained optimization of the mean return of defined as the mean of the portfolio modeled by the random variable  $\langle Z, u \rangle$  defined by:

$$\langle Z, u \rangle = \sum_{i=1}^m u_i Z_i.$$

The risk measure depends on a level  $\alpha$  for which the  $\alpha$ -quantile is negative. To underline this change with respect to the previous section we use the notation  $V@R_\alpha(\alpha)$  for the  $\alpha$  quantile.  $CV@R_\alpha$  (noted  $\vartheta_\alpha$  above) is the mean value of the *loss* when  $\langle Z, u \rangle$  is below  $V@R_\alpha$ , namely:

$$CV@R_\alpha(u) = \mathbb{E}[-\langle Z, u \rangle \mid \langle Z, u \rangle \leq V@R_\alpha(u)].$$

With our convention this risk measure is positive.

Then, for any fixed positive level  $M$ , the optimization problem of  $u \mapsto \mathbb{E}[\langle Z, u \rangle]$  with  $\text{CV@R}_\alpha$  constraints we are interested in writes

$$\mathcal{P}_M = \arg \min_{u \in \Delta_m} \left\{ - \sum_{i=1}^m u_i \mathbb{E}[Z_i] : \text{CV@R}_\alpha(u) \leq M \right\}.$$

In our work we consider an equivalent unconstrained penalized optimization problem deriving from a Lagrangian formulation. For any  $\lambda > 0$ , we search for the solution  $u_\lambda^*$  of the following (convex) optimization problem:

$$\mathcal{Q}_\lambda := \arg \min_{u \in \Delta_m} \left\{ - \sum_{i=1}^m u_i \mathbb{P}[Z_i] + \lambda \text{CV@R}_\alpha(u) \right\}.$$

We will use, as presented above, the convex representation of  $\text{CV@R}_\alpha$  introduced by [RU00, KPU01] and write similarly as in (4.5) that

$$\text{CV@R}_\alpha(u) = \min_{\theta \in \mathbb{R}} \psi_\alpha(u, \theta),$$

where  $\psi$  is the convex coercive Lipschitz continuous and differentiable function defined by

$$\psi_\alpha(u, \theta) = \theta + \frac{1}{1-\alpha} \mathbb{E}[\langle Z, u \rangle - \theta]_+, \tag{4.13}$$

where  $[x]_+ = \max(0, x)$ .

As a consequence, we can rewrite the optimization problem as searching for  $x_\lambda^* = (u_\lambda^*, \theta_\lambda^*)$  of

$$\mathcal{Q}_\lambda = \arg \min_{(u, \theta) \in \Delta_m \times \mathbb{R}} \{p_\lambda(u, \theta)\}, \tag{4.14}$$

where the key function  $p_\lambda$  is defined by:

$$p_\lambda(u, \theta) = - \sum_{i=1}^m u_i \mathbb{E}[Z_i] + \lambda \psi_\alpha(\theta, u). \tag{4.15}$$

Note that by construction  $\theta_\lambda^* = \text{CV@R}_\alpha(u_\lambda^*)$ .

To solve the minimization problem (4.14) we are led to use stochastic approximation theory, since the (convex) function  $p_\lambda$  is written as an expectation. However, there are specific difficulties. First, we have to deal with minimization over the simplex  $\Delta_m$ . Second, the random variables  $Z$  involved in  $p_\lambda$  cannot generally be simulated exactly, so it will be necessary to control the bias coming from the stochastic simulation as well as its impact on the optimization strategy. Finally, note that although  $p_\lambda$  is differentiable, it is the expectation of a non-differentiable function of  $(u, \theta)$  that will require special attention in the following algorithms.

### 4.3.2 The biased stochastic gradient descent

To handle the minimization on the simplex  $\Delta_m$  we chose to use the Mirror descent (MD) introduced by Nemirovski [NY83]. It allows to handle constrained optimization problems especially without adding some supplementary projection step by “pushing” the frontiers of the simplex at an infinite distance from any point strictly inside  $\Delta_m$ .

**Deterministic mirror descent** To explain the main idea, let me come back to a deterministic setting. Recall that when we consider the question of minimizing a convex smooth  $f$ , the deterministic gradient descent writes :

$$x_{k+1} = x_k - \eta_k \nabla f(x_k),$$

it is equivalent to the *proximal problem* that consists in minimizing at each step the first order Taylor expansion of the function penalized by the  $L^2$  norm :

$$x_{k+1} = \arg \min_x \left\{ \langle x, \nabla f(x_k) \rangle + \frac{1}{2\eta_k} \|x - x_k\|^2 \right\}.$$

The mirror descent considers a different penalization using a Bregman divergence function  $D$  [BT03] by

$$x_{k+1} = \arg \min_x \left\{ \langle x, \nabla f(x_k) \rangle + \frac{1}{\eta_k} D(x, x_k) \right\}.$$

The choice of divergence used depends on the geometry of the problem and must allow the minimization problem to be solved explicitly so that it can be solved algorithmically without too much difficulty. In our setting a natural choice is to consider the Bregman divergence associated to the strongly convex negative entropy on  $\Delta_m$  and the  $L^2$  norm on  $\mathbb{R}$  namely

$$D_\Phi(u, v) = \Phi(u) - \Phi(v) - \langle \nabla \Phi(v), u - v \rangle.$$

where

$$\Phi(u, \theta) = \sum_{i=1}^m u_i \log(u_i) + \frac{\theta^2}{2}.$$

Coming back to our optimization problem (4.14) and denote by  $x = (u, \theta)$ , then the deterministic mirror descent writes

$$x_{k+1} = \arg \min_{x \in \Delta_m \times \mathbb{R}} \left\{ \langle \nabla p_\lambda(x_k), x \rangle + \frac{1}{\eta_{k+1}} D_\Phi(x, x_k) \right\}.$$

that can be solved explicitly as

$$x_{k+1} = \begin{pmatrix} u^{k+1} \\ \theta^{k+1} \end{pmatrix} \quad \text{with} \quad \begin{cases} u^{k+1} = \frac{u^k e^{-\eta_{k+1} \partial_u p_\lambda(u^k, \theta^k)}}{\|u^k e^{-\eta_{k+1} \partial_u p_\lambda(u^k, \theta^k)}\|_1} \\ \theta^{k+1} = \theta^k - \eta_{k+1} \partial_\theta p_\lambda(u^k, \theta^k) \end{cases}, \quad (4.16)$$

where the first equation has to be understood within a  $m$  dimensional vector structure.

Following the work of [BT03, NY83], it can be shown (see Proposition 1 in [LNS12]) that an averaged version of the sequence  $(x_k)_{k \geq 1}$  defined by

$$\tilde{x}_n = \left( \sum_{k=0}^n \eta_k \right)^{-1} \sum_{k=0}^n \eta_k x_k,$$

satisfies the next error bound

$$p_\lambda(\tilde{x}_n) - p_\lambda(x_\lambda^*) \leq \frac{\{\Delta_\Phi^0\}^2 + L^2 \sum_{k=1}^n \eta_k^2}{2 \sum_{k=1}^n \eta_k},$$

where  $\{\Delta_\Phi^0\}^2 = \frac{(\theta_0 - V @ R_\alpha(u_\lambda^*))^2}{2} + \log m$  and  $L = \arg \max_{x \in \Delta_m \times \mathbb{R}} \{\|\nabla p_\lambda(x)\|\}$ .

**Our stochastic biased mirror descent** We denote by  $\partial_u$  and  $\partial_\theta$  the partial derivatives with respect to  $u$  and  $\theta$ , and using that  $p_\lambda$  is the expectation of a convex function, we then verify that:

$$\partial_u p_\lambda(u, \theta) = -\mathbb{E}[Z] + \frac{\lambda \mathbb{E}[Z \mathbf{1}_{\langle Z, u \rangle \geq \theta}]}{1 - \alpha}, \quad (4.17)$$

and

$$\partial_\theta p_\lambda(u, \theta) = \lambda \left[ 1 - \frac{1}{1 - \alpha} \mathbb{E}[\mathbf{1}_{\langle Z, u \rangle \geq \theta}] \right]. \quad (4.18)$$

We assume that we observe a sequence of mutually independent random variables  $(\hat{Z}^k)_{k \geq 0}$  that are also sampled independently from the previous positions of the algorithm. The expressions (4.17) and (4.18) lead to a natural (possibly biased) stochastic approximation of  $\nabla p_\lambda$  with the help of the sequence  $(\hat{Z}^k)_{k \geq 0}$ . Assuming that the algorithm is at step  $k$  at position  $(U_k, \theta_k)$ , we introduce the stochastic approximation of the sub-gradients:

$$\begin{cases} \hat{g}_{k+1,1} &= -\hat{Z}^{k+1} + \frac{\lambda}{1-\alpha} \hat{Z}^{k+1} \mathbf{1}_{\langle \hat{Z}^{k+1}, U_k \rangle \geq \theta_k} \\ \hat{g}_{k+1,2} &= \lambda \left[ 1 - \frac{1}{1-\alpha} \mathbf{1}_{\langle \hat{Z}^{k+1}, U_k \rangle \geq \theta_k} \right] \end{cases}. \quad (4.19)$$

The method we propose is then defined in Algorithm 1.

**Data:** Step-size sequence  $(\eta_k)_{k \in \mathbb{N}}$  and  $U_0 \in \mathbb{R}$ ,  $\theta_0 \in \mathbb{R}$ ;  $\alpha \in (0, 1)$

**Result:** Two sequences:  $X_k = (U_k, \theta_k)_{k \geq 0}$

**for**  $k = 0, \dots$  **do**

Simulate the random  $Z^{k+1}$  satisfying (4.20) and (4.21);

Compute a stochastic approximation  $\hat{g}_{k+1}$  of  $\nabla p_\lambda(U_k, \theta_k)$  with

$$\begin{cases} \hat{g}_{k+1,1} &= -\hat{Z}^{k+1} + \frac{\lambda}{1-\alpha} \hat{Z}^{k+1} \mathbf{1}_{\langle \hat{Z}^{k+1}, U_k \rangle \geq \theta_k} \\ \hat{g}_{k+1,2} &= \lambda \left[ 1 - \frac{1}{1-\alpha} \mathbf{1}_{\langle \hat{Z}^{k+1}, U_k \rangle \geq \theta_k} \right] \end{cases}.$$

Update the algorithm

$$X_{k+1} = \arg \min_{x \in x \in \Delta_m \times \mathbb{R}} \left\{ \langle \hat{g}_{k+1}, x - X_k \rangle + \frac{1}{\eta_{k+1}} \mathcal{D}_\Phi(x, X_k) \right\} \text{ using (4.16)}$$

$$X_{k+1} = (U_{k+1}, \theta_{k+1}), \quad \begin{cases} U^{k+1} &= \frac{U^k e^{-\eta_{k+1} \hat{g}_{k+1,1}}}{\|U^k e^{-\eta_{k+1} \hat{g}_{k+1,1}}\|_1} \\ \theta^{k+1} &= \theta^k - \eta_{k+1} \hat{g}_{k+1,2} \end{cases}.$$

**Algorithm 1:** Biased SMD

**Assumptions and convergence** We now describe the necessary assumptions on the sequence  $(\hat{Z}^k)_{k \geq 0}$  to build a consistent SMD algorithm. We introduce two sequences  $(\delta_{k+1})_{k \geq 0}$  and  $(v_{k+1})_{k \geq 0}$  that translate the fact that we may observe some biased realizations of the assets  $Z$  at step  $k$ , the perfect simulation framework being translated by  $\delta_{k+1} = v_{k+1} = 0$  for all integer  $k$ .

**Assumption (H<sub>1</sub>) Assumptions on the biased simulations**

Let  $\mathcal{L}(\hat{Z}^{k+1})$  and  $\mathcal{L}(Z)$  denote the distributions of  $\hat{Z}^{k+1}$  and  $Z$  respectively. We assume that the sequence  $(\hat{Z}^k)_{k \geq 0}$  satisfies both:

$$\mathcal{W}_1(\mathcal{L}(\hat{Z}^{k+1}), \mathcal{L}(Z)) \leq \delta_{k+1}, \quad (4.20)$$

and

$$\forall u \in \mathbf{\Delta}_m, \quad \forall \theta \in \mathbb{R}, \quad \left\| \mathbb{E} \left[ \langle Z, u \rangle \mathbf{1}_{\langle Z, u \rangle \geq \theta} - \langle \hat{Z}^{k+1}, u \rangle \mathbf{1}_{\langle \hat{Z}^{k+1}, u \rangle \geq \theta} \mid \mathcal{F}_k \right] \right\| \leq v_{k+1}. \quad (4.21)$$

where  $\mathcal{W}_1$  stands for the Wasserstein-1 distance.

We emphasize that both  $(\delta_{k+1})_{k \geq 0}$  and  $(v_{k+1})_{k \geq 0}$  might heavily depend on  $m$  the dimension of the vector  $Z$ . A specific example will be detailed below.

The next result states the asymptotic almost sure convergence of the sequence  $(X_k)_{k \geq 0}$  constructed in Algorithm 1.

**Theorem 4.3.1** (Almost sure convergence of the biased SMD). *Assume that  $\sum_{k \geq 0} \eta_{k+1} = +\infty$  and  $\sum_{k \geq 0} \eta_{k+1}^2 < +\infty$ , and that the bias sequences  $(\delta_{k+1})_{k \geq 0}$  and  $(v_{k+1})_{k \geq 0}$  satisfy*

$$\sum_{k \geq 0} \eta_{k+1} (\sqrt{\delta_{k+1}} + v_{k+1}) < +\infty,$$

then the Cesaro average  $\bar{X}_k^\eta = \left( \sum_{i=0}^k \eta_i \right)^{-1} \left( \sum_{i=0}^k \eta_i X_i \right)$  is almost surely convergent and

$$p_\lambda(\bar{X}_k^\eta) \longrightarrow \min(p_\lambda) \quad a.s.$$

The proof of this theorem follows the standard scheme developed by [NY83] that relies on standard results that are valid for any Bregman divergence  $\mathcal{D}_\Phi$  and on Robbins- Sigmund theorem.

Our work also provides conditions for the convergence of the non averaged sequence  $X_k$ . In that case, we follow the roadmap of [Ben99] that introduces the asymptotic pseudo-trajectories for stochastic algorithms, and [MS18] that adapts these tools to the mirror descent.

**Non asymptotic results** Obtaining non asymptotic guaranties on the convergence speed of the algorithm requires to control the sequence of  $\mathbb{E}(\mathcal{D}_\Phi(x^*, X_k))$  for which we obtain a recursion inequality of the form

$$\mathbb{E}(\mathcal{D}_\Phi(x^*, X_{k+1})) \leq \mathbb{E}(\mathcal{D}_\Phi(x^*, X_k))(1 + a_{k+1}) + b_{k+1},$$

where  $a_{k+1} = 2\eta_{k+1} \left( 2\sqrt{\delta_{k+1}} + \delta_{k+1} + \frac{\lambda v_{k+1}}{1-\alpha} \right)$  and  $b_{k+1} = C \left( \eta_{k+1}^2 + a_{k+1}/2 \right)$ .

For sake of conciseness, I will only state here a result in finite time horizon. Namely, let us consider the specific case of the SMD with a constant step-size sequence stopped at iteration  $n$ , that is

$$\begin{aligned} \eta_{k+1} &= \eta > 0, \quad \forall 0 \leq k \leq n. \\ \eta_{k+1} &= 0, \quad \forall k > n. \end{aligned}$$

We will also assume a constant upper bound of the bias in the simulation of the random variables  $\hat{Z}^k$ : we denote by  $\omega$  its resulting impact in the SMD. More precisely we consider fixed values of  $\delta_{k+1}$  and  $v_{k+1}$  such that:

$$2\sqrt{\delta_{k+1}} + \delta_{k+1} + \frac{\lambda v_{k+1}}{1 - \alpha} = \omega > 0, \quad \forall 1 \leq k \leq n.$$

This is legitimate since we can reduce this bias with the use of an arbitrarily small discretization step-size, which of course harms the computational cost.

**Proposition 4.3.2.** *For a given  $n \in \mathbb{N}$ , if  $(\eta, \omega)$  are chosen such that  $\eta = \frac{\Delta_\Phi^0}{2\sqrt{n+1}}$  and  $\omega = \frac{1}{\sqrt{n+1}\Delta_\Phi^0}$  with  $\{\Delta_\Phi^0\}^2 = \frac{(\theta_0 - V@R_\alpha(u_\lambda^*))^2}{2} + \log m$ , then there exists  $C > 0$  large enough such that:*

$$\mathbb{E} \left[ p_\lambda \left( \bar{X}_n^\eta \right) \right] - p_\lambda(x_\lambda^*) \leq C \frac{|\theta_0 - V@R_\alpha(u_\lambda^*)| + \sqrt{\log m}}{\sqrt{n+1}}.$$

Therefore, these values may be seen as purely theoretical as we do not exactly know the value of  $\Delta_\Phi^0$ . Note that the choice:  $\eta = 1/2\sqrt{n+1}$  and  $\omega = 1/\sqrt{n+1}$  yields:

$$\mathbb{E} \left[ p_\lambda \left( \bar{X}_n^\eta \right) \right] - p_\lambda(x_\lambda^*) \leq C \frac{|\theta_0 - V@R_\alpha(u_\lambda^*)|^2 + \log m}{\sqrt{n+1}}.$$

This degrades the dependency on the size of the portfolio but keeps the appropriate speed in terms of  $n$  the number of iterations.

### 4.3.3 Approximation of a portfolio

**Model for the portfolio dynamics** We consider the situation where  $Z$  contains  $m = m' + 1$  assets: a family  $\mathbf{S} = (S^1, \dots, S^{m'})$  of  $m' = m - 1$  geometric Brownian motions that encode some risky assets in the portfolio. We refer to Pitman and Yor [Pit82] and to Gulisashvili and Stein [GS10] for several details on these classical processes used (among others) for portfolio modelling.

We assume in addition that one of the asset  $Y$  is risk-less, which allows to model debt obligations, or treasury bonds. In mathematical terms, we consider a stochastic rate  $r$ , modeled by a Cox-Ingersoll-Ross (shortened as CIR below) [CIR85] process and we use this rate for the growth ratio of  $Y$ . Recall that the CIR follows a diffusion process given by:

$$dr_t = a(b - r_t)dt + \sigma_0 \sqrt{r_t} dB_0(t), \tag{4.22}$$

where  $(B_0(t))_{t \geq 0}$  stands for a standard real Brownian motion. The parameter  $b$  stands for the long-time mean of the short rate while  $a$  quantifies the strength of the mean-reversion effect. The volatility  $\sigma_0$  is multiplied by  $\sqrt{r_t}$ . We assume that the CIR parameters satisfy

$$ab > \sigma_0^2 \quad \text{and} \quad a > 2\sqrt{2}\sigma_0,$$

ensuring the positivity of  $r_t$ , a control of the  $L^2$  moment of the weak error rate as well as exponential integrability of the integral of the CIR for all time  $t$ . We refer to [Gla03] for further details.



The assets  $Z_t = (Y_t, S_t^1, \dots, S_t^{m'})$  are then described by the following system of stochastic differential equations:

$$\forall t \geq 0 \quad \begin{cases} dY_t &= r_t Y_t dt, \\ dS_t^i &= \mu_i S_t^i dt + \sigma_i S_t^i dB_i(t), \quad \forall i \in \{1, \dots, m'\}, \end{cases} \quad (4.23)$$

where  $\mathbf{B} = (B_0, B_1, \dots, B_{m'})$  refers to a multivariate Brownian motion with correlated components. For the sake of simplicity, these components are assumed to satisfy:

$$\mathbb{E}[B_i(t)B_j(t)] = \rho_{i,j}t.$$

The correlation matrix is the symmetric definite positive matrix, denoted by  $\Sigma = (\rho_{i,j})_{1 \leq (i,j) \leq m}$ .

**Simulation** Let us now present our discretization scheme on a discrete grid  $(kh)_{\{1 \leq k \leq N\}}$  of  $[0, T]$ . The step  $h$  should be chosen in order to satisfy the requirement on the biased simulations (4.20)-(4.21).

Geometric Brownian motion can be exactly simulated from independent Gaussian random variables with variances  $\sqrt{T}$   $W(T) = (W_0(T), W_1(T), \dots, W_{m'}(T))$  Using the Cholewsky decomposition for the matrix  $\Sigma$  as  $LL^T = \Sigma$  we then perform  $B(T) = LW(T)$ . Finally for  $i \in \{1, \dots, m'\}$  we set

$$S_T^i = S_0^i \exp \left( \left( \mu_i T - \frac{(\sigma_i)^2}{2} \right) + \sigma_i B^i(T) \right).$$

The main difficulty comes from discretizing the CIR process, as the coefficients in the SDE are not uniformly elliptic and bounded, as assumed in the seminal works of Bally and Talay [BT96]. Besides, a classical explicit Euler scheme generates positivity issues (because of the square root). However, many authors, notably Alfonsi [Alf05, Alf10] proposed implicit Euler schemes and provided weak and strong error rates in the previously mentioned works. The drift-implicit Euler scheme on a discrete time grid  $(kh)_{0 \leq k \leq N}$  can be written by considering the SDE satisfied by  $y_t = \sqrt{r_t}$  which leads to

$$\hat{y}_{(k+1)h} = \hat{y}_{kh} + \left( \frac{4ab - \sigma_0^2}{8\hat{y}_{(k+1)h}} - \frac{a}{2}\hat{y}_{(k+1)h} \right) h + \frac{\sigma_0}{2} \Delta B_0^{(k)},$$

where  $\Delta B_0^{(k)} = B_0((k+1)h) - B_0(kh)$ . This implicit scheme can be solved explicitly on  $\mathbb{R}_+$  from iteration  $k$  to iteration  $k+1$ , which ensures the positivity of the scheme  $(\hat{r}_{kh})_{0 \leq k \leq N}$ . In particular, the update from  $kh$  to  $(k+1)h$  is given by:

$$\hat{r}_{(k+1)h} = \hat{y}_{(k+1)h}^2 = \left( \frac{\sqrt{\hat{r}_{kh}} + \frac{\sigma_0}{2} \Delta B_0^{(k)}}{2(1 + \frac{ah}{2})} + \sqrt{\frac{\left( \sqrt{\hat{r}_{kh}} + \frac{\sigma_0}{2} \Delta B_0^{(k)} \right)^2}{4(1 + \frac{ah}{2})^2} + \frac{(4ab - \sigma_0^2)h}{8(1 + \frac{ah}{2})}} \right)^2. \quad (4.24)$$

We then use the integral representation  $Y_t = Y_0 \exp \left( \int_0^t r_s ds \right)$  and propose to approximate  $Y_t$  with a Riemann integral approximation between 0 and  $t$ :

$$\hat{I}_h := \frac{1}{N} \sum_{k=1}^N \hat{r}_{kh}.$$

This leads to the definition of our approximation:

$$\hat{Y}_t^{(h)} := Y_0 \exp(\hat{I}_h) = Y_0 \exp\left(\frac{1}{N} \sum_{k=1}^N \hat{r}_{kh}\right). \quad (4.25)$$

**Control of the discretization** Using results from [Alf13] proved in [DNS12] providing an upper bound of the weak error associated to the implicit Euler scheme(4.24) we obtain the control of (4.20) and (4.21) as a function of  $h$ .

**Proposition 4.3.3.** *Using the above discretization scheme presented above,*

- *a constant  $C$  exists (dependent on the CIR parameters) such that:*

$$\mathcal{W}_1(\mathcal{L}(\hat{Z}_1), \mathcal{L}(Z_1)) = \mathcal{W}_1(\mathcal{L}(\hat{Y}_1^{(h)}), \mathcal{L}(Y_1)) \leq C\sqrt{h}.$$

- *for any  $\epsilon > 0$ , there exists a constant  $K_\epsilon$  independent of  $h$  and  $m$  such that:*

$$\left\| \mathbb{E} \left[ \langle Z_1, w \rangle \mathbf{1}_{\langle Z_1, w \rangle \geq \theta} - \langle \hat{Z}_1, w \rangle \mathbf{1}_{\langle \hat{Z}_1, w \rangle \geq \theta} \right] \right\|_2 \leq K_\epsilon \sqrt{m} e^{\frac{\{\sigma^+\}^2 m^2}{4\epsilon^2}} h^{\frac{1}{6} - \epsilon}.$$

We finally aggregate the optimization procedure described in Algorithm 1 with our sampling scheme and address the problem of adapting the step size of the discretization scheme to the step size of the SMD approximation. We focus on the case of constant step at finite time horizon  $n$  as in Proposition 4.3.2. Recall that we assume a constant step-size sequence  $\eta$  and discretization step-size  $h$  which impacts the choice of  $\omega$ . Corollary 4.3.2 combined with Propositions 4.3.3 induce that  $h$  should be chosen as:

$$h^{1/4} + h^{\frac{1}{6} - \epsilon} \sim n^{-1/2},$$

which entails that we could choose a discretization step-size close to  $n^{-3}$ .

## 4.4 Perspectives

### 4.4.1 Deviations inequalities for the quantile and superquantile

A first natural question would be to obtain deviation inequalities for the Robbins-Monro algorithms (4.1) on the form

$$\mathbb{P}(|\theta_n - \theta_\alpha| > \epsilon_n) \leq e^{-\tau_n},$$

and similarly for  $\vartheta_n$ . One idea would be to follow the proof scheme developed by [Woo72], who used characteristic and moment generating functions to derive large deviations for stochastic gradient descent. A major problem here arises from the fact that the gradient descent for the quantile is convex but not strongly convex, so one may need to use the same Lyapunov function as for non-asymptotic bounds (4.10). A first step in this direction is the recent work of [CKW23], who obtained deviation inequalities for the averaged algorithm  $\bar{\theta}_n$ , but the rates obtained do not seem optimal and to my knowledge there are no results for the superquantiles.

A second interesting research direction arises from a discussion with Sebastien Gadat, Xavier Gendre and Thierry Klein on the application of stochastic mirror descent to the joint estimation of indices for sensibility analysis. In the context of sensitivity analysis, where one observes a real output  $Y$  written as

$$Y = g(X_1, \dots, X_p),$$

where  $(X_i)_{1 \leq i \leq p}$  are independent real-valued random variables, various indices such as the Sobol indices [Sob01] or the Cramer von Mises indices [GKL18] have been defined to assess the relative importance of one or a set of input variables. These indices, often associated with an analog of a Hoeffding decomposition of variance, are positive and sum to 1. For example, considering Sobol indices associated with a random variable  $Y$  with finite variance, the Hoeffding decomposition of the variance gives

$$\text{Var}(Y) = \sum_{u \subset \{1, \dots, p\} \setminus \emptyset} V_u, \quad (4.26)$$

with

$$V_u = \sum_{v \subset u \setminus \emptyset} (-1)^{|u|-|v|} \text{Var}(\mathbb{E}[Y|X_i, i \in v]). \quad (4.27)$$

The Sobol indices with respect to  $(X_i)_{i \in u}$  are then defined by

$$S_u = \frac{V_u}{\text{Var}(Y)},$$

where  $\bar{u} = \{1, \dots, p\} \setminus u$ . Hence dividing both side of (4.26) by  $\text{Var}(Y)$  we get

$$1 = \sum_{u \subset \{1, \dots, p\} \setminus \emptyset} S_u. \quad (4.28)$$

This last equation motivates us to construct a stochastic mirror descent on the simplex to estimate simultaneously all the  $2^p - 1$  Sobol indices  $\mathbf{S} = (S_u, u \subset \{1, \dots, p\} \setminus \emptyset)$  at once.

**Some words on our strategy** We will rather focus on closed Sobol indices which are conditioned sobol indices defined as

$$S_u^{cl} = \frac{\text{Var}(\mathbb{E}[Y|X_i, i \in u])}{\text{Var}(Y)}.$$

Closed Sobol indices are linked to Sobol indices through a linear relationship

$$\mathbf{S} = M\mathbf{S}^{cl},$$

where the square matrix  $M$  of size  $q = 2^p - 1$  is triangular with diagonal terms equal to 1. More precisely  $\forall u, v \subset \{1, \dots, p\} \setminus \emptyset$

$$M_{u,v} = (-1)^{|u|-|v|} \mathbf{1}_{v \subset u}.$$

The *Pick-Freeze trick* considers two independent copies  $(X'_1, \dots, X'_p)$  and  $(X_1, \dots, X_p)$  and built for any  $u \subset \{1, \dots, p\} \setminus \emptyset$

$$Y^u = g(X_1^u, \dots, X_p^u),$$

with  $X_i^u = X_i$  if  $i \in u$  and  $X_i^u = X'_i$  if  $i \notin u$ . As a consequence

$$\text{Var}(\mathbb{E}[Y|X_i, i \in u]) = \text{Cov}(Y, Y^u).$$

and thus

$$S_u^{cl} = \frac{\text{Cov}(Y, Y^u)}{\text{Var}(Y)}. \quad (4.29)$$

This allows to define Monte Carlo estimators which are widely studied [Sob01, SRA<sup>+</sup>08, GJK<sup>+</sup>16].

Since the indices writes as an expectation, we aim at considering an alternative method based on Robbins Monro algorithm. For any single index  $S_u^{cl}$ , we can built a function  $\Psi_u(x_u)$  minimal in  $S_u^{cl}$ :

$$\Psi_u(x) = \frac{1}{2} \mathbb{E}[(Y - \mathbb{E}[Y])x_u - (Y^u - \mathbb{E}[Y])]^2.$$

Now the function from  $\mathbb{R}^q$  to  $\mathbb{R}$  defined for  $x = (x_u)_{u \subset \{1, \dots, p\} \setminus \emptyset}$  by

$$\Psi^a(x) = \sum_{u \subset \{1, \dots, p\} \setminus \emptyset} a_u \Psi_u(x_u) \quad (4.30)$$

where  $a = (a_u)_{u \subset \{1, \dots, p\} \setminus \emptyset}$  is a probability distribution with positive weights. The function  $\Psi^a$  is minimized for  $x = \mathbf{S}^{cl}$ , and we can write

$$\mathbf{S}^{cl} = \operatorname{argmin}_{s \in \mathbb{R}^q} \left\{ \Psi^a(s), \sum_{u \in \{1, \dots, p\} \setminus \emptyset} [Ms]_u = 1 \right\}.$$

Our goal is then to study the SMD associated with this minimization problem on the simplex, to give bounds on the convergence rate, and to compare this method with existing methods that estimate the indices one by one. There will be several other questions related to more general problems, such as the optimal choice of weights  $(a_u)$  (which may change with time), or obtaining a central limit theorem for the SMD. These questions are related to the ANR-PRCE-24 GATSBII coordinated by Thierry Klein.



---

## Bibliography

---

- [Alf05] A. Alfonsi. On the discretization schemes for the cir (and bessel squared) processes. Monte Carlo Methods and Applications, 11:355–384, 2005.
- [Alf10] A. Alfonsi. High order discretization schemes for the cir process: application to affine term structure and heston models. Mathematics of Computation, 79(269):209–237, 2010.
- [Alf13] A. Alfonsi. Strong order one convergence of a drift implicit euler scheme: Application to the cir process. Statistics & Probability Letters, 83(2):602–607, 2013.
- [AN72] K. B. Athreya and P. E. Ney. Branching processes. Springer-Verlag Berlin, Mineola, NY, 1972. Reprint of the 1972 original [Springer, New York; MR0373040].
- [Asm03] Soren Asmussen. Applied Probability and Queues, volume 51 of Stochastic Modelling and Applied Probability. Springer, New York, NY, 2003.
- [Bah66] R. R. Bahadur. A Note on Quantiles in Large Samples. The Annals of Mathematical Statistics, 37(3):577–580, June 1966. Publisher: Institute of Mathematical Statistics.
- [Ban08] Vincent Bansaye. Proliferating parasites in dividing cells: Kimmel’s branching model revisited. The Annals of Applied Probability, 18(3):967–996, June 2008. Publisher: Institute of Mathematical Statistics.
- [Ban09] Vincent Bansaye. Cell Contamination and Branching Processes in a Random Environment with Immigration. Advances in Applied Probability, 41(4):1059–1081, 2009. Publisher: Applied Probability Trust.
- [BBGM20] Emmanuel Bacry, Martin Bompain, Stéphane Gaïffas, and Jean-Francois Muzy. Sparse and low-rank multivariate hawkes processes. Journal of Machine Learning Research, 21(50):1–32, 2020.
- [BCF<sup>+</sup>23] Maxime Boutin, Manon Costa, Colin Fontaine, Adrien Perrard, and Violaine Llaurens. Influence of mimicry on extinction risk in Aculeata: a theoretical approach. Peer Community Journal, 3, 2023.
- [BCG21] Bernard Bercu, Manon Costa, and Sébastien Gadat. Stochastic approximation algorithms for superquantiles estimation. Electronic Journal of Probability, 26(none):1–29, January 2021. Publisher: Institute of Mathematical Statistics and Bernoulli Society.
- [BCJ<sup>+</sup>08] Željko Bajzer, Thomas Carr, Krešimir Josić, Stephen J Russell, and David Dingli. Modeling of cancer virotherapy with recombinant measles viruses. Journal of theoretical Biology, 252(1):109–122, 2008.

- [BCS19] Anton Bovier, Loren Coquille, and Charline Smadi. Crossing a fitness valley as a metastable transition in a stochastic population model. The Annals of Applied Probability, 29(6):3541–3589, 2019.
- [BDHM13] Emmanuel Bacry, Sylvain Delattre, Marc Hoffmann, and Jean-François Muzy. Some limit theorems for hawkes processes and application to financial statistics. Stochastic Processes and their Applications, 123(7):2475–2499, 2013.
- [Ben99] M. Benaïm. Dynamics of stochastic approximation algorithms. Séminaire de probabilités de Strasbourg, 33:1–68, 1999.
- [Ber04] B Bercu. On the convergence of moments in the almost sure central limit theorem for martingales with statistical applications. Stochastic Processes and their Applications, 111(1):157–173, May 2004.
- [BFP09] O. Bardou, N. Frikha, and G. Pages. Computing VaR and CVaR using stochastic approximation and adaptive unconstrained importance sampling. Monte Carlo Methods and Applications, 15(3):173–210, 2009. Publisher: Walter de Gruyter GmbH & Co. KG.
- [BHS21] Anna Bonnet, Miguel Martinez Herrera, and Maxime Sangnier. Maximum likelihood estimation for hawkes processes with self-excitation or inhibition. Statistics & Probability Letters, 179:109214, 2021.
- [BKL<sup>+</sup>10] Matt Biesecker, Jung-Han Kimn, Huitian Lu, David Dingli, and Željko Bajzer. Optimization of virotherapy for cancer. Bulletin of mathematical biology, 72(2):469–489, 2010. Publisher: Springer.
- [BLL97] S. Bonhoeffer, M. Lipsitch, and B. R. Levin. Evaluating treatment protocols to prevent antibiotic resistance. Proceedings of the National Academy of Sciences of the United States of America, 94(22):12106–12111, October 1997.
- [BLL04] Carl T. Bergstrom, Monique Lo, and Marc Lipsitch. Ecological theory suggests that antimicrobial cycling will not reduce antimicrobial resistance in hospitals. Proceedings of the National Academy of Sciences, 101(36):13285–13290, September 2004. Publisher: Proceedings of the National Academy of Sciences.
- [BM96] Pierre Brémaud and Laurent Massoulié. Stability of nonlinear hawkes processes. The Annals of Probability, pages 1563–1588, 1996.
- [BM14] Emmanuel Bacry and Jean-François Muzy. Hawkes model for price and trades high-frequency dynamics. Quantitative Finance, 14(7):1147–1166, 2014.
- [BMP90] A. Benveniste, M. Métivier, and P. Priouret. Adaptive Algorithms and Stochastic Approximations. Springer-Verlag, Berlin, New-York, Stochastic Modelling and Applied Probability, 1990.
- [BO05] N. H. Barton and Sarah P. Otto. Evolution of Recombination Due to Random Drift. Genetics, 169(4):2353–2370, April 2005.
- [Bor97] Vivek S. Borkar. Stochastic approximation with two time scales. Systems & Control Letters, 29(5):291–294, February 1997.

- [BPM10] Robert E. Beardmore and Rafael Pena-Miller. Antibiotic cycling versus mixing: the difficulty of using mathematical models to definitively quantify their relative merits. Mathematical biosciences and engineering: MBE, 7(4):923–933, October 2010.
- [BPMGI17] Robert Eric Beardmore, Rafael Peña-Miller, Fabio Gori, and Jonathan Iredell. Antibiotic Cycling and Antibiotic Mixing: Which One Best Mitigates Antibiotic Resistance? Molecular Biology and Evolution, 34(4):802–817, April 2017.
- [BRBH21] Yannick Bessy-Roland, Alexandre Boumezoued, and Caroline Hillairet. Multivariate hawkes process for cyber insurance. Annals of Actuarial Science, 15(1):14–39, 2021.
- [Bru17] Hermann Brunner. Volterra Integral Equations: An Introduction to Theory and Applications. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press, Cambridge, 2017.
- [BS06] Reinhard Bürger and Kristan A Schneider. Intraspecific competitive divergence and convergence under assortative mating. The American Naturalist, 167(2):190–205, 2006. Publisher: The University of Chicago Press.
- [BS17] Sylvain Billiard and Charline Smadi. The interplay of two mutations in a population of varying size: a stochastic eco-evolutionary model for clonal interference. Stochastic Processes and their Applications, 127(3):701–748, 2017.
- [BT96] V. Bally and D. Talay. The law of the Euler scheme for stochastic differential equations. Probability theory and related fields, 104(1):43–60, 1996.
- [BT03] Amir Beck and Marc Teboulle. Mirror descent and nonlinear projected subgradient methods for convex optimization. Operations Research Letters, 31(3):167–175, 2003.
- [BT07] Charles Bordenave and Giovanni Luca Torrisi. Large Deviations of Poisson Cluster Processes. Stochastic Models, 23(4):593–625, November 2007. Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/15326340701645959>.
- [BTT86] A. Ben-Tal and M. Teboulle. Expected Utility, Penalty Functions, and Duality in Stochastic Nonlinear Programming. Management Science, 32(11):1445–1466, 1986.
- [CBGB<sup>+</sup>10] A Chaput-Bardy, A Grégoire, M Baguette, A Pagano, and J Secondi. Condition and phenotype-dependent dispersal in a damselfly, *Calopteryx splendens*. PLoS One, 5(5):e10694, 2010. Publisher: Public Library of Science.
- [CCC22] Patrick Cattiaux, Laetitia Colombani, and Manon Costa. Limit theorems for hawkes processes including inhibition. Stochastic Processes and their Applications, 149:404–426, 2022.
- [CCC23] Patrick Cattiaux, Laetitia Colombani, and Manon Costa. Asymptotic deviation bounds for cumulative processes. Stochastic Processes and their Applications, 163:85–105, September 2023.
- [CCDRB15] Julien Chevallier, María José Cáceres, Marie Doumic, and Patricia Reynaud-Bouret. Microscopic approach of a time elapsed neural model. Mathematical Models and Methods in Applied Sciences, 25(14):2669–2719, 2015.
- [CCF22] Pierre Cordelier, Manon Costa, and Jérôme Fehrenbach. Slow–Fast Model and Therapy Optimization for Oncolytic Treatment of Tumors. Bulletin of Mathematical Biology, 84(6):64, May 2022.



- [CCL<sup>+</sup>21] Camille Coron, Manon Costa, Fabien Laroche, H el ene Leman, and Charline Smadi. Emergence of homogamy in a two-loci stochastic population model. ALEA: Latin American Journal of Probability and Mathematical Statistics, 18:469–508, 2021.
- [CCL<sup>+</sup>23] Camille Coron, Manon Costa, H el ene Leman, Violaine Llaurens, and Charline Smadi. Origin and persistence of polymorphism in loci targeted by disassortative preference: a general model. Journal of Mathematical Biology, 86(1):4, 2023.
- [CCLS18] Camille Coron, Manon Costa, H el ene Leman, and Charline Smadi. A stochastic model for speciation by mating preferences. Journal of mathematical biology, 76:1421–1463, 2018.
- [CEM21] Manon Costa, Christ ele Etchegaray, and Sepideh Mirrahimi. Survival criterion for a population subject to selection and mutations; Application to temporally piecewise constant environments. Nonlinear Analysis: Real World Applications, 59:103239, June 2021.
- [CG21] Manon Costa and S ebastien Gadat. Non asymptotic controls on a recursive superquantile approximation. Electronic Journal of Statistics, 15(2):4718–4769, January 2021. Publisher: Institute of Mathematical Statistics and Bernoulli Society.
- [CGGR19] Manon Costa, S ebastien Gadat, Pauline Gonnord, and Laurent Risser. Cytometry inference through adaptive atomic deconvolution. Journal of Nonparametric Statistics, 31(2):506–547, April 2019. Publisher: Taylor & Francis eprint: <https://doi.org/10.1080/10485252.2019.1599376>.
- [CGH24] Manon Costa, S ebastien Gadat, and Lorick Huang. Cv@ r penalized portfolio optimization with biased stochastic mirror descent. arXiv preprint arXiv:2402.11999, 2024.
- [CGMT20] Manon Costa, Carl Graham, Laurence Marsalle, and Viet-Chi Tran. Renewal in Hawkes processes with self-excitation and inhibition. Advances in Applied Probability, 52(3):879–915, September 2020. Publisher: Applied Probability Trust.
- [Cha06] Nicolas Champagnat. A microscopic interpretation for adaptive dynamics trait substitution sequence models. Stochastic processes and their applications, 116(8):1127–1160, 2006. Publisher: Elsevier.
- [Chi06] Carmen Chicone. Ordinary Differential Equations with Applications. Number 34 in Texts in Applied Mathematics. Springer-Verlag New York, 2006.
- [CHLM16] M. Costa, C. Hauzy, N. Loeuille, and S. M el eard. Stochastic eco-evolutionary model of a prey-predator community. Journal of Mathematical Biology, 72(3):573–622, 2016.
- [Chu54] K. L. Chung. On a Stochastic Approximation Method. The Annals of Mathematical Statistics, 25(3):463–483, September 1954. Publisher: Institute of Mathematical Statistics.
- [CIR85] J.C. Cox, J.E. Ingersoll, and S.A. Ross. A theory of the term structure of interest rates. Econometrica, 53:385–407, 1985.
- [CKS21] Loren Coquille, Anna Kraut, and Charline Smadi. Stochastic individual-based models with power law mutation rate on a general finite trait space. Electronic Journal of Probability, 26:1–37, 2021.

- [CKW23] Likai Chen, Georg Keilbar, and Wei Biao Wu. Recursive Quantile Estimation: Non-Asymptotic Confidence Bounds. Journal of Machine Learning Research, 24(91):1–25, 2023.
- [CL22] Felix Cheysson and Gabriel Lang. Spectral estimation of hawkes processes from count data. The Annals of Statistics, 50(3):1722–1746, 2022.
- [CM11] N. Champagnat and S. Méléard. Polymorphic evolution sequence and evolutionary branching. Probability Theory and Related Fields, 151(1-2):45–94, 2011.
- [CMM24a] Manon Costa, Pascal Maillard, and Anthony Muraro. (Almost) complete characterization of the stability of a discrete-time Hawkes process with inhibition and memory of length two. Journal of Applied Probability, pages 1–22, May 2024.
- [CMM24b] Manon Costa, Pascal Maillard, and Anthony Muraro. (stability of discrete-time hawkes process with inhibition: towards a general condition. Arxiv preprint, 2024.
- [CMT21] Nicolas Champagnat, Sylvie Méléard, and Viet Chi Tran. Stochastic analysis of emergence of evolutionary cyclic behavior in population dynamics with transfer. The Annals of Applied Probability, 31(4):1820–1867, 2021.
- [Cos16] Manon Costa. A piecewise deterministic model for a prey-predator community. The Annals of Applied Probability, pages 3491–3530, 2016. Publisher: JSTOR.
- [CSSBW17] S. Chen, A. Shojaie, E. Shea-Brown, and D. Witten. The multivariate Hawkes process in high dimensions: Beyond mutual excitation. arXiv:1707.04928v2, 2017.
- [DE97] P. Dupuis and R.S. Ellis. A weak convergence approach to the theory of Large deviations. 1997.
- [DFH16] S. Delattre, N. Fournier, and M. Hoffmann. Hawkes processes on large networks. Annals of Applied Probability, 26(1):216–261, 2016.
- [DJN14] Mark F. Dybdahl, Christina E. Jenkins, and Scott L. Nuismer. Identifying the molecular basis of host-parasite coevolution: Merging models and mechanisms. The American Naturalist, 184(1):1–13, 2014.
- [DL96] U. Dieckmann and R. Law. The dynamical theory of coevolution: a derivation from stochastic ecological processes. Journal of mathematical biology, 34(5-6):579–612, 1996.
- [DLP22] Céline Duval, Eric Luçon, and Christophe Pouzat. Interacting hawkes processes with multiplicative inhibition. Stochastic Processes and their Applications, 148:180–226, 2022.
- [DM11] Rick Durrett and John Mayberry. Traveling waves of selective sweeps. The Annals of Applied Probability, 21(2):699–744, 2011.
- [DML95] U. Dieckmann, P. Marrow, and R. Law. Evolutionary cycling in predator-prey interactions: population dynamics and the red queen. Journal of Theoretical Biology, 176(1):91–102, 1995.
- [DMPS18] Randal Douc, Eric Moulines, Pierre Priouret, and Philippe Soulier. Markov chains. Springer Series in Operations Research and Financial Engineering. Springer, Cham, 2018.

- [DNS12] S. Dereich, A. Neuenkirch, and L. Szpruch. An Euler-type method for the strong approximation of the cox ingersoll ross process. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 468(2140):1105–1115, 2012.
- [DRR20] Sophie Donnet, Vincent Rivoirard, and Judith Rousseau. Nonparametric Bayesian estimation for multivariate Hawkes processes. The Annals of Statistics, 48(5):2698–2727, October 2020. Publisher: Institute of Mathematical Statistics.
- [Duf97] M. Duflo. Random iterative models, volume 34 of Applications of Mathematics (New York). Springer-Verlag, Berlin, 1997.
- [DVJ06] D. J. Daley and D. Vere-Jones. An Introduction to the Theory of Point Processes: Volume I: Elementary Theory and Methods. Springer Science & Business Media, April 2006.
- [EK86] N. Ethier and T.Q. Kurtz. Markov Processes Characterization and Convergence. John Wiley and Sons, 1986.
- [Fab68] Vaclav Fabian. On Asymptotic Normality in Stochastic Approximation. The Annals of Mathematical Statistics, 39(4):1327–1332, August 1968. Publisher: Institute of Mathematical Statistics.
- [Fel91] William Feller. An Introduction to Probability Theory and Its Applications, volume 2 of Wiley Series in Probability and Statistics. 2nd edition edition, 1991.
- [FL16] Nicolas Fournier and Eva Löcherbach. On a toy model of interacting neurons. Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, 52(4):1844–1876, November 2016. Publisher: Institut Henri Poincaré.
- [FLO06] René Ferland, Alain Latour, and Driss Oraichi. Integer-Valued GARCH Process. Journal of Time Series Analysis, 27(6):923–942, November 2006. Publisher: Wiley-Blackwell.
- [Fos53] F. G. Foster. On the Stochastic Matrices Associated with Certain Queuing Processes. The Annals of Mathematical Statistics, 24(3):355–360, September 1953. Publisher: Institute of Mathematical Statistics.
- [GB98] S Gavrillets and C R B Boake. On the evolution of premating isolation after a founder event. The American Naturalist, 152(5):706–716, 1998. Publisher: JSTOR.
- [GCA<sup>+</sup>19] Pauline Gonnord, Manon Costa, Arnaud Abreu, Michael Peres, Loïc Ysebaert, Sébastien Gadat, and Salvatore Valitutti. Multiparametric analysis of CD8+ T cell compartment phenotype in chronic lymphocytic leukemia reveals a signature associated with progression toward therapy. Oncoimmunology, 8(4):e1570774, February 2019.
- [Gho71] J. K. Ghosh. A New Proof of the Bahadur Representation of Quantiles and an Application. The Annals of Mathematical Statistics, 42(6):1957–1961, December 1971. Publisher: Institute of Mathematical Statistics.
- [GJK<sup>+</sup>16] F. Gamboa, A. Janon, T. Klein, A. Lagnoux, and C. Prieur. Statistical inference for Sobol pick-freeze Monte Carlo method. Statistics, 50(4):881–902, July 2016. Publisher: Taylor & Francis \_eprint: <https://doi.org/10.1080/02331888.2015.1105803>.

- [GK75] V. F. Gaposkin and T. P. Krasulina. On the Law of the Iterated Logarithm in Stochastic Approximation Processes. Theory of Probability & Its Applications, 19(4):844–850, September 1975. Publisher: Society for Industrial and Applied Mathematics.
- [GKL18] Fabrice Gamboa, Thierry Klein, and Agnès Lagnoux. Sensitivity Analysis Based on Cramér–von Mises Distance. SIAM/ASA Journal on Uncertainty Quantification, 6(2):522–548, January 2018. Publisher: Society for Industrial and Applied Mathematics.
- [GKZ94] Israel M. Gelfand, Mikhail M. Kapranov, and Andrei V. Zelevinsky. Discriminants, Resultants, and Multidimensional Determinants. Birkhäuser, Boston, MA, 1994.
- [Gla03] P. Glasserman. Monte Carlo Methods in Financial Engineering. Stochastic Modelling and Applied Probability. Springer, 2003.
- [GP23] Sébastien Gadat and Fabien Panloup. Optimal non-asymptotic analysis of the Ruppert–Polyak averaging stochastic algorithm. Stochastic Processes and their Applications, 156:312–348, February 2023.
- [GPS18] Sébastien Gadat, Fabien Panloup, and Sofiane Saadane. Stochastic heavy ball. Electronic Journal of Statistics, 12(1):461–529, January 2018. Publisher: Institute of Mathematical Statistics and Bernoulli Society.
- [Gra21] Carl Graham. Regenerative properties of the linear hawkes process with unbounded memory. The Annals of Applied Probability, 31(6):2844 – 2863, 2021.
- [Gre92] Hans-Rolf Gregorius. A two-locus model of speciation. Journal of theoretical Biology, 154(3):391–398, 1992.
- [GS10] A. Gulisashvili and E. Stein. Asymptotic behavior of the stock price distribution density and implied volatility in stochastic volatility models. Applied Mathematics and Optimization, 61(3):287–315, 2010.
- [GZ21] Fuqing Gao and Lingjiong Zhu. Precise deviations for Hawkes processes. Bernoulli, 27(1):221–248, February 2021. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- [Haw71] Alan G Hawkes. Spectra of some self-exciting and mutually exciting point processes. Biometrika, 58(1):83–90, 1971.
- [Haw18] Alan G Hawkes. Hawkes processes and their applications to finance: a review. Quantitative Finance, 18(2):193–198, 2018.
- [Hei03] Andréas Heinen. Modelling Time Series Count Data: An Autoregressive Conditional Poisson Model. MPRA Paper, University Library of Munich, Germany, July 2003.
- [HHKR22] Caroline Hillairet, Lorick Huang, Mahmoud Khabou, and Anthony Réveillac. The malliavin-stein method for hawkes functionals. ALEA, 19:1293–1328, 2022.
- [HK22] Lorick Huang and Mahmoud Khabou. The non-linear discrete-time hawkes process. hal preprint hal-03220800, 2022.
- [HO74] Alan G Hawkes and David Oakes. A cluster process representation of a self-exciting process. Journal of Applied Probability, 11(3):493–503, 1974.

- [HRBR15] Niels Richard Hansen, Patricia Reynaud-Bouret, and Vincent Rivoirard. Lasso and probabilistic inequalities for multivariate point processes. *Bernoulli*, 21(1):83–143, February 2015. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- [HS05] Marcel P Haesler and Ole Seehausen. Inheritance of female mating preference in a sympatric sibling species pair of lake victoria cichlids: implications for speciation. *Proceedings of the Royal Society of London B: Biological Sciences*, 272(1560):237–245, 2005.
- [HTPW97] Hope Hollocher, Chau-Ti Ting, Francine Pollack, and Chung-I Wu. Incipient speciation by sexual isolation in *Drosophila melanogaster*: variation in mating preference and correlation between sexes. *Evolution*, pages 1175–1181, 1997. Publisher: JSTOR.
- [HWE<sup>+</sup>07] O P Höner, B Wachter, M L East, W J Streich, K Wilhelm, T Burke, and H Hofer. Female mate-choice drives the evolution of male-biased dispersal in a social mammal. *Nature*, 448:797–802, 2007.
- [HX24] Ulrich Horst and Wei Xu. Functional Limit Theorems for Hawkes Processes, January 2024. arXiv:2401.11495 [math, q-fin, stat].
- [JR09] Adam G. Jones and Nicholas L. Ratterman. Mate choice and sexual selection: What have we learned since Darwin? *Proceedings of the National Academy of Sciences*, 106(supplement\_1):10001–10008, June 2009. Publisher: Proceedings of the National Academy of Sciences.
- [KC64] Motoo Kimura and James F Crow. The number of alleles that can be maintained in a finite population. *Genetics*, 49(4):725, 1964. Publisher: Genetics Society of America.
- [Ker77] Gotz Kersting. Almost sure approximation of the robbins-monro process by sums of independent random variables. *Ann. Probab.*, 5(6):954–965, 1977.
- [Kim97] Marek Kimmel. Quasistationarity in a Branching Model of Division-Within-Division. In Krishna B. Athreya and Peter Jagers, editors, *Classical and Modern Branching Processes*, pages 157–164. Springer, New York, NY, 1997.
- [Kir16] Matthias Kirchner. Hawkes and  $\text{inar}(\infty)$  processes. *Stochastic Processes and their Applications*, 126(8):2494–2525, 2016.
- [KPU01] P. Krokmal, J. Palmquist, and S. Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4(2):43–68, 2001.
- [KR07] Elizabeth Kelly and Stephen J Russell. History of oncolytic viruses: genesis to genetic engineering. *Molecular therapy*, 15(4):651–659, 2007.
- [KT04] Vijay R. Konda and John N. Tsitsiklis. Convergence rate of linear two-time-scale stochastic approximation. *The Annals of Applied Probability*, 14(2):796–819, May 2004. Publisher: Institute of Mathematical Statistics.
- [L<sup>+</sup>17] Eva Löcherbach. Spiking neurons: interacting Hawkes processes, mean field limits and oscillations. In *Journées MAS 2016 de la SMAI—Phénomènes complexes et hétérogènes*, volume 60 of *ESAIM Proc. Surveys*, pages 90–103. EDP Sci., Les Ulis, 2017.

- [Lan81] Russell Lande. Models of speciation by sexual selection on polygenic traits. Proceedings of the National Academy of Sciences, 78(6):3721–3725, June 1981. Publisher: Proceedings of the National Academy of Sciences.
- [LFDE13] Pepijn Luijckx, Harris Fienberg, David Duneau, and Dieter Ebert. A matching-allele model explains host resistance to parasites. Current Biology, 23(12):1085–1088, 2013.
- [LGT78] RC Lewontin, LR Ginzburg, and SD Tuljapurkar. Heterosis as an explanation for large amounts of genic polymorphism. Genetics, 88(1):149–169, 1978. Publisher: Oxford University Press.
- [LL54] JJ Levin and Norman Levinson. Singular perturbations of non-linear systems of differential equations and an associated boundary layer equation. Journal of rational mechanics and analysis, 3:247–270, 1954.
- [LLT21] Patrick J. Laub, Young Lee, and Thomas Taimre. The Elements of Hawkes Processes. Springer International Publishing, Cham, 2021.
- [LM22] Eva Löcherbach and Pierre Monmarché. Metastability for systems of interacting neurons. Annales de l’Institut Henri Poincaré, Probabilités et Statistiques, 58(1):343–378, February 2022. Publisher: Institut Henri Poincaré.
- [LMZ11] Raphaël Lefevre, Mauro Mariani, and Lorenzo Zambotti. Large deviations for renewal processes. Stochastic Processes and their Applications, 121(10):2243–2271, October 2011.
- [LNS12] G. Lan, A. Nemirovskij, and A. Shapiro. Validation analysis of mirror descent stochastic approximation method. Math. Program., Ser. A, (134):425–458, 2012.
- [LR79] T. L. Lai and H. Robbins. Adaptive design and stochastic approximation. Ann. Statist., 7(6):1196–1221, 1979.
- [LTMB<sup>+</sup>18] Régis C Lambert, Christine Tuleau-Malot, Thomas Bessaih, Vincent Rivoirard, Yann Bouret, Nathalie Leresche, and Patricia Reynaud-Bouret. Reconstructing the functional connectivity of multiple spike trains using hawkes models. Journal of neuroscience methods, 297:9–21, 2018.
- [LV14] Remi Lemonnier and Nicolas Vayatis. Nonparametric markovian learning of triggering kernels for mutually exciting and mutually inhibiting multivariate hawkes processes. In Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2014, Nancy, France, September 15-19, 2014. Proceedings, Part II 14, pages 161–176. Springer, 2014.
- [LZLS23] Mengya Liu, Fukang Zhu, Jianfeng Li, and Chuning Sun. A Systematic Review of INGARCH Models for Integer-Valued Time Series. Entropy, 25(6):922, June 2023.
- [MBJ<sup>+</sup>21] Ludovic Maisonneuve, Thomas Beneteau, Mathieu Joron, Charline Smadi, and Violaine Llaurens. When do opposites attract? A model uncovering the evolution of disassortative mating. The American Naturalist, 198(5):000–000, 2021. Publisher: The University of Chicago Press Chicago, IL.
- [MG21] David V McLeod and Sylvain Gandon. Understanding the evolution of multiple drug resistance in structured populations. eLife, 10:e65645, June 2021. Publisher: eLife Sciences Publications, Ltd.

- [MGG02] C Matessi, A Gimelfarb, and S Gavrillets. Long-term buildup of reproductive isolation promoted by disruptive selection: how far does it go? *Selection*, 2(1-2):41–64, 2002. Publisher: Akadémiai Kiadó.
- [MGM<sup>+</sup>96] J. A.J. Metz, S. A.H. Geritz, G. Meszéna, F. J.A. Jacobs, and J.S. Van Heerwaarden. Adaptive dynamics, a geometrical study of the consequences of nearly faithful reproduction. *Stochastic and spatial structures of dynamical systems*, 45:183–231, 1996.
- [MLC92] P. Marrow, R. Law, and C. Cannings. The coevolution of predator–prey interactions: Esss and red queen dynamics. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 250(1328):133–141, 1992.
- [MMOD12] L K M’Gonigle, R Mazzucco, S P Otto, and U Dieckmann. Sexual selection enables long-term coexistence despite ecological equivalence. *Nature*, 484(7395):506–509, 2012.
- [MP06] A. Mokkadem and M. Pelletier. Convergence rate and averaging of nonlinear two-time-scale stochastic approximation algorithms. *The Annals of Applied Probability*, 16, November 2006.
- [MS18] P. Mertikopoulos and M. Staudigl. On the convergence of gradient-like flows with noisy gradient input. *SIAM Journal on Optimization*, 28(1):163–197, 2018.
- [MS21] Aline Marguet and Charline Smadi. Long time behaviour of continuous-state nonlinear branching processes with catastrophes. *Electronic Journal of Probability*, 26(none):1–32, January 2021. Publisher: Institute of Mathematical Statistics and Bernoulli Society.
- [MS24] Aline Marguet and Charline Smadi. Spread of parasites affecting death and division rates in a cell population. *Stochastic Processes and their Applications*, 168:104262, February 2024.
- [MT09] S. P. Meyn and R. L. Tweedie. *Markov Chains and Stochastic Stability*. Cambridge Mathematical Library. Cambridge University Press, 2009.
- [Nag77] Thomas Nagylaki. *Selection in One- and Two-Locus Systems*, volume 15 of *Lecture Notes in Biomathematics*. Springer, Berlin, Heidelberg, 1977.
- [Nag93] T. Nagylaki. The Evolution of Multilocus Systems under Weak Selection. *Genetics*, 134(2):627–647, June 1993.
- [NCL21] Phuong Linh Nguyen, Manon Costa, and Nicolas Loeuille. Implications of drift and rapid evolution on negative niche construction, April 2021. Pages: 2021.04.26.441094 Section: New Results.
- [Nei75] Masatoshi Nei. *Molecular population genetics and evolution*. North-Holland Publishing Company., 1975.
- [NY83] A. Nemirovskij and D. Yudin. *Problem complexity and method efficiency in optimization*. Wiley-Interscience, 1983.
- [Oga78] Y. Ogata. The asymptotic behaviour of maximum likelihood estimators for stationary point processes. *Annals of the Institute of Statistical Mathematics*, 30:243–261, 1978.
- [OSN08] Sarah P Otto, Maria R Servedio, and Scott L Nuismer. Frequency-Dependent Selection and the Evolution of Assortative Mating. *Genetics*, 179(4):2091–2112, August 2008.

- [Pel98] Mariane Pelletier. On the almost sure asymptotic behaviour of stochastic algorithms. Stochastic Process. Appl. 78, 2:217–244, 1998.
- [PGY<sup>+</sup>18] Andrei Papkou, Thiago Guzella, Wentao Yang, Svenja Koepper, Barbara Pees, Rebecca Schalkowski, Mike-Christoph Barg, Philip C. Rosenstiel, Henrique Teotónio, and Hinrich Schulenburg. The genomic basis of red queen dynamics during rapid reciprocal host–pathogen coevolution. Proceedings of the National Academy of Sciences, 116(3):923–928, 2018.
- [Pit82] M. Pitman, J. Yor. A decomposition of Bessel bridges. Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete, 59(4):425–457, 1982.
- [PJ92] B. T. Polyak and A. Juditsky. Acceleration of Stochastic Approximation by Averaging. SIAM Journal on Control and Optimization, 30(4):838–855, 1992.
- [PK97] R. J. H. Payne and D. C. Krakauer. Sexual selection, space, and speciation. Evolution, 51(1):1–9, 1997.
- [Raa19] Mads Bonde Raad. Renewal Time Points for Hawkes Processes, June 2019. arXiv:1906.02036 [math].
- [RBR07] Patricia Reynaud-Bouret and Emmanuel Roy. Some non asymptotic tail estimates for Hawkes processes. Bulletin of the Belgian Mathematical Society - Simon Stevin, 13(5):883–896, January 2007. Publisher: The Belgian Mathematical Society.
- [RBRGTM14] Patricia Reynaud-Bouret, Vincent Rivoirard, Franck Grammont, and Christine Tuleau-Malot. Goodness-of-fit tests and nonparametric adaptive estimation for spike train analysis. The Journal of Mathematical Neuroscience, 4:1–41, 2014.
- [Rit07] Michael G. Ritchie. Sexual Selection and Speciation. Annual Review of Ecology, Evolution, and Systematics, 38(1):79–102, 2007.
- [RM51] Herbert Robbins and Sutton Monro. A Stochastic Approximation Method. The Annals of Mathematical Statistics, 22(3):400–407, September 1951. Publisher: Institute of Mathematical Statistics.
- [Ros17] Gil G. Rosenthal. Mate Choice: The Evolution of Sexual Decision Making from Microbes to Humans. Princeton University Press, 2017.
- [RR14] R. Rockafellar and J. Royset. Random variables, monotone relations, and convex analysis. Mathematical Programming, 148(1):297–331, 2014.
- [RS71] H. Robbins and D. Siegmund. A CONVERGENCE THEOREM FOR NON NEGATIVE ALMOST SUPERMARTINGALES AND SOME APPLICATIONS\*. In Jagdish S. Rustagi, editor, Optimizing Methods in Statistics, pages 233–257. Academic Press, January 1971.
- [RU00] R. T. Rockafellar and S. Uryasev. Optimization of Conditional Value-at-Risk. The Journal of Risk, 2(3), 2000.
- [RU02] R. Tyrrell Rockafellar and Stanislav Uryasev. Conditional value-at-risk for general loss distributions. Journal of Banking & Finance, 26(7):1443–1471, July 2002.



- [Rup88] D. Ruppert. Efficient estimations from a slowly convergent Robbins-Monro process. Technical Report, 781, Cornell University Operations Research and Industrial Engineering, 1988.
- [RY23] Luis Iván Hernández Ruíz and Kouji Yano. A cluster representation of the renewal Hawkes process, April 2023. arXiv:2304.06288 [math].
- [Sac58] Jerome Sacks. Asymptotic Distribution of Stochastic Approximation Procedures. The Annals of Mathematical Statistics, 29(2):373–405, June 1958. Publisher: Institute of Mathematical Statistics.
- [SB14] Maria R Servedio and Reinhard Bürger. The counterintuitive role of sexual selection in species maintenance and speciation. Proceedings of the National Academy of Sciences, 111(22):8113–8118, 2014. Publisher: National Acad Sciences.
- [Sch88] P L Schwagmeyer. Scramble-competition polygyny in an asocial mammal: Male mobility and mating success. The American Naturalist, 131:885–892, 1988.
- [Ser10] M R Servedio. Limits to the evolution of assortative mating by female choice under restricted gene flow. Proceedings of the Royal Society of London B: Biological Sciences, 278(1703):179–187, 2010.
- [SHK<sup>+</sup>17] DN Santiago, JPW Heidbuechel, WM Kandell, R Walker, J Djeu, CE Engeland, D Abate-Daga, and H Enderling. Fighting cancer with mathematics and viruses. Viruses, 9:239, 2017.
- [SLL18] Charline Smadi, Hélène Leman, and Violaine Llaurens. Looking for the right mate in diploid species: How does genetic dominance affect the spatial differentiation of a sexual trait? Journal of theoretical biology, 447:154–170, 2018. Publisher: Elsevier.
- [SMP<sup>+</sup>19] Thorsten Stefan, Louise Matthews, Joaquin M Prada, Colette Mair, Richard Reeve, and Michael J Stear. Divergent allele advantage provides a quantitative model for maintaining alleles with a wide range of intrinsic merits. Genetics, 212(2):553–564, 2019. Publisher: Oxford University Press.
- [Sob01] I. M Sobol. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Mathematics and Computers in Simulation, 55(1):271–280, February 2001.
- [SRA<sup>+</sup>08] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, and S. Tarantola. Global sensitivity analysis: the primer. Wiley Online Library, 2008.
- [SRR22] Deborah Sulem, Vincent Rivoirard, and Judith Rousseau. Scalable variational bayes methods for hawkes processes. arXiv preprint arXiv:2212.00293, 2022.
- [SRR24] Déborah Sulem, Vincent Rivoirard, and Judith Rousseau. Bayesian estimation of non-linear Hawkes processes. Bernoulli, 30(2):1257–1286, May 2024. Publisher: Bernoulli Society for Mathematical Statistics and Probability.
- [SST20] H. Schenk, H. Schulenburg, and A. Traulsen. How long do red queen dynamics survive under genetic drift? a comparative analysis of evolutionary and eco-evolutionary models. BMC Evol Biol, 20(8), 2020.

- [Sto70] William F. Stout. A martingale analogue of Kolmogorov's law of the iterated logarithm. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 15(4):279–290, December 1970.
- [Tak62] L. Takács. Introduction to the theory of queues. University Texts in the Mathematical Sciences. Oxford University Press, New York, 1962.
- [Thu14] Måns Thulin. The cost of using exact confidence intervals for a binomial proportion. Electronic Journal of Statistics, 8(1):817–840, January 2014. Publisher: Institute of Mathematical Statistics and Bernoulli Society.
- [WFS16] Spencer Wheatley, Vladimir Filimonov, and Didier Sornette. The Hawkes process with renewal immigration & its estimation with an EM algorithm. Computational Statistics & Data Analysis, 94:120–135, February 2016.
- [Wod03] Dominik Wodarz. Gene therapy for killing p53-negative cancer cells: use of replicating versus nonreplicating agents. Human gene therapy, 14(2):153–159, 2003.
- [Woo72] Michael Woodroffe. Normal approximation and large deviations for the robbins monro process. Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete, 21(4):329–338, December 1972.
- [Zhu13] Lingjiong Zhu. Central limit theorem for nonlinear Hawkes processes. J. Appl. Probab., 50(3):760–771, 2013.